DISSECTING THE DIFFERENTIAL MOLECULAR SIGNATURE IN PATIENTS FROM AFRICAN-ORIGIN WITH TRIPLE-NEGATIVE BREAST CANCER

ANA TERESA DE SOUSA E MENEZES MATIAS

Tese para obtenção do grau de Doutor em Ciências da Saúde

na Especialidade em Biomedicina

na Faculdade de Ciências Médicas | NOVA Medical School da Universidade NOVA de Lisboa

março, 2021

DISSECTING THE DIFFERENTIAL MOLECULAR SIGNATURE IN PATIENTS FROM AFRICAN-ORIGIN WITH TRIPLE-NEGATIVE BREAST CANCER

Ana Teresa de Sousa e Menezes Matias

António Jacinto, Investigador Principal e Prof. Associado, NMS-FCM

Guadalupe Cabral, PhD, Professora Auxiliar Convidada, NMS-FCM

Sofia Braga, PhD, Assistente Hospitalar Graduada, NMS-FCM, Hospital Prof. Doutor Fernando Fonseca, Hospital Vila Franca de Xira e Cuf Oncologia

Patrícia Homem de Brito, PhD, Investigadora Independente, UCIBIO, FCT-UNL, e Professora Auxiliar Convidada, NMS-FCM

Tese para obtenção do grau de Doutor em Ciências da Saúde

na Especialidade em Biomedicina

março, 2021

The research described here was performed at the Chronic Diseases Research Center (CEDOC), NOVA Medical School – Faculdade de Ciências Médicas, between 2016 and 2021. Its execution was supported by a PhD fellowship from the Portuguese Foundation for Science and Technology (PD/BD/114053/2015). This work was accomplished under the supervision of Dr António Jacinto, Dr Guadalupe Cabral, Dr Sofia Braga and Dr Patrícia Homem de Brito.

Este trabalho de investigação foi realizado no Centro de Estudos de Doenças Crónicas (CEDOC) da NOVA Medical School – Faculdade de Ciências Médicas, entre 2016 e 2021, ao abrigo de uma bolsa de Doutoramento financiada pela Fundação para a Ciência e Tecnologia (PD/BD/114053/2015). Esta tese de doutoramento foi supervisionada pelo Doutor António Jacinto, Doutora Guadalupe Cabral, Doutora Sofia Braga e pela Doutora Patrícia Homem de Brito.

“Nothing in life is to be feared, it is only to be understood.

Now is the time to understand more, so that we may fear less.”

— Marie Skłodowska Curie

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

And here I am, writing the acknowledgments. It is an incredibly feeling - a mixture between accomplishment, nostalgia, and relief.

Many people were part of this journey, directly or indirectly. I thank everybody who helped me to get here and without whom this work would not be possible.

First of all, I would like to thank my supervisor António Jacinto for giving me the opportunity to be part of the amazing Tissue Repair and Inflammation lab and for supervising my PhD work. I would like to specially thank you for allowing me to explore a completely different approach than what is usual in the lab.

I would like to thank my co-superiors Guadalupe Cabral, Sofia Braga e Patrícia Brito. Firstly, I must thank Guadalupe "Lupita" for the kind words, proximity and friendship. To Sofia for the enthusiasm and mentorship of this project. And last but not least, to Patrícia, for all the late hour meetings and encouragement. To you three, thank you for your advice, availability and support throughout the years and specially for trusting me and my work, even when I doubted myself. I have learned a lot from you and your passion for science truly inspired and motivated me!

I would also like to thank Ana Teresa Maia and her group at Universidade do Algarve for receiving me and introducing me to the wonderful world of applied bioinformatics and R environment. Besides Ana Teresa herself, I must thank Ramiro Magno, Joana Xavier and Ana Fernandes for their guidance and teachings.

To the Pathological Anatomy Service from Hospital Professor Doutor Fernando Fonseca (HFF), in particularly to Dr. Gabriela Gasparinho, Dr. Lucília Gonçalves, Dr. Mário Lagaillarde and Dr. Ana Rodrigues, for the important collaboration in this work. A big thanks to Dr. Marta Vaz, not only for helping in collecting patients' medical information, essential to identify the patients of interest, but also for always being so helpful and friendly. And, last but not least, to all the breast cancer patients that participated in this work.

To Patrícia Alves and Bruno Alexandre from iBET, for kindly allowing me the use of their Ingenuity Pathway Analysis (IPA) software.

To Ana Farinho, for all the sympathy and important suggestions regarding the immunohistochemistry protocol and its optimization, and to Telmo Pereira, for his precious help with the microscopy and suggestions regarding imaging analysis.

I would like to thank the incredible Tissue Morphogenesis and Repair lab members! To the present and the former ones, thank you for your positivism in times of need and for the wonderful chocolate hours! To my dear SuperHuman girls, Diana "Di" Saraiva, Rute Salvador and Bruna Correia, for the "onco"-companionship throughout the years; to the amazing Fish&Chicks, Ana Sofia "Sofs" Brandão, Jorge "Borbs" Borbinha, José Marques, Rita Gorgulho, Diogo Paramos, Valdir Semedo, Nídia Sousa, Nuno Silva, Pedro Seco and Sílvia Coelho; to the incredible FireFlies Susana Ponte, Mafalda Pinho and Patrícia Rodrigues. And last but not least, to the dear TRI postdocs, Lara Carvalho, Carolina Crespo, Ana Teresa Tavares and Raquel Lourenço. To you all, thank you for your good energy, joy, teachings, friendship and passion for science. You're the best!

I would also like to thank the incredible CEDOC community. Thank you all for your support and good times for the past 5 years (and counting).

To the members of the coordination of the Health Sciences - Biomedicine PhD Program for accepting me and to FCT for funding this PhD thesis.

To my friends, you were a central part of this journey.

To Inês and Mariana, for the companionship in this journey that started more than 10 years ago. Thank you for listening and for sharing the emotional roller coaster we all have been through in these past years. It is reassuring to know that we are not alone.

To my “little friends” Ana Irene, Cláudio e João. What almost 25 years of friendship has brought together, a PhD cannot end.

To my strange and loud friends, companions in the best and worst moments. To my dear “má onda” girls, Ana “Hara” and Bárbara; to my “little sisters” Ana “Mizu” and Joana “Juju”; to the boys Filipe, Rui “Muta”, Francisco, José, João Luís, Guilherme, André Lonet, Ricardo e Daniel. Thank you for your friendship, support, silliness and for the long, nonsensical and essential talks.

Um grande agradecimento à minha família, em especial:

Aos meus pais, Paulo e Cristina, que independente de tudo, sempre me apoiaram incondicionalmente em todas as decisões que tomei. Muito obrigada por tudo!

À minha irmã, pela sua incrível sapiência e palavras de apoio, dada a sua idade. E também pela parvoíce, indispensável desde sempre, mas, particularmente, nestes últimos anos. Obrigada, Fedelha!

À minha avó Jerónima, que vindo de meios muito humildes nunca deixou de desejar e de fazer o melhor pelas netas. Não chegaria onde estou agora sem o seu apoio.

E porque o melhor fica para o fim, um enorme agradecimento a ti, Rúben. Muito obrigada pelo extraordinário apoio, por me conheceres melhor do que ninguém e, simplesmente, por existires. Mesmo percebendo pouco ou nada deste trabalho, esta tese também é tua!

THESIS CONTRIBUTIONS

All Chapters and figures integrating this thesis were written and assembled by the author.

All computational and experimental procedures and analysis were performed by the author.

THESIS PUBLICATIONS

The work presented in this thesis is published in *bioRxiv*:

Ana T. Matias, Ana Jacinta-Fernandes, Ana-Teresa Maia, Sofia Braga, António Jacinto, M. Guadalupe Cabral, Patrícia H. Brito. Differential expression between African-ancestry and White patients diagnosed with Triple-Negative Breast Cancer: EGFR, Myc, Bcl2 and β-Catenin as ancestry-associated markers. November 15, 2020. doi: https://doi.org/10.1101/2020.11.13.381608.

And is currently under revision in *Breast Cancer Research and Treatment*:

Matias AT, Jacinta-Fernandes A, Maia A-T, Braga S, Jacinto A, Cabral MG & Brito PH. Differential expression between African-ancestry and White patients diagnosed with Triple-Negative Breast Cancer: EGFR, Myc, Bcl2 and β-Catenin as ancestry-associated markers.

An abstract with part of the project's results is also publish in a supplement issue of *Annals of Oncology*:

A.T. Matias, A. Jacinta-Fernandes, R. Magno, J. Xavier, M.G. Cabral, A. Jacinto, A.T. Maia, S. Braga. Differential molecular signature in patients from African origin with Triple-Negative Breast Cancer, Annals of Oncology. 2018; Supplement 8(29).

Preliminary results of this work were presented in poster format in different meetings:

Ana Teresa Matias, Ana Jacinta-Fernandes, Ramiro Magno, Joana Xavier, António Jacinto, Sofia Braga & Ana-Teresa Maia. Differential molecular signature in patients from African origin with Triple-Negative Breast Cancer. In: 3rd ASPIC International Congress, Lisbon, Portugal, 2018.

Matias, A.T., Cabral, G., Jacinto, A., Maia, A.-T., Braga, S. Differential molecular signature in patients from African origin with Triple-Negative Breast Cancer. In: EACR-AACR-ISCR Conference: The Cutting Edge of Contemporary Cancer Research 2018, Jerusalem, Israel, 2018.

Matias, A.T., Jacinta-Fernandes, A., Magno, R., Xavier, J.M., Cabral, G., Jacinto, A., Maia, A.-T., Braga, S. Differential molecular signature in patients from African origin with Triple-Negative Breast Cancer. In: ESMO 2018 Congress, Munich, Germany, 2018.

During the course of this PhD, I also contributed to the following publications:

Saraiva DP, **Matias AT**, Braga S, Jacinto A & Cabral MG, Establishment of a 3D co-culture with MDA-MB-231 breast cancer cell line and patient-derived immune cells for application in the development of immunotherapies. Front. Oncol 2020; 10:1543.

Joana Sales-Dias, Gabriela Silva, Diogo Casal, Sara Alves, Cristina Casalou, M. Clara Barreto, Giacomo Domenici, **Ana Teresa Matias**, Andreia Ferreira, Ana R. Lemos, Carolina Matos, Maria Raquel Moita, Catarina Brito, Pedro M. F. Sousa, Guadalupe Cabral, Sofia Braga, Tiago M. Bandeiras, Paula Videira, Ana Barbas. Development of an anti-DLL1 IgG that inhibits ER+ breast tumour growth. Manuscript in preparation for submission.

I am also one of the authors of the following book chapter:

H. Gervásio, J.V. Paulo, J.L. Coelho, S.A. Braga, M.V. Batista e **A.T. Matias**. *Capítulo 6 - Cancro da mama triplo negativo*. In F. Cardoso (Coordenação), 100 perguntas chave no cancro de mama. 2ª Edição, Permanyer Portugal, Lisboa, 2017: 41-52.

ABSTRACT

Breast cancer (BC) is the most commonly occurring cancer in women and the second most common cancer worldwide. BCs comprise a diverse group of tumors, which are characterized by a wide spectrum of clinical, pathologic, and molecular features. This wide range of factors accounts for disparities in response to therapy and outcomes among patients. The main determinant for the prognosis of a BC patient is the expression of estrogen receptor (ER), progesterone receptor (PR) and Human Epidermal Growth Factor Receptor 2 (HER2). Patients presenting these biomarkers have specific targeted treatments and favorable prognosis. However, patients with the triple-negative breast cancer (TNBC) subtype do not express the conventional biomarkers ER, PR, and HER2, being "negatively defined". Also, TNBCs are not only remarkably inter- and intra-heterogeneous, as they usually present a more aggressive behavior, with large locally advanced breast lesion or metastatic disease developing shortly after adjuvant chemotherapy. Thus, TNBC is considered one of the most relevant unmet medical problems in BC management and women with this subtype have worse outcome compared to other BC patients.

Epidemiologically, TNBC comprises 10-20% of BC cases globally. Nonetheless, TNBC prevalence and mortality is higher in women of African-ancestry, particularly in younger patients. Even when accounting for socioeconomic factors and independently of other risk factors, African-ancestry patients present a faster and more aggressive clinical course of TNBC, comparing with other populations. However, Ancestry-specific molecular mechanisms involved in the increased aggressiveness and earlier age of onset of TNBC in Black-origin patients remain poorly clarified.

The present PhD thesis aimed to unravel African ancestry-associated transcriptomic patterns in TNBC, particularly when controlling for menopause status or pathological staging, as well as to identify signaling pathways and network regulators potentially involved in the higher incidence and aggressiveness of TNBC in African-ancestry women in comparison with White women.

We started by exploring the clinicopathological data from a group of BC patients comprised by African-ancestry and White women, from The Cancer Genome Atlas (TCGA), and observed a significantly higher TNBC incidence in the African-American (AA) group. Given the reported younger age of onset and poorer prognosis of African-ancestry BC cases, patients were then grouped according to their menopause status and pathological stage of the disease. Interestingly, we observed particularly in premenopausal and stage II disease a significantly lower survival probability of African-American patients comparing with their White counterparts. These observations prompt us to perform differential gene expression analysis (DGEA), using RNA-sequencing data, between African-ancestry and White TNBC patients, controlling for the menopause status and/or pathological stage of the disease. Remarkably, despite studies consistently highlighting the earlier age of TNBC onset in African-ancestry patients, our results suggest that differences in transcription profile between patients groups tend to be greater when controlling for disease's pathological stage than when controlling for menopause status. To exclude the identified differences which can be observed in normal mammary tissue, we performed identical DGEA with RNA-sequencing data from normal-adjacent BC tissue from AA and White patients from TCGA. Finally, we also compared the expression profile of TNBCs and normal breast tissue from AA patients, to identify gene expression alterations that the breast tissue acquires in order to develop TNBC in this group of patients. With thousands of differentially expressed genes recovered among the different groups of patients, DGEA results showed that there are indeed gene expression ancestry-associated differences to justify TNBC's epidemiological and clinical

observations. Thus, we proceeded to refine DGEA results, in order to find potential leading molecules involved in the ancestry-associated TNBC discrepancy.

Gene set enrichment analysis (GSEA), using collections of diseases and pathways from Kyoto Encyclopedia of Genes and Genomes (KEGG), was performed to identify altered pathways and biological processes between the different groups of AA and White patients, taking into account patients' menopausal status and disease's pathological stage. Among the enriched gene sets observed, we highlighted Hippo signaling pathway as the most downregulated gene set identified in AA patients, comparing with matching White patients, in the following established groups: all TNBC, postmenopausal, stage II, and postmenopausal/stage II. This downregulation is potentially promoting tumorigenesis in AA patients, since downregulation of Hippo pathway promotes the transcription of genes involved in cell proliferation and competition, cell death inhibition, epithelial-to-mesenchymal transition and tumor metastasis. Furthermore, gene sets from the cellular community and cell motility gene collections are also downregulated in AA patients from the previously mentioned groups, suggesting that the integrity of cell-cell contacts and actin cytoskeleton organization, fundamental in epithelial tissue homeostasis, are impaired, promoting tumor cells dissociation and subsequent metastasis, in agreement with the poorer prognosis of these patients. Also, since BC gene set, comprised by genes involved in the tumorigenic process in the breast, is consistently upregulated in AA patients in the same comparisons, we hypothesize that AA patients have a higher activation of components and processes involved in BC development, which may explain TNBC's faster progression and aggressive behavior of these particular patients. Finally, we highlight that MAPK signaling pathway, which regulate a variety of cellular processes, including proliferation, differentiation, apoptosis and stress responses, playing a crucial role in the survival and development of cancer cells, is upregulated specifically in stage II disease AA patients, suggesting that this pathway may contribute to the observed discrepancy in the survival rate between the AA and White TNBC patients in this stage of the disease.

To identify potential driving molecules involved in the ancestry-associated TNBC discrepancy, we also performed Ingenuity Pathway Analysis (IPA)'s core analysis with focus on network design, using the complete lists of genes obtained through DGEA. Networks in IPA produced a map of known interactions between molecules, where highly interconnected networks are likely to represent significantly altered biological functions and where central network regulators, i.e., differentially expressed genes with more direct and indirect connections in its network, are potential drivers for ancestry-associated biological differences. IPA's network design highlighted a central role for EGFR, Myc and Bcl2 genes in discriminating differences in transcription profiles of AA and White TNBC patients. These findings triggered our interest for the following reasons. EGFR, which promote cell proliferation, motility, and survival, via activation of various downstream signaling pathways, including MAPK and PI3K-AKT-mTOR, and is frequently reported as overexpressed in several cancers, including TNBC, is downregulated in AA patients, comparing with matching White patients in our study. This result is in line with a previous report, in which EGFR expression in BC and TNBC has a nonlinear relation with disease outcome, with both lower and higher expression associated with worse outcome, compared to intermediate expression levels. Thus, we suggest that underexpression of EGFR may promote a worse outcome specifically in African-ancestry TNBC patients. Myc, which is involved in cell growth, proliferation, metabolism, differentiation, stress pathways, drug resistance and apoptosis, and is stated as being overexpressed in TNBC in studies mostly composed by White patients or cell lines, has here an even higher expression in the African-ancestry patients, which may contribute to the poorer outcome of this group of patients. Finally, Bcl2, which acts by promoting cell survival instead of

driving cell proliferation, is downregulated specifically in premenopausal African-ancestry patients. In line with previous studies, Bcl2's negative expression in TNBC was significantly associated with high proliferation and an increased risk of death and recurrence. Interestingly, *BCL2* is also significantly downregulated in TNBC compared to normal-adjacent BC tissue in AAs, suggesting that this transcription impairment may contribute to TNBC development in premenopausal African-ancestry patients.

Thus, we selected EGFR, Myc and Bcl2 for further validation in samples from a group of TNBC patients followed in a Portuguese hospital, through immunohistochemistry (IHC). Additionally, we also included β-Catenin in this validation, which we observed to be downregulated in AA patients presenting postmenopausal status and stage II disease, due to the crescent number of studies about Wnt/β-catenin signaling as a therapeutic target in many tumors, as well as because β-Catenin expression has been consensually reported as being required for TNBC development by controlling numerous tumor-associated properties, including migration, stemness, anchorage-independent growth and chemosensitivity.

Remarkably, even though DGEA were performed with USA patients' RNA-sequencing data and IHC in patients' samples from a Portuguese hospital, gene and protein differential expression between African-ancestry and White patients show equivalent results. Since AA population is predominantly of Niger-Kordofanian/western Africa ancestry, and most of the African-ancestry patients followed in Portugal are from Western Africa countries, our results suggest that at least patients with Western Africa ancestry, besides having an higher incidence of TNBC, as reported in other studies, also share some TNBC-associated molecular features, distinct from White patients.

Taken together, this PhD thesis provides novel insights into ancestry-associated gene expression patterns between African-ancestry and White TNBCs, particularly when controlling for patients' menopause status or pathological stage of the disease at diagnosis. In particular, we observed that EGFR, Myc, Bcl2 and β-catenin gene and protein differential expression match in geographically and culturally distinct populations, suggesting these markers as being important indicators of TNBC's ancestry-associated development. Thus, we expect that these findings have a significant impact in the prognosis and selection of therapeutic strategies specific for African-ancestry TNBC patients.

Keywords: Triple-negative breast cancer. Ancestry-associated disease risk . African ancestry . Differential gene expression . RNA-sequencing data

RESUMO

O cancro da mama (CM) é o cancro que ocorre mais frequentemente em mulheres e o segundo mais comum a nível mundial. Os CM englobam um grupo diverso de tumores, que são caracterizados por um amplo espetro de características clínicas, patológicas e moleculares. Esta ampla gama de fatores leva a disparidades na resposta a terapias por parte dos doentes. O principal determinante no prognóstico de um doente com CM é a expressão do recetor de estrogénio (RE), do recetor de progesterona (RP) e do *Human Epidermal Growth Factor Receptor 2* (HER2). Doentes que apresentem estes biomarcadores têm terapias alvo específicas e um prognóstico mais favorável. No entanto, doentes que apresentem o subtipo de cancro da mama triplo-negativo (CMTN) não expressam os biomarcadores convencionais RE, RP e HER2, e são, por isso, "definidos negativamente". Além disso, os CMTN são notavelmente inter- e intra-heterogéneos, e apresentam habitualmente um comportamento mais agressivo, com grandes lesões localmente avançadas ou desenvolvimento de metástases imediatamente após quimioterapia adjuvante. Assim, os CMTN são considerados um dos mais relevantes problemas médicos em CM e mulheres com este subtipo têm piores resultados clínicos do que doentes com outros subtipos de CM.

Epidemiologicamente, o CMTN perfaz 10-20% dos casos globais de CM. No entanto, a prevalência e mortalidade do CMTN são mais elevadas em mulheres de ascendência africana, particularmente em doentes mais jovens. Mesmo quando considerando fatores socioeconómicos e independentemente de outros fatores de risco, mulheres de ascendência africana apresentam uma evolução mais rápida e agressiva do CMTN, comparativamente a doentes de outras populações. No entanto, os mecanismos moleculares envolvidos nesta discrepância associada à ancestralidade das doentes e relacionados com uma maior agressividade e idade mais jovem destas doentes ainda está pouco clarificado.

A presente tese de doutoramento tem como objetivo principal desvendar os padrões transcriptómicos associados à ancestralidade de doentes de ascendência africana com CMTN, particularmente quando controlando para a condição menopáusica das doentes e para o estadio patológico da doença, e identificar vias de sinalização e reguladores de redes moleculares que estão potencialmente envolvidos na incidência mais elevada e agressividade do CMTN em mulheres de ascendência africana, comparando com doentes caucasianas.

Começamos por explorar os dados clinicopatológicos de um grupo de doentes com CM composto por mulheres de ascendência africana e caucasiana, retirados do *The Cancer Genome Atlas* (TCGA), onde observámos uma incidência significativamente mais elevada de CMTN entre as doentes Afro-Americanas (AA). Considerando dados anteriormente publicados em como as doentes com CM de ascendência africana são diagnosticadas numa idade mais jovem e têm pior prognóstico, as doentes foram então agrupadas de acordo com a sua condição menopáusica e com o estadio patológico da doença. Curiosamente, observámos uma probabilidade de sobrevivência significativamente mais baixa especificamente em doentes AA no estado pré-menopáusico e com a doença no estadio II, quando comparando com as suas homólogas caucasianas. Estas observações levaram-nos a realizar análises de expressão génica diferencial (*differential gene expression analysis* - DGEA), usando dados de sequenciação de RNA, entre doentes AA com CMTN e doentes caucasianas, controlando para a condição menopáusica e/ou para o estadio patológico da doença. Notavelmente, e apesar de diferentes estudos sublinharem o facto de que doentes com CMTN e ascendência africana são diagnosticadas numa idade mais jovem, os nossos resultados sugerem que as diferenças no perfil de

transcrição entre as duas populações são maiores quando controlando para o estadio patológico da doença do que quando controlando para a condição menopáusica da doente. Para excluir os genes diferencialmente expressos entre as duas populações que também o são em tecido mamário normal, realizamos uma DGEA comparando dados se sequenciação de RNA de tecido normal adjacente a CM de doentes AA e caucasianas do TCGA. Finalmente, também comparámos o perfil de expressão de CMTN e tecido mamário normal adjacente a CM de doentes AA, de modo a identificarmos alterações na expressão génica que o tecido mamário adquire para dar origem ao CMTN nesta população. Com milhares de genes diferencialmente expressos identificados nos diferentes grupos de doentes, os resultados da DGEA mostraram que, de facto, existe uma expressão génica diferencial associada à ancestralidade das doentes com CMTN.

Posteriormente, efetuámos uma análise de enriquecimento de grupos de genes (*Gene set enrichment analysis* – GSEA), usando coleções de doenças e vias moleculares da *Kyoto Encyclopedia of Genes and Genomes* (KEGG), de modo a identificar vias de sinalização e processos biológicos alterados entre os diferentes grupos de doentes AA e caucasianas com CMTN, tendo em consideração a condição menopáusica e o estadio patológico da doença. Entre os grupos de genes enriquecidos, destacamos a via de sinalização Hippo como sendo a que está mais negativamente regulada nos seguintes grupos de doentes AA, comparando com doentes caucasianas: todos os CMTN, pós-menopausa, estadio II, e pós-menopausa e estadio II. Esta regulação negative potencia a tumorigénese em doentes AA, uma vez que a regulação negativa da via de sinalização Hippo promove a transcrição de genes envolvidos na proliferação e competição celular, inibição da morte celular, transição epitélio-mesenquimal e formação de metástases. Adicionalmente, grupos de genes das coleções comunidade celular e motilidade celular estão também negativamente regulados em doentes AA dos grupos anteriormente mencionados, sugerindo que a integridade dos contactos célula-célula e que a organização do citoesqueleto de actina, fundamentais na homeostase do tecido epitelial, estarão enfraquecidos, promovendo a dissociação de células tumorais e subsequente surgimento de metástases, em concordância com o pior prognóstico destas doentes. Além disso, uma vez que o grupo Cancro da Mama, que inclui genes envolvidos no processo tumoral na mama, estão consistentemente regulados positivamente em doentes AA nas mesmas comparações, também sugerimos que estas doentes têm uma atividade aumentada de componentes e processos envolvidos no desenvolvimento do CM, que podem explicar a progressão mais rápida e a uma maior agressividade do CMTN nestas doentes. Finalmente, destacamos que a via de sinalização MAPK, responsável pela regulação de uma panóplia de processos celulares, incluindo proliferação, diferenciação, apoptose e respostas a stress, tendo um papel fundamental na sobrevivência e desenvolvimento das células cancerígenas, está positivamente regulada especificamente nas doentes AA que apresentam a doença no estadio II, sugerindo que esta via poderá contribuir para a discrepância observada relativamente à sobrevivência das doentes das duas populações diagnosticadas neste estadio da doença.

De modo a identificar potenciais moléculas responsáveis pela discrepância associada à ancestralidade das doentes com CMTN, através do *Ingenuity Pathway Analysis* (IPA), realizámos uma análise com foco no design de redes moleculares, usando a lista completa de genes obtidos nas diferentes DGEA. As redes moleculares produzidas pelo IPA refletem um mapa de interações conhecidas entre as moléculas, onde redes moleculares altamente interconectadas são mais prováveis de representar alterações significativas nas funções biológicas e onde os reguladores centrais de cada rede molecular, i.e., genes diferencialmente expressos com mais ligações diretas e indiretas na sua rede, são potencialmente responsáveis pelas diferenças biológicas observadas entre as duas

populações. O design das redes moleculares do IPA destacou o papel central do EGFR, do Myc e do Bcl2 nas diferenças transcriptómicas associadas à ancestralidade. Estas descobertas despertaram o nosso interesse pelas seguintes razões. O EGFR, descrito como promotor da proliferação, motilidade e sobrevivência celular através da ativação de várias vias de sinalização, incluindo as vias MAPK e PI3K-AKT-mTOR, e sendo frequentemente descrito como estando superexpresso em vários tipos de cancro, incluindo o CMTN, está subexpresso nas doentes AA, quando comparando com doentes caucasiana homólogas. Este resultado está em linha com um estudo anterior, na qual a expressão de EGFR em CM e CMTN tem uma relação não-linear com o resultado da doença, sendo que uma expressão elevada ou diminuída do EGFR está associada a um pior resultado, comparando com os níveis intermédios de expressão. Assim, sugerimos que a baixa expressão de EGFR poderá promover um pior resultado especificamente nas doentes com CMTN de ascendência africana. O Myc, que está envolvido no crescimento celular, proliferação, metabolismo, diferenciação, vias de stress, resistência a fármacos e apoptose, e que é descrito como estando superexpresso em CMTN em estudos maioritariamente compostos por doentes de origem caucasiana ou linhas celulares, tem uma expressão génica ainda mais elevada nas doentes AA, o que poderá contribuir para uma pior evolução da doença neste grupo de doentes. Finalmente, o Bcl2, que é descrito como sendo um promotor da sobrevivência celular, está subexpresso especificamente nas doentes AA pré-menopáusicas. Em linha com estudos anteriores, a baixa expressão de Bcl2 em CMTN foi significativamente associada a uma proliferação mais elevada e a um aumento do risco de morte e recorrência. Notavelmente, também observámos uma subexpressão do *BCL2* em CMTN comparativamente ao tecido normal adjacente ao CM em doentes AA, sugerindo que esta alteração na transcrição do gene poderá contribuir para o desenvolvimento de CMTN em doentes pré-menopáusicas de ascendência africana.

Assim, selecionamos os candidatos EGFR, Myc e Bcl2 para a sua expressão ser validada em amostras de doentes com CMTN seguidas num hospital português, através de imunohistoquímica (IHC). Adicionalmente, também incluímos a β-Catenina nesta validação, sendo que anteriormente tínhamos observado que o seu gene se encontrava subexpresso em doentes AA pós-menopáusicas e diagnosticadas com a doença no estadio II. Decidimos incluir a β-Catenina devido ao número crescente de estudos sobre a via de sinalização Wnt/β-catenina como alvo terapêutico em vários tipos de tumores, bem como pelo facto de que a elevada expressão da β-catenina tem sido consensualmente descrita como sendo necessária para o desenvolvimento do CMTN através do controlo de várias propriedades tumorais, nomeadamente a migração, estaminalidade, crescimento celular e quimiossensibilidade.

Notavelmente, apesar de a DGEA ter sido realizada usando dados de sequenciação de RNA de doentes dos EUA, e de para a IHC termos usado amostras de doentes de um hospital português, a expressão diferencial génica e proteica entre doentes de ascendência africana e doentes caucasianas originaram resultados equivalentes. Uma vez que a população AA tem predominantemente ascendência *Niger-Kordofanian/*África Ocidental, e a sendo que as doentes com ascendência africana seguidas em Portugal têm origem em países da África Ocidental, os nossos resultados sugerem que pelo menos as doentes com ancestralidade na África Ocidental, além de terem uma incidência mais elevada de CMTN, como descrito noutros estudos, também partilham algumas características moleculares associadas ao CMTN distintas das das doentes caucasianas.

Assim, esta tese de doutoramento mostra uma nova perspetiva relativamente aos padrões de expressão génica do CMTN associados à ancestralidade de doentes de ascendência africana e caucasiana, particularmente quando considerando a condição menopáusica ou o estadio patológico

da doença no diagnóstico. Particularmente, observamos que a expressão diferencial génica e proteica do EGFR, Myc, Bcl2 e β-catenina corresponde em populações de doentes geograficamente e culturalmente distintas, sugerindo que estes marcadores são indicadores importantes do desenvolvimento do CMTN associado à ancestralidade das doentes. Assim, esperamos que estas descobertas venham a ter um impacto significativo no prognóstico e nas opções terapêuticas especificas para doentes com ascendência africana diagnosticadas com CMTN.

Palavras-chave: Cancro da mama triplo-negativo. Risco de doença associado à ancestralidade. Ascendência africana . Expressão génica diferencial . Dados de sequenciação de RNA

LIST OF ABBREVIATIONS

Abbreviation	Full form
AA	African-American
ADIPOQ	Adiponectin
AJ	Adherens Junctions
AJCC	American Joint Committee on Cancer
ALDHA1	Aldehyde Dehydrogenase 1 Family Member A1
ALK	Anaplastic lymphoma kinase
AKT1	AKT Serine/Threonine Kinase 1
APOBEC	Apolipoprotein B mRNA editing catalytic polypeptide-like
ATM	ATM Serine/Threonine Kinase
ATP	Adenosine triphosphate
BARD1	BRCA1 Associated RING Domain 1
BC	Breast cancer
BCL2	B-Cell CLL/Lymphoma 2 apoptosis regulator
BCSC	Breast cancer stem cells
BCV	Biological coefficients of variation
BHLH	Basic helix-loop-helix
BL1/2	Basal-like 1 or 2 TNBC subtype
BLIA	Basal-like immune-activated TNBC subtype
BLIS	Basal-like immune-suppressed TNBC subtype
BMI	Body Mass Index
BRCA1/2	Breast cancer DNA repair associated proteins
BRIP1	BRCA1 Interacting Protein C-Terminal Helicase 1
BSA	Bovine serum albumin

°C	Degree Celsius
CAV1	Caveolin 1
CCNE	G1/S-Specific Cyclin-E1
CDC	Center for Disease Control and Prevention
CDH1	Cadherin 1, Type 1, E-Cadherin (Epithelial)
CDK6	Cyclin Dependent Kinase 6
CHEK2	Checkpoint Kinase 2
CNA	Copy number aberration
COX2	Cyclooxygenase 2
CPM	Counts-per-million
CR	Cox-Reid profile-adjusted likelihood (method)
CRYBB2	Crystallin Beta B2
CSC	Cancer stem cells
CT	Chemotherapy
CTNNA1	α-Catenin
CTNNB1	β-Catenin
CXCR4	C-X-C Motif Chemokine Receptor 4
DAB	3,3'-Diaminobenzidine
DARC/ACKR1	Atypical Chemokine Receptor 1 (Duffy Blood Group)
DCIS	Ductal carcinoma in situ
ddH_2O	Double-distilled water
DEG	Differentially expressed gene
DFS	Disease-free survival
DGEA	Differential gene expression analysis
DSB	Double-strand breaks
DSS	Disease-specific survival

DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EMT	Epithelial-mesenchymal transition
EpCAM	Epithelial Cellular Adhesion Molecule
EPCR	Endothelial Protein C-Receptor
EPHA2	Ephrin Type-A Receptor 2
ER	Estrogen receptor
ERK1/2	Extracellular signal-regulated kinases 1 and 2
ESR1	Estrogen Receptor 1
EZH2	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit
FAK	Focal Adhesion Kinase
FANCA	Fanconi Anemia Complementation Group A
FC	Fold-change
FDR	False discovery rate
FFPE	Formalin-fixed paraffin-embed
FGFR1	Fibroblast Growth Factor Receptor 1
FOXA1	Forkhead Box Protein A1
FOXO3A	Forkhead Box Protein O3
GATA3	GATA Binding Protein 3
GDC	Genomic Data Commons
GJ	Gap Junctions
GLM	Generalized linear model
GSEA	Gene set enrichment analysis
GWAS	Genome-Wide Association Studies

h	Hour(s)
H_2O_2	Hydrogen peroxide
HER2	Human Epidermal Growth Factor Receptor 2
HFF	Hospital Professor Doutor Fernando Fonseca
HR	Hormone receptor
HRD	Homologous recombination deficiency
HR-MAS-NMR	High-resolution Magic angle spinning Nuclear magnetic resonance
HRP	Horseradish Peroxidase
HSP90	Heat shock protein 90 alpha family
HT	Hormone therapy
ICI	Immune checkpoint inhibitor
ICSBCS	International Center for the Study of Breast Cancer Subtypes
ID	Ensembl Gene identifier
IDC	Invasive ductal carcinoma
IGF1	Insulin Like Growth Factor 1
IGF1R	Insulin Like Growth Factor 1 Receptor
IHC	Immunohistochemistry
IL	Interleukin
ILC	Invasive lobular carcinoma
IM	Immunomodulatory TNBC subtype
INPP4B	Inositol Polyphosphate-4-Phosphatase Type II B
IPA	Ingenuity Pathway Analysis
IQR	Interquartile range
JAK2	Janus Kinase 2
JAK/STAT	Janus kinase/signal transducers and activators of transcription
JAMs	Junctional Adhesion Molecules

KEGG	Kyoto Encyclopedia of Genes and Genomes
KIFC1	Kinesin Family Member C1
KMT2C	Lysine Methyltransferase 2C
LAR	Luminal androgen receptor TNBC subtype
LCIS	Lobular carcinoma in situ
LOH	Loss of heterozygosity
LMP	Last menstrual period
M	Mesenchymal TNBC subtype according to Lehmann et al. 2011
mAb	Monoclonal antibody
MAPK	Mitogen-activated protein kinase
Mb	Megabase
MD	Mean-difference (plot)
MDM2	MDM2 Oncogene, E3 Ubiquitin Protein Ligase
MDS	Multi-dimensional scaling (plot)
MEK1	Mitogen-Activated Protein Kinase Kinase 1
MES	Mesenchymal TNBC subtype according to Burstein et al. 2015
MET	MET Proto-Oncogene, Receptor Tyrosine Kinase
MHC	Major histocompatibility complex
min	Minute(s)
mL	milliLitre(s)
mM	miliMolar
MLL3	Myeloid/Lymphoid Or Mixed-Lineage Leukemia Protein 3
MMR	Mismatch repair
mRNA	messengerRNA
MSL	Mesenchymal stem-like TNBC subtype
MTOR	Mechanistic Target Of Rapamycin Kinase

MYC	MYC Proto-Oncogene, BHLH transcription factor
NA	Not available
NACT	Neoadjuvant chemotherapy
NBN	Cell Cycle Regulatory Protein P95
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
NF1	Neurofibromin 1
NFKB	Nuclear Factor Kappa B
NGF	Nerve Growth Factor
NGS	Next-generation sequencing
NK	Natural killer
OGN	Osteoglycin
OR	Odds ratio
OS	Overall survival
PALB2	Partner And Localizer Of BRCA2
PAM	PI3K-Akt-mTOR pathway
PARP	Poly(ADP-Ribose) Polymerase
PARPi	PARP inhibitors
PBS	Phosphate Buffered Saline
pCR	Pathologic complete response
PD-1	Programmed cell death 1
PD-L1	Programmed death-ligand 1
PDGF	Platelet Derived Growth Factor
PFS	Progression-free survival
PGR	Progesterone Receptor
PI3K	Phosphatidylinositol 3-kinase

PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PLIN1	Perilipin 1
PLK	Polo Like Kinase
PR	Progesterone receptor
PTEN	Phosphatase And Tensin Homolog
QLF	Quasi-likelihood F (test)
RAD51C	DNA Repair Protein RAD51 Homolog 3
RB1	Retinoblastoma-Associated Protein
RGB	Red-Green-Blue color system
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing
RT	Room temperature
RTK	Receptor tyrosine kinase
SAP	Serviço de Anatomia Patológica
SEER	Surveillance Epidemiology and End Results
seq	Sequencing
SERD	Selective estrogen receptor downregulator
SERM	Selective estrogen receptor modulator
SNP	Single-nucleotide polymorphism
SPDEF	SAM Pointed Domain Containing ETS Transcription Factor
STAT3	Signal Transducer And Activator Of Transcription 3
STK11	Serine/Threonine Kinase 11
T2D	Type 2 diabetes
TCF4	Transcription Factor 4
TCGA	The Cancer Genome Atlas
TCV	Technical coefficients of variation

TDM1	Ado-Trastuzumab Emtansine
TGF-β	Transforming growth factor β
TH1/TH2	T helper type-1/2
THY1	Thy-1 Cell Surface Antigen
TILs	Tumor infiltrating lymphocytes
TJ	Tight junctions
TMM	Trimmed mean of M-values
TNBC	Triple-negative breast cancer
TNC	Tenascin C
TNF	Tumor Necrosis Factor
TP53	Tumor Protein P53
TREAT	T-tests relative to a threshold (method)
TWIST1	Twist Family BHLH Transcription Factor 1
UK	United Kingdom
UNS	Unstable TNBC subtype
USA	United Stated of America
VEGFR	Vascular Endothelial Growth Factor Receptor
W	White
WA	White-American
WHO	World Health Organization
WHR	Waist/hip ratio
XBP1	X-Box Binding Protein 1
ZEB1	Zinc Finger E-Box Binding Homeobox 1

LIST OF TABLES

LIST OF FIGURES

Chapter I

Introduction

1. CANCER

Cancer is one of the leading causes of death worldwide, being responsible for almost 10 million deaths in 2020. Globally, about 1 in 5 men or women develop the disease, and 1 in 8 men and 1 in 11 women die from it (Sung et al., 2021). Cancer or malignant tumors are broad terms used for a group of distinct diseases in which the accumulation of genetic mutations in cells causes neoplasia. Those abnormal cells are then able to invade adjacent tissue or to migrate and invade distant tissues. Tumorigenesis, the biological process by which normal cells are transformed into malignant cancer cells, which are further able to invade other tissues, has been the subject of a large research in the biomedical sciences for many decades (DeVita and Rosenberg, 2012; Douglas Hanahan and Robert A. Weinberg, 2000; Hanahan and Weinberg, 2011). This multistep process reflects the effect of genetic alterations that influence key cellular pathways that drive normal cells into highly malignant derivatives. Such genetic alterations may lead to the inactivation of tumor suppressor genes, genes involved in DNA repair, oncogene activation and/or alteration of epigenetic modulation, which are responsible for the neoplastic process (Szychot et al., 2013).

Proto-oncogenes transformation into oncogenes reflects a dominant mutational trait, which normally affects a single allele, and usually involves amplifications, missense mutations, chromosomal rearrangements and epigenetic event. Oncogenes become activated by structural alterations resulting from mutation or gene fusion, by juxtaposition to enhancer elements, or by amplification. Translocations and mutations can occur as initiating events or during tumor progression, while amplification usually occurs during progression. Proto-oncogenes, such as *EGFR*, *MYC*, *BCL2* or *CTNNB1* encode proteins that control cell proliferation and/or apoptosis usually leading to uncontrolled cellular growth when mutated (Carlo M. Croce, 2008; You and Jones, 2012).

In turn, tumor suppressor genes, such as *RB1* or *TP53*, normally require inactivation of both alleles to give rise to tumorigenesis. The loss of functional suppressor proteins deprives the cell of crucial brakes, which prevent inappropriate growth. According to Knudson (Knudson, 1971), most tumor suppressor genes are subject to recessive inactivation. In other words, the inactivation of a single allele is not enough for the gene to lose its tumor suppressive activity, being necessary the inactivation of the second allele (two-hit hypothesis of Knudson). In general, the recessive loss-of-function alterations affecting tumor suppressor genes, include large chromosomal deletions (Loss Of Heterozygosity – LOH) (Jones and Laird, 1999), missense, nonsense, base insertions and/or deletion mutations (Weinberg, 1996) and epigenetic events, such as DNA methylation (You and Jones, 2012).

Repair genes constitute a subset of the tumor-suppressor gene class and are responsible for DNA damage repair, being frequently inactivated in human cancers (Douglas Hanahan and Robert A. Weinberg, 2000; Hanahan and Weinberg, 2011). Due to their cellular function, repair genes, such as *BRCA1/2* or *PARP1*, constitute a unique subset of the tumor-suppressor genes because they have a more passive role in cell growth regulation. Instead, their inactivation in tumor cells results in an increased rate of mutations in other cellular genes, including proto-oncogenes and other tumor-suppressor genes (Ronen and Glickman, 2001).

Hanahan and Weinberg proposed that cancer cells can acquire up to ten hallmarks, or biological competences, during the multistep process of tumorigenesis (Figure 1) (Douglas Hanahan and Robert A. Weinberg, 2000; Hanahan and Weinberg, 2011). These 10 distinct and complementary capabilities allow the development of cancer and its dissemination by metastasis.

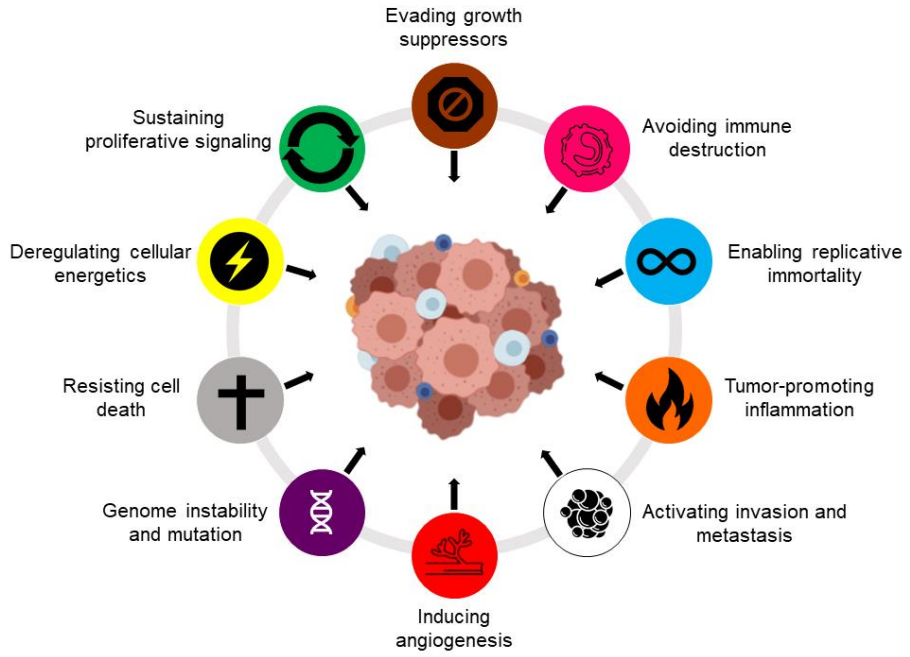


Figure 1 The Hallmarks of Cancer. Adapted from Hanahan & Weinberg, 2011. Created with BioRender.

Cancer cells may acquire capabilities such as self-sufficiency in growth signals, allowing independence from external proliferative stimuli; surpassing anti-proliferative cellular signals, in order to maintain the continuous growth; adjacent tissue colonization and metastasis; replicative immortality by upregulation of telomerase; sustained angiogenesis, with the capability of creating and sustain new blood vessels for oxygen and nutrients supply; resisting apoptosis, which is a major cellular response to neoplastic events (Douglas Hanahan and Robert A. Weinberg, 2000). Later, Hanahan and Weinberg postulated more hallmarks of cancer, namely deregulation of cellular energetic metabolism, mainly to support high glycolytic rates; and immune destruction avoidance. Two additional mechanisms have been proposed as the main inducers of the cancer hallmarks, namely genomic/chromosomal instability and mutation, which generates random mutations at high frequency; and tumor-promoting inflammation, allowing the recruitment of important molecules for the tumoral microenvironment, including growth, survival and proangiogenic factors, as well as extracellular matrix-modifying enzymes that facilitate angiogenesis, invasion, and metastasis, and inductive signals that lead to activation of epithelial-mesenchymal transition (EMT) and other hallmark-facilitating programs (Hanahan and Weinberg, 2011).

Thus, the tumorigenic process is caused by different insults that continuously act on replicative cells, leading to transformative alterations in their (epi)genetics, chromosomal numbers and arrangements, and heterotypic interactions. Then, these cells, in a path towards malignancy, undergo cycles of evolutionary clonal selection leading to the acquisition of cancer-competent traits, the hallmarks of cancer (Fouad and Aanei, 2017) (Figure 2).

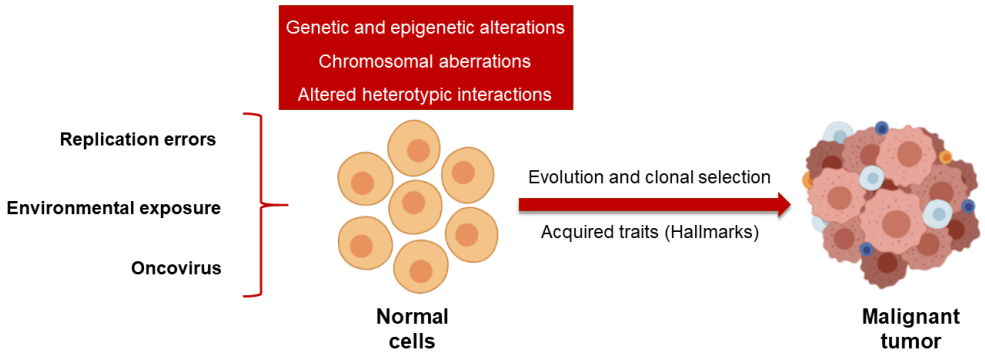


Figure 2 Tumorigenesis process. Adapted from Fouad & Aanei, 2017. Created with BioRender.

2. BREAST CANCER

2.1. Incidence and epidemiology

Female breast cancer (BC) was the most commonly diagnosed cancer worldwide in 2020, with an estimated 2.3 million new cases (11.7%), surpassing lung cancer (11.4) (Sung et al., 2021), and it was the fifth most deadly, with 684 996 deaths (6.9%) (Sung et al., 2021). In women, BC accounts for 24.5% of new cases of cancer and 15.5% of cancer deaths, the highest cause of cancer-specific death in females (Sung et al., 2021). In other words, among women, BC accounts for 1 in 4 cancer cases and for 1 in 6 cancer deaths, ranking first for incidence in the vast majority of countries (159 of 185 countries) and for mortality in 110 countries (Sung et al., 2021). Roughly 1% of BC occur in males, and 90% of these are related with the expression of proteins related with female hormones, such as the estrogen receptor (Korde et al., 2010).

Age of onset is an independent risk factor for BC, and cumulative evidence has suggested that BC in younger women shows differences in type, grade and aggressiveness, in comparison to the disease in older women. Although BC in women under 40 years are rare, representing less than 1% of all cases, tumors are frequently larger in size and of a higher grade (Anders et al., 2011; Peng et al., 2011). Nonetheless, the median age of BC diagnosis is 62 and around one quarter are women with more than 75 years, according to the Surveillance Epidemiology and End Results (SEER) registry (https://seer.cancer.gov/, USA, last consulted on February 6, 2021). As the population continues to age, 2015 projections estimate that invasive BC cases will double by 2030, with most of diagnosed cases in women with ages between 70-84, according to the Division of Cancer Epidemiology and Genetics from the National Cancer Institute (Rosenberg, 2015).

2.2. Breast cancer risk

Around the world, there are variations in BC incidence, mortality, and survival, which may be a result of several underlying complex factors, such as age, patient's ancestry, diet, and lifestyles, as well as reproductive issues (Bray et al., 2012). Increasing risk factors for BC development include exogenous hormones, such as contraceptive pills or postmenopausal hormone replacement therapy (Chlebowski et al., 2010; Collaborative Group on Hormonal Factors in Breast Cancer, 1997), reproductive factors,

namely earlier age of menarche, later age of menopause, nulliparity, late age of first child birth (Farhat et al., 2011), obesity and high insulin levels (Calle and Kaaks, 2004; Hvidtfeldt et al., 2012). Furthermore, BC is twice as common among first-degree relatives of BC patients than in women with no family history of the disease. However, the two most prevalent BC susceptibility genes, *BRCA1* and *BRCA2*, account for less than 20% of familial clustering of BC (King et al., 2003), demonstrating how heterogeneous BC can be. Interestingly, prognosis of *BRCA1/2* mutation carriers with BC appears similar to prognosis for sporadic BC patients (Daly et al., 2021; Goodwin et al., 2012). Thus, individuals with a personal or family history of hereditary breast cancer should be referred for genetic evaluation (Daly et al., 2021).

2.3. Biologic characteristics and pathology

2.3.1. Normal breast anatomy and histology

Female breasts (Figure 3) are mostly composed of glandular tissue, a specialized tissue responsible for milk production, and fatty tissue. The glandular tissue is organized into 15 to 20 sections, called lobes. Within each lobe are smaller structures, called lobules, where milk is produced. The milk travels through a network of ducts. The ducts connect and come together into larger ducts, which eventually exit the skin in the nipple. Connective tissue and ligaments provide support to the breast and give it its shape. Nerves provide sensation to the breast. The breast also contains blood vessels, lymph vessels, and lymph nodes (Henry et al., 2020).

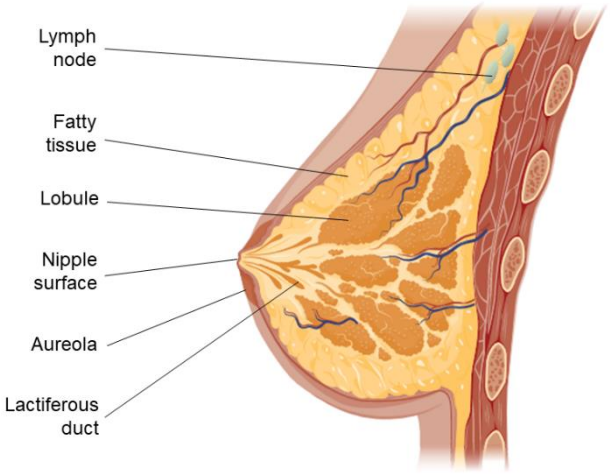


Figure 3 Normal breast anatomy. Created with BioRender.

2.3.2. Breast cancer classification

BC form a heterogeneous group of tumors, which are characterized by a wide spectrum of clinical, pathologic, and molecular features. This wide spectrum of factors accounts for variations in response to therapy and outcomes among patients. To aid in patients' prognosis and therapeutics, BC has different and complementary classifications. BC classifications consider 1) the location in the breast in which the tumorigenesis occurred; 2) hormone receptors and Human Epidermal Growth Factor Receptor 2 (HER2) expression level; and 3) the molecular profile of the cancer.

2.3.2.1. Breast cancer classification according to cancer invasiveness and location

Carcinomas, which are tumors that start in the epithelial cells, are the most common among BCs, being histologically heterogeneous. Breast carcinomas are usually called adenocarcinomas because they commonly arise from glandular cells in the ducts or in the lobules. Ductal carcinoma in situ (DCIS), also known as intraductal carcinoma, is a non-invasive or pre-invasive BC, located in the ducts and comprising about 80% of non-invasive BCs. On the other hand, lobular carcinoma in situ (LCIS) makes roughly 20% of non-invasive BCs (Henry et al., 2020).

The term invasive or infiltrating BC is used to describe any type of BC that trespassed the basement membrane, spreading into the surrounding breast tissue. The invasive ductal carcinoma (IDC) accounts for approximately 85% of BCs and, in contrast, invasive lobular carcinoma (ILC) accounts for roughly 10% of BCs (Sims et al., 2007). The overall prognosis of classic invasive ductal and lobular cancers is thought to be similar, although there may be some differences in likelihood of response to chemotherapy (CT) and endocrine therapy (Filho et al., 2015). Table 1 displays a summary of the BC histological subtypes according to the location of the cancer and its invasiveness capacity (Henry et al., 2020; Sims et al., 2007). Besides ductal and lobular invasive carcinomas, there are a number of less common types of breast carcinomas, including mucinous, tubular, and papillary, which are generally hormone receptor positive (Dieci et al., 2014).

Table 1 Breast cancer histological subtypes according to cancer invasiveness and location.

Histological subtypes	**Ductal**	**Lobular**
Pre-invasive cancer	Ductal carcinoma in situ (DCIS)	Lobular carcinoma in situ (LCIS)
- 25% - Cells limited to basement membrane	- 80% - May spread through ducts and impair duct structure - 1% progress to invasive cancer - Usually unilateral	- 20% - Does not impair duct structure - 1% progress to invasive cancer - Can be bilateral
Invasive cancer	Invasive ductal carcinoma (IDC)	Invasive lobular carcinoma (ILC)
- 75% - Extension beyond basement membrane	- 85% - Causes fibrous response, producing a palpable mass on examination - Metastasis through lymphatics and blood	- 10% - Minimal fibrous response, presents less often with palpable mass - Metastasis through abdominal viscera to gastrointestinal tract, ovaries and uterus - Almost always estrogen receptor positive

2.3.2.2. Breast cancer classification according to hormone receptor and HER2 expression

Estrogen receptor (ER), progesterone receptor (PR) and HER2 are considered the conventional BC biomarkers, being a major determinant of adjuvant and metastatic BC therapy (Henry et al., 2020). The expression of these markers is evaluated through immunohistochemistry (IHC). A tumor

is hormone receptor-positive if at least 1% of the cells tested have estrogen and/or progesterone receptors. Otherwise the test will say the tumor is hormone receptor negative.

Cancer-specific therapeutic targets such as ER and HER2 led to the successful development of therapies that are effective in the subset of patients whose tumors harbor these targets. Antiestrogen therapy in tumors that are ER and/or PR positive and anti-HER2 therapy in tumors with HER2 amplification are strategies that have demonstrated a survival benefit beyond that provided by cytotoxic therapy alone (Henry et al., 2020).

2.3.2.2.1. Estrogen and Progesterone receptor-positive breast cancers

ER and PR are the most widely studied markers in breast tissue. Estrogen plays a key role in the development of both normal breast epithelium and BC, and the modulation of estrogen concentrations and ER signaling are key therapeutic modalities for the majority of BCs. From 1995 to 2012, ER/PR+ BC comprised 62-76% of all BC cases diagnosed in the USA, with higher incidence in non-Hispanic White patients and lower incidence in non-Hispanic African-American patients (Desantis et al., 2016). When compared with hormone receptor negative tumors, hormone receptor positive BCs exhibit stronger clinical responses to hormonal treatment (Abe et al., 1998), better differentiated morphologic appearance (Henderson and Patek, 1998), and incidence rates that rise continuously with aging rather than slowing after menopause (Althuis et al., 2004).

The introduction of adjuvant systemic therapy led to a significant improvement in post-surgical survival and a reduction in disease relapse, especially in women with early BC and those with ER+ tumors, who may receive hormone therapy (HT) alone or in combination with cytotoxic therapy. Such hormone receptor-positive tumors are sensitive to targeted HT, which is the least toxic of the treatments available for BC (Aebi et al., 2011; Perez, 2007).

ERs are nuclear proteins that act as transcription factors regulating the expression of estrogen-responsive genes, and selective ER modulators (SERMs) are drugs that block signaling at the level of ERs. SERMs are estrogen agonists or antagonists depending on their interaction with ERs, target tissues, and post-translational effects (Osborne et al., 2000). They can be divided into three groups: 1) triphenylethylene nonsteroidal derivatives, such as tamoxifen, 2) other nonsteroidal compounds, such as Toremifene and Raloxifene, and 3) steroidal compounds with more complete antiestrogenic activity, currently named selective ER downregulators (SERDs), such as Fulvestrant (Abe et al., 1998; Lumachi et al., 2013; Perez, 2007).

On the other hand, PR expression is activated by ER, being expressed in approximately 50% of all BCs and in roughly 65% of ER+ cases. Survival analyses for BC patients given Tamoxifen therapy showed that PR+ cases had 24% higher relative probability for BC specific survival as compared to PR- patients (Liu et al., 2010).

2.3.2.2.2. HER2 breast cancers

The oncogene HER2 is overexpressed in approximately 20% of BCs as a result of amplification of the HER2/*neu* gene. HER2 protein overexpression or gene amplification is observed through IHC or in situ hybridization, respectively (Henry et al., 2020).

HER2 functions as a transmembrane receptor tyrosine kinase (RTK), activated on dimerization with another member of the epidermal growth factor family of receptors, including EGFR (ERBB1), ERBB3 (HER3), and ERBB4 (HER4). The dimerization domain of a partner receptor, such as EGFR or HER3, is exposed on binding of a ligand, such as heregulin, or through ligand-independent mechanisms (Baselga and Swain, 2009). The HER2-HER3 heterodimer is the most potent of these dimer pairs. This dimer pair is able to activate oncogenic intracellular signaling pathways, particularly phosphatidylinositol 3-kinase (PI3K)-Akt-mTOR pathway, which in turn promotes proliferation and cell survival (Gala and Chandarlapaty, 2014).

Clinically, BCs marked by HER2 amplification have a more aggressive outcome and decreased survival as compared with HER2-non amplified cases (Slamon et al., 1987). Nonetheless, nowadays, most BC cases with amplification of HER2 are treatable. Trastuzumab, a HER2-targeted humanized monoclonal antibody (mAb), started a new era in the treatment of invasive breast carcinomas with HER2 amplification by promoting the downregulation of HER2 dimerization and, consequently, the growth factor signaling cascades, promoting antibody-dependent cytotoxicity (Hudis, 2007). Following trastuzumab, lapatinib, pertuzumab and TDM1 were also developed as anti-HER2 therapy (Engebraaten et al., 2013).

2.3.2.2.3. Triple-negative breast cancers

Triple-negative breast cancer (TNBC) is a remarkably inter- and intra-heterogeneous BC subtype that neither expresses ER, PR, or has HER2 amplification. Therefore, TNBC lacks the conventional molecular targets and, as such is “negatively defined”. Due to the absence of specific treatments, women with TNBC have worse outcome compared to patients with other BC subtypes. Today, TNBC is considered one of the most relevant unmet medical problems in BC care (Masuda and Masuda, 2016; Schmadeka et al., 2014). Epidemiologically, TNBC is estimated to account for 10-20% of BC patients worldwide. However, there is evidence that TNBC prevalence greatly differs among populations (P Boyle, 2012).

Since TNBC is the BC subtype focused on this project, it will be described in more detail later (Section 3).

2.3.2.3. Breast cancer classification according to the molecular profile

Genome-wide RNA transcriptional profiling in combination with bioinformatic approaches have led to the development of molecular or intrinsic classification, which is based on gene profiles involving quantitative gene expression. This BC classification is often referred to as the "intrinsic" subtyping because BCs are defined by its intrinsic properties rather than by behavior. Array-based expression profiling studies allowed BC intrinsic classification according to six gene clusters: luminal subtype A, luminal subtype B, HER2-positive, basal-like and normal breast-like subtypes (Brenton et al., 2005; Perou et al., 2000; Sørlie et al., 2003, 2001), and later, the claudin-low subtype (Prat et al., 2010).

2.3.2.3.1. Luminal subtypes

The luminal BC subtypes make up the hormone receptor-expressing BCs and have expression patterns reminiscent of the luminal epithelial part of the breast. These patterns include expression of

luminal cytokeratins 8/18 and genes associated with ER (Perou et al., 2000). Within the luminal cluster there are two subtypes, luminal A and luminal B. Although both are hormone receptor expressing, the luminal subtypes have distinguishing features. Luminal A has, in general, higher expression of ER-related genes and lower expression of proliferative markers, such as the nuclear marker Ki-67 (Perou et al., 2000; Sørlie et al., 2003). Next-generation sequencing technology also showed that luminal A and B BCs, although having in common ER expression, have specific oncogenic drivers, rather than only having different proliferative capacity. Specifically, Luminal B BCs display increased expression of growth factor receptor genes, having higher grade and approximately 20% being HER2 positive (Ades et al., 2014).

Luminal A tumors are usually treated with endocrine therapy alone, while the more proliferative luminal B tumors may benefit from CT added to endocrine therapy (Sørlie et al., 2003).

2.3.2.3.2. HER2 subtype

Most tumors that are clinically HER2-positive, being observed through IHC or in situ hybridization, will fall within the HER2 molecular subtype. However, some Luminal B cases, although being ER+, also have HER2 overexpression or amplification (Sørlie et al., 2003, 2001).

The HER2 molecular subtype has a high proportion, 40% to 80%, of *TP53* mutations and is significantly more likely to be grade 3 than luminal A tumors, thus, carrying a poor prognosis (Carey et al., 2006; Sørlie et al., 2001). Despite its poor prognosis, HER2 molecular subtype has also demonstrated sensitivity to anthracycline and taxane-based neoadjuvant chemotherapy (NACT), with significantly higher pathologic complete response than luminal BCs (46% vs 7%, p = 0 .001) (Rouzier et al., 2005). Furthermore, HER2 molecular subtype has also molecularly targeted agents, such as the mAb Trastuzumab (Hudis, 2007). However not all BCs from the molecular subtype HER2 respond to Trastuzumab. *PTEN* loss or abrogation and *CXCR4* upregulation have been implicated in Trastuzumab resistance and may provide targets for combination strategies for even better approaches in the future (Brenton et al., 2005; Nagata et al., 2004).

2.3.2.3.3. Basal-like subtype

The basal-like molecular subtype of BC was so named due to the expression profile mimicking that of the basal epithelial cells of other parts of the body and normal breast myoepithelial cells. Basal-like BC is commonly known as TNBC because the majority of cases lack expression of ER and PR, have no overexpression and/or amplification of HER2, have strong expression of basal cytokeratins 5, 6, and 17, and also express proliferation related genes (Brenton et al., 2005; Prat et al., 2013). However, not all TNBCs are identified as basal-like by gene expression, and not all basal-like tumors are TNBCs (Prat et al., 2013), being that the discordance rate between the two definitions is 20%-30% (Cheang et al., 2015; Prat and Perou, 2011).

Following the molecular classification of BC, approximately 20% of all cases are basal-like. However, and as with TNBC, the basal-like subtype incidence has ancestry-associated discrepancies (Carey et al., 2006), which will be further discussed in Section 4.

IHC profiling using tissue microarray has identified that a group of tumors characterized by basal cytokeratin expression are also characterized by low expression of BRCA1 (Abd El-Rehim et al., 2005).

Indeed a basal phenotype is one of the hallmark features of "BRCA-ness", i.e., sporadic cancers that look like those from *BRCA1/2* mutation carriers, and might have important implications for management (Turner et al., 2004). Also, basal-like tumors are more likely to have aggressive features such as *TP53* mutations and a markedly higher likelihood of being grade III than luminal A BCs (Carey et al., 2006).

The poor prognosis experienced by patients with basal-like BCs is not from initial chemoresistance, but rather reflects the fewer treatment options available for ER-, PR-, and HER2-negative tumors and/or the intrinsic biology of this subtype. Nonetheless, basal-like response to anthracycline-based or combination anthracycline and taxane-based NACT is higher among basal-like BC than non-basal like (Carey et al., 2006; Rouzier et al., 2005).

Table 2 shows a summary of the invasive BC subtypes according to the intrinsic molecular subtyping (Section 2.3.2.3.).

Table 2 Breast cancer subtypes according to the intrinsic molecular subtyping.

Parameter	Luminal A	Luminal B HER2-	Luminal B HER2+	HER2	Basal-like
ER and/or PR	Positive	Positive	Positive	Negative	Negative
Ki-67	Low	High	Low or high	Low or high	Low or high
HER2	Negative	Negative	Amplified or overexpressed	Amplified or overexpressed	Negative

ER – Estrogen receptor
PR – Progesterone receptor
Ki-67 – nuclear marker of cell proliferation
HER2 – Human growth factor receptor 2

2.3.2.3.4. Normal-like breast cancer

Normal-like is an additional intrinsic BC subtype. Normal-like BCs show high expression of many genes known to be expressed by adipose tissue, which is the main component of a normal breast, and other nonepithelial cell types. Normal-like BC is similar to luminal A, being hormone-receptor positive and HER2 negative in most cases. However normal-like BC may have a higher Ki-67 expression and, usually, a worse prognosis than luminal A. These tumors may also show expression of basal epithelial genes and low expression of luminal epithelial genes (Dai et al., 2015). In a cohort of patients with no metastasis in the lymph-node (negative lymph-node), normal-like BCs comprised less than 8% of all cases, displaying lower to high grade disease and an intermediate outcome (Smid et al., 2008).

Overall, this is the most controversial molecular subtype of BC, in which the authors of the intrinsic molecular BC subtyping admit that is unclear whether these tumors represent poorly sampled tumor tissue or a distinct, clinically important group (Perou et al., 2000; Sørlie et al., 2001). Furthermore, Fan et al. (Fan et al., 2006) have suggested a 70-gene signature to classify BCs into 4 groups, including Luminal A and B, HER2 and basal-like, and where the normal-like subtype was not identified according to Sørlie's subtyping (Sørlie et al., 2003, 2001).

2.3.2.3.5. Claudin-low breast cancer

Lastly, additional gene-expression analyses later revealed the existence of another subtype, the claudin-low, present in 7% to 14% of all BC (Prat et al., 2010). Roughly 70% of claudin-low tumors are TNBC, with high representation of metaplastic and medullary BCs. Although claudin-low and basal-like subtypes share low luminal and HER2 gene expression, claudin-low tumors do not highly express proliferation markers. Specifically, claudin-low BCs are characterized by low levels of cell adhesion proteins and elevated expression of immune-related genes (e.g., *CD4* and *CD79a*). The mesenchymal features, including elevated expression of CD44, vimentin, and N-cadherin, and low epithelial differentiation, namely low *CD24* gene expression, resemble the mammary stem cell-like phenotype (CD44+CD24−/low) that can be acquired by EMT (Prat et al., 2010). In retrospectives studies, claudin-low BCs were associated with lower (39%) pathologic complete response (pCR) rates compared with basal-like subtype (73%), and worse prognosis than luminal-A tumors but similar survival as luminal-B, HER2-enriched, or basal-like tumors (Prat et al., 2010). Formation of cancer stem cells (CSCs) is induced by TGF-β in claudin-low cell lines (Bruna et al., 2012) and in CT-resistant TNBC TGF-β signaling and other stem cell markers are overexpressed (Bhola et al., 2013). Thus, inhibition of TGF-β signaling may represent a potential therapeutic strategy to help prevent the development of chemo-refractory disease, particularly in the claudin-low BC intrinsic subtype.

2.3.3. Breast cancer staging and grading

Accurate cancer staging is essential for therapeutic decision making and prognostic information. The American Joint Committee on Cancer (AJCC) staging TNM method is the most widely used method of BC staging (Giuliano et al., 2017). AJCC staging system is based on anatomic findings, namely primary tumor size (T), regional lymph nodes status (N), and metastases (M). Currently, in the specific case of BC staging, AJCC also includes the expression of biological biomarkers, such as ER, PR and HER2, due to their recognized impact on patients' treatment and outcome. Table 3 shows a generalized overview of the TNM method applied in BC according to AJCC (Giuliano et al., 2017).

Table 3 Generalized overview of breast cancer TNM staging system according to the 8th edition of AJCC (Giuliano et al., 2017). Each categorical column indicates the anatomical valuation criteria by row for advancing cancer stage, thus reflecting tumor progression.

T **Tumor size**	**N** **Lymph node**	**M** **Metastasis**
TX Primary tumor cannot be assessed	NX Regional lymph nodes cannot be assessed	M0 No evidence of distant metastases
T0 No evidence of primary tumor	N0 No regional lymph node metastasis identified	M1 Distant metastases detected
T1 Tumor ≤ 20 mm in greatest dimension	N1 Tumor spread to closest or small number (1-3) of regional lymph nodes	
T2 Tumor > 20mm but ≤ 50 mm in greatest dimension	N2 Tumor spread to an extend between N1 and N3	
T3 Tumor > 50mm in greatest dimension	N3 Tumor spread to more distant or numerous (> 10) lymph nodes	
T4 Tumor of any size that invades to other organs		

TNM method is often translated in clinical/pathological stages. This happens not only to simplify the diagnosis information given to the patient, according to the severity of the disease, but also when T, and/or N, and/or M have not been explicitly recorded in the medical records.

Staging can be “clinical” or “pathological”. Clinical staging is based on the results of tests done before surgery, such as physical examinations and imaging scans. Thus, clinical staging contributes for the initial diagnosis of the patient. On the other hand, pathological staging is based on what is found during surgery. Clinical stage is often indicated with a lowercase “c” before the TNM classification and the pathological stage is indicated with a lowercase “p” (Giuliano et al., 2017). Clinical and pathological staging may be different in the cases of a major cancer progression between diagnosis and surgery or when the patient is subjected to neoadjuvant treatments, which usually involve one or more CT medicines, before surgery. These neoadjuvant treatments are performed before surgery to help shrinking the tumor so it can be more easily removed. Furthermore, staging may be also evaluated after patient has received adjuvant treatments, before the surgery. This so-called post-therapy stage is indicated with a lowercase “y” before the TNM classification.

Table 4 illustrates the TNM method converted into clinical/pathological stages.

Table 4 American Joint Commission on Cancer TNM stage groups.

When T is…	And N is…	And M is…	The clinical stage is…
Tis	N0	M0	0
T1	N0	M0	IA
T0	N1mi	M0	IB
T1	N1mi	M0	IB
T0	N1	M0	IIA
T1	N1	M0	IIA
T2	N0	M0	IIA
T2	N1	M0	IIB
T3	N0	M0	IIB
T1	N2	M0	IIIA
T2	N2	M0	IIIA
T3	N1	M0	IIIA
T3	N2	M0	IIIA
T4	N0	M0	IIIB
T4	N1	M0	IIIB
T4	N2	M0	IIIB
Any T	N3	M0	IIIC
Any T	Any N	M1	IV

Tis – Ductal carcinoma in situ (DCIS)
N1mi – Micrometastases in the lymph nodes (approximately 200 cells, larger than 0.2 mm, but none larger than 2.0 mm)

Besides TNM method and staging, the histologic grading of a cancer, which defines morphologic information of the metastatic potential of the tumor, also contributes to patient's diagnosis. Essentially, in a grade I (G1) cancer, cells are well differentiated, being morphologically similar to normal cells and not growing rapidly; in a grade II (G2) cancer, cells are less differentiated, being morphologically different than normal cells, and with a faster proliferation and; in a grade III (G3) cancer, cells are poorly differentiated and proliferation markers are highly expressed. Frequently, a higher morphological grading coincides with a higher staging (Ivshina et al., 2006).

Together with biological factors, including hormone receptor status and HER2 expression, staging and grading are essential for BC patient's diagnosis and treatment decision. Nowadays, we are able to identify groups of BC patients and to adjust specific therapeutic modalities. Thus, even patients with a more advanced staging or grading may have a good prognosis and skip systemic CT, which may trigger toxic side effects. Nonetheless, as shown in Figure 4, the targetable BC subtypes, such as the hormone receptor positive or HER2 overexpression cases, are also the ones frequently found with lower grading and best prognosis. Particularly, TNBC has higher grade and should be considered a priority in the BC translational research. In the next section, TNBC is covered in more detail.

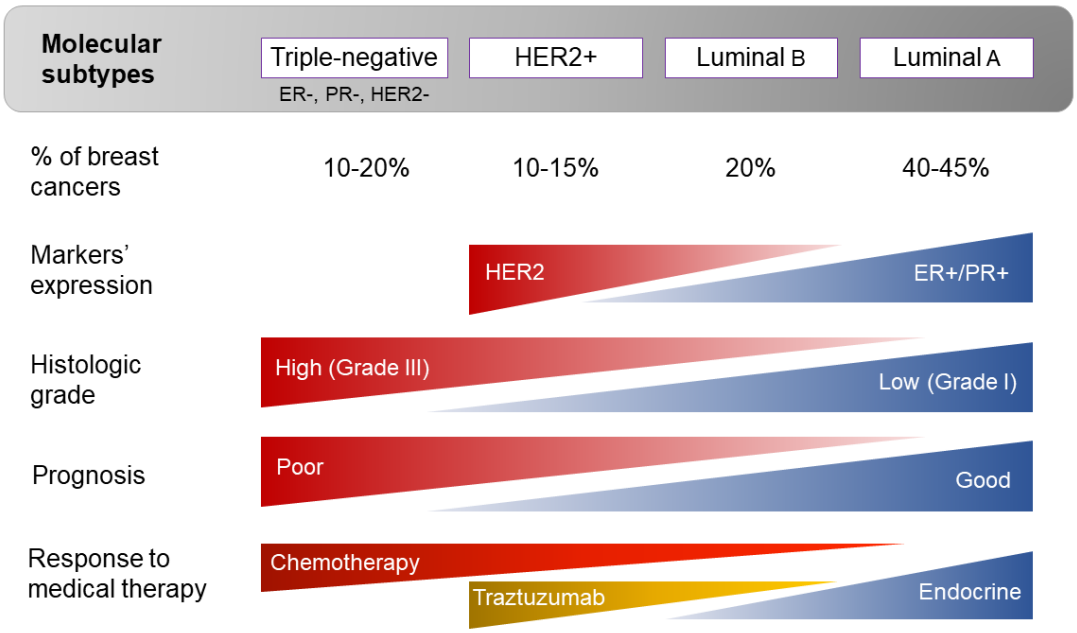


Figure 4 Summary of breast cancer molecular subtypes according to its incidence, expression of conventional biomarkers, grading, prognosis and response to treatment.

3. TRIPLE-NEGATIVE BREAST CANCER

As aforementioned in section 2.3.2.2.3., TNBC denomination is due to the fact that these cancers do not express the conventional BC therapeutic targets ER, PR and HER2. The definition of negative ER and PR status by IHC is not concordant in the literature, with some definitions considering ER expression to be significant only if at least 10% of tumor cells express the receptors. However, the St. Gallen guidelines (Gnant et al., 2015), the American Society of Clinical Oncology (Denduluri et al., 2016), and the American College of Pathology (WOLFF et al., 2007) have defined TNBC as the BC with less than 1% of tumor cells expressing ER and PR via IHC. The lack of the conventional BC targets is one of the reasons for TNBC being the most challenging BC subtype to treat. To date, therapies directed to specific molecular targets are slowly achieving clinically meaningful improvements in outcomes of patients with TNBC (Kang and Syed, 2020; Schmid et al., 2020), but systemic CT remains the standard of care (Garrido-Castro et al., 2019; Saraiva et al., 2017).

TNBC has an aggressive behavior with presentation of de novo metastatic BC, large locally advanced breast lesion or metastatic disease developing shortly after adjuvant CT (Collett et al., 2005; Seewaldt and Scott, 2007). TNBC frequently metastasizes to the viscera, liver, lung or brain (Saraiva et al., 2017). Furthermore, TNBC may also present an oligometastatic phenotype closer to ER+ BC, with only lymph node and bone disease (Wei et al., 2008).

Considering the age at diagnosis, there is increasing evidence that TNBC has a bimodal distribution with the first incidence peak in premenopausal patients and a second peak after 70 years of age (Saraiva et al., 2017). Prognosis of stage-matched premenopausal TNBC is worse than older age TNBC. One can speculate on the underlying biology that explains this difference in outcome. Premenopausal TNBC would be a disease with a few very powerful molecular drivers, more akin to

single hit neoplasms, while geriatric TNBC would be a disease of generalized chromosomal instability, a hallmark of ageing tissues and of geriatric cancer (Curtin et al., 2005; Saraiva et al., 2017).

The molecular mechanisms that drive TNBC recurrence have not been fully elucidated. TNBC has a higher rate of recurrence in the first 5 years and a lower rate of recurrence afterwards. Nevertheless, there appears to be two distinct recurrence peaks. On one hand, although many patients with early stages of TNBC are cured with CT, in those who develop metastatic disease, median overall survival (OS) with current treatment options is 13 to 18 months (André and Zielinski, 2012). Thus, the peak risk of relapse is at the third year and in this scenario the clinical outcome worsens. Also, survival after metastasis relapse is reduced when compared with other BC subtypes (Kumar and Aggarwal, 2016). On the other hand, the pattern of late recurrence is generally associated with less aggressive disease, frequently with bone metastases (Dent et al., 2007; Saraiva et al., 2017; Wei et al., 2008). The apparent different paths of tumor progression in TNBC might be driven by confounding aspects. These differences may be either just a proxy for age of incidence (Tse et al., 2009), or patients that have late recurrence and good prognosis may represent cases that are false negatives for ER (Gong et al., 2007).

Briefly, in addition to the fact of not having the conventional therapeutic targets, TNBC has a highly aggressive clinical course, with earlier age of onset, higher grade, greater metastatic potential, and poorer clinical outcomes, as shown by the higher relapse and lower survival rates (Dent et al., 2007). Furthermore, TNBC has a higher incidence in patients from African ancestry. Such issue will be later discussed in this chapter (Section 4). TNBC features are summarised in Table 5.

Table 5 Clinical, epidemiological and therapeutic features of TNBC. Adapted from (Saraiva et al., 2017).

Characteristics	Worse outcome	Better outcome
Age of presentation	Young	Old
Stage at presentation	Advanced	Early
Growth rate	Fast	Slow
First site of metastasis	Liver and brain	Lymph nodes and bone
Chemotherapy response	Resistant	Sensitive
Body mass index	High	Low
Patients' ancestry	African-ancestry	Caucasian/White

3.1. Molecular heterogeneity of TNBC

Patients with TNBC experience a wide range of treatment outcomes, from a rapid recovery with minimal therapy to a highly resistant and recurrent disease. This diversity is further highlighted by the high prevalence of rare histopathological subtypes, such as metaplastic (in which 90% are TNBCs), medullary (95%), adenoid cystic (90-100%), and apocrine (40-60%) carcinomas (Rakha et al., 2007). Although being a single diagnosis, TNBC response rates and presentation greatly differ among patients.

Major effort has been devoted over the past decade to classify TNBC into distinct clinical and molecular subtypes that could guide treatment decisions. Characterization of genomic, transcriptomic, proteomic, epigenomic, and microenvironmental alterations has expanded our knowledge of TNBC (Garrido-Castro et al., 2019). The Cancer Genome Atlas (TCGA) Research Network analyzed primary BCs using different platforms, including genomic DNA copy number arrays, DNA methylation, messenger RNA arrays, reverse phase protein arrays and exome, and RNA and microRNA sequencing

(Koboldt et al., 2012). By integrating information across platforms, the authors were able to examine the molecular heterogeneity of the tumors. This TCGA analysis revealed that the most frequent loss-of-function and gain-of-function alterations in TNBC involve genes associated with DNA damage repair and phosphatidylinositol 3-kinase (PI3K) signaling pathways, respectively. Alterations in DNA damage repair genes include loss of *TP53, RB1*, and *BRCA1* function. Aberrant activation of the PI3K pathway occurs due to loss of negative regulators such as the lipid phosphatases *PTEN* or *INPP4B* (Andre et al., 2009; Gewinner et al., 2009) or activating mutations in *PIK3CA*, along with other genes in the PI3K/TOR signaling network (Abramson et al., 2015; Saal et al., 2005). Another study sequenced and analyzed 104 TNBC tumors and confirmed the high rate of *TP53* mutations; however, the study showed that 12% of cases did not have somatic mutations in any established "driver" gene, suggesting that primary TNBCs are indeed mutationally heterogeneous from the outset (Shah et al., 2012).

All things considered, is now becoming well appreciated that TNBC, which continues to be defined by the biomarkers it lacks, is not one disease but a constellation of molecularly, morphologically, and behaviorally diverse entities. The next sections highlight some of the efforts performed in the last decade in dissecting TNBC complexity.

3.1.1. Somatic genetic alterations in TNBC

Cancers harbor numerous somatic genetic alterations, though only a small proportion of them confer clear fitness advantage, also known as "cancer drivers" (Vogelstein et al., 2013). Large-scale exome and targeted sequencing studies in primary BC have revealed the presence of many alterations in putative cancer-driver genes in TNBC (Koboldt et al., 2012; Pereira et al., 2016; Weisman et al., 2016). The average mutation rate in basal-like BC is among the highest in breast tumors, with 1.68 mutations per megabase (Mb) (Koboldt et al., 2012). Different genomic classifications in BC have been proposed by grouping next-generation sequencing (NGS)-detected alterations in known cancer-driver genes according to the intracellular pathways in which they are involved, such as PI3K/AKT and RAS/MAPK signaling, DNA damage repair, and cell-cycle or transcriptional regulation (Balko et al., 2014; Koboldt et al., 2012; Pereira et al., 2016). Most somatic mutations in TNBC occur in tumor suppressor genes, such as *TP53*, *RB1*, and *PTEN*, which have not been successfully targeted therapeutically to date. Although less prevalent, oncogenic alterations in the PI3K/AKT pathway have also been described in basal-like BCs, being *PIK3CA* mutated in 7%, *AKT3* amplification in 28%, and *PTEN* mutation or loss, in 35% of all cases (Koboldt et al., 2012). Consistent with findings in untreated TNBCs, targeted sequencing of residual disease post-NACT showed that >90% of patients had at least one altered pathway (Balko et al., 2014). However, only the following three alterations were found to be significantly prognostic for OS: *JAK2* amplification and *BRCA1* truncation or mutation predicted poor OS, and *PTEN* alteration predicted better OS (Balko et al., 2014).

Furthermore, knowing that TNBCs are dominated by copy number aberrations (CNAs) with in *cis* or in *trans* associated gene expression changes (Ciriello et al., 2013; Curtis et al., 2012a), Patel and colleagues integrated CNA and gene expression-driven gene dependency identification and functional validation approach in order to identify novel malignant cell selective addictions and potential targetable genes or pathways in TNBC. Among the 37 identified genes, biological processes, and clusters of interacting proteins, that merit further investigation, authors highlight the newly mechanistically validated *KIFC1*, a potentially druggable mitotic kinesin, as a highly selective malignant cell target, with mechanistic evidence of synergy in combination with cisplatin (Patel et al., 2018).

3.1.2. BRCA and "BRCA-ness" in TNBC

Cancers that lack functional *BRCA1* or *BRCA2* have a deficiency in HR repair of DNA double-strand breaks (DSB), leading to dependence on alternative mechanisms to repair these lesions, and genomic instability (Moynahan et al., 1999; Tutt et al., 2001; Yoshida and Miki, 2004). Drugs that generate DSBs, such as alkylating agents, namely platinum or mitomycin C, or PARP inhibitors, cause persistent DNA damage in HR-deficient cells and, consequently, induction of cell-cycle arrest and apoptosis (Bhattacharyya et al., 2000; Farmer et al., 2005).

Germline mutations in *BRCA1* or *BRCA2* (BRCA1/2) are present in approximately 10% of patients with TNBC (Hartman et al., 2012). Furthermore, somatic mutations and epigenetic alterations that inactivate BRCA1/2 and other DNA-repair genes have also been identified in TNBC (Foulkes, 2003; Turner et al., 2004). Given that HR deficiency exposes specific therapeutic vulnerabilities, the detection of tumors with this so-called "BRCA-ness" phenotype has clinical implications. Most *BRCA1*-related TNBCs are basal-like (Foulkes, 2003), and there is a marked resemblance in phenotype and biology between sporadic basal-like tumors and *BRCA1*-associated cancers (Turner et al., 2004).

Despite these similarities, targeting the HR pathway in basal-like BC and TNBC has revealed conflicting data in the metastatic and neoadjuvant settings, with reports of resistance to DNA-damaging drugs, such as platinum or alaprib, which may be due to epigenetic changes such as loss of *BRCA1* promoter hypermethylation via *BRCA1* locus fusion rearrangements, with subsequent *BRCA1* reexpression (Ter Brugge et al., 2016).

Several strategies to exploit potential synthetic lethality in HR-deficient tumors are being explored across solid tumors, including clinical trials combining PARP inhibitors with PI3K/AKT inhibitors, immune-checkpoint inhibition, and HSP90 inhibitors (Garrido-Castro et al., 2019).

3.1.3. TNBC and intrinsic breast cancer subtypes

All intrinsic BC subtypes (Section 1.3.2.3) can be found within IHC/in situ hybridization defined TNBC disease, however, basal-like tumors exhibit the greatest overlap with TNBC. Between 50% and 75% of TNBC have basal phenotype, and approximately 80% of basal-like tumors are ER-/PR- and HER2- (Prat et al., 2013).

Characterization of intrinsic subtypes using a 50-gene assay, established as the PAM50 BC subtype predictor (Prat et al., 2012), developed by the same team as the intrinsic BC subtypes (Prat et al., 2010; Sørlie et al., 2003, 2001), has provided independent predictive information of pathologic complete response (pCR) to neoadjuvant therapy across all BC subtypes. However, when restricting analyses to TNBC, none of the PAM50 signatures at the time of diagnosis have significantly correlated with pCR (Prat et al., 2014). Specifically, in basal-like TNBC, high expression of cell cycle-related genes (e.g., *CCNE* and *FANCA*) and low levels of estrogen signaling-related genes (e.g., *FOXA1* and *PGR*) were associated with pCR, while high expression of EMT genes (e.g., *TWIST1* and *ZEB1*) was significantly enriched in residual disease (Prat et al., 2014).

Although not being part of the PAM50 predictor, most claudin-low tumors are TNBC, comprising approximately 70% of all claudin-low cases. However, and contrary to basal-like TNBCs, claudin-low TNBCs are not highly proliferative. Claudin-low TNBCs are rich in quiescent cells, high expression of mesenchymal markers, such as CD44, vimentin, and N-cadherin, and having lower expression of CD24

(Prat et al., 2010). Specifically, claudin-low TNBCs highly express the phenotype of BC stem cells, namely CD44+CD24−/low (Fillmore and Kuperwasser, 2007). CD44+CD24−/low cells have been associated with an enhanced capacity for detaching from the primary tumor and metastasizing, supported by a dominant EMT gene-expression profile (Jiagge et al., 2018).

3.1.4. TNBC molecular subtyping

With evolving transcriptomic studies, the heterogeneity of TNBC has been further dissected. To better understand the molecular underpinnings of TNBC, Lehmann et al. first compiled an extensive number of TNBC gene expression profiles from different BC datasets and initiated the molecular subtyping race of TNBC (Lehmann et al., 2011). This study revealed that TNBC subtypes are characterized by distinct patterns of molecular alterations in terms of RNA expression, somatic mutations, and copy number variations, that tend to cluster in genes implicated in specific pathways. Thus, each subtype has a unique biology that responds differentially to current therapies. With a 2188-gene expression algorithm, Lehmann et al. reported the following 6 TNBC subtypes: two basal-like TNBC subtypes, one with cell cycle and DNA damage response gene expression profile (BL1), and the other enriched in growth factor signaling and myoepithelial markers (BL2); two mesenchymal subtypes with high expression of genes involved in differentiation and growth factor pathways (M and MSL); an immunomodulatory (IM) type; and a luminal subtype driven by androgen signaling, called luminal androgen receptor (LAR), mostly composed of molecular apocrine TNBC cases. Furthermore, an unstable subtype (UNS) was also identified (Lehmann et al., 2011).

To validate these results, differential gene expression was analyzed to identify TNBC cell line models representative of each subtype, and the predicted “driver” signaling pathways were pharmacologically targeted in these preclinical models as proof of concept that the analysis of distinct gene expression signatures can inform therapy selection. Interestingly, cell line models representing each of the TNBC subtypes also displayed different sensitivities to targeted therapeutic agents (Lehmann et al., 2011).

With this six-subtype classification, Lehmann’s group developed the “TNBCtype”, a web-based subtyping tool for TNBC tumor specimens using gene enrichment metadata and classification methods (Chen et al., 2012). Importantly, evidence supportive of the clinical utility of Lehmann’s classification has already been demonstrated, based on the ability of the algorithm to predict differential responsiveness to the current standard of care, namely taxane- and anthracycline-based CT, for TNBC cases (Masuda and Masuda, 2016).

However, Lehmann’s TNBC classification (Lehmann et al., 2011) had two main drawbacks. One of them was regarding the unstable subtype, not accounted in the TNBCtype tool. Additionally, since IHC is the clinical standard procedure used to define TNBC, through detection of ER, PR and HER2, in the study by Lehmann et al., when IHC-confirmed TNBCs were analyzed, only 5 clustered subtypes were observed (Burstein et al., 2015; Lehmann et al., 2011). Therefore, while this study greatly advanced the understanding of TNBCs, stable subtypes as well as subtype-specific molecular targets, that can preferably be identified through IHC in the clinical context, were still lacking.

Following Lehmann's TNBC subtyping (Lehmann et al., 2011), Burstein and colleagues, through RNA and DNA profiling, developed an 80-gene signature for TNBC classification. Authors were able to distinguish the following 4 stable molecularly defined TNBC subtypes: LAR, mesenchymal (MES), basal-like immune-suppressed (BLIS), and basal-like immune-activated (BLIA) (Burstein et al., 2015). MES, BLIS, and BLIA are characterized by distinct clinical prognosis, with BLIS tumors having the worst and BLIA tumors having the best outcome. These results also demonstrated subtype-specific gene amplification and gene expression, suggesting the possibility of using in situ hybridization techniques to identify these TNBC subsets. Essentially, Burstein's LAR and M subtypes overlap with Lehmann's TNBCtype LAR and MES subtypes, respectively, whereas BLIS and BLIA contained mixtures of the other 4 Lehmann subgroups (Burstein et al., 2015). Moreover, BLIS and BLIA completely overlap with PAM50's basal-like subtype, while M subtype is composed of PAM50's basal-like and normal-like subtypes and the LAR subtype may be of all PAM50 subtypes (Burstein et al., 2015). These observations are summarized in Figure 5.

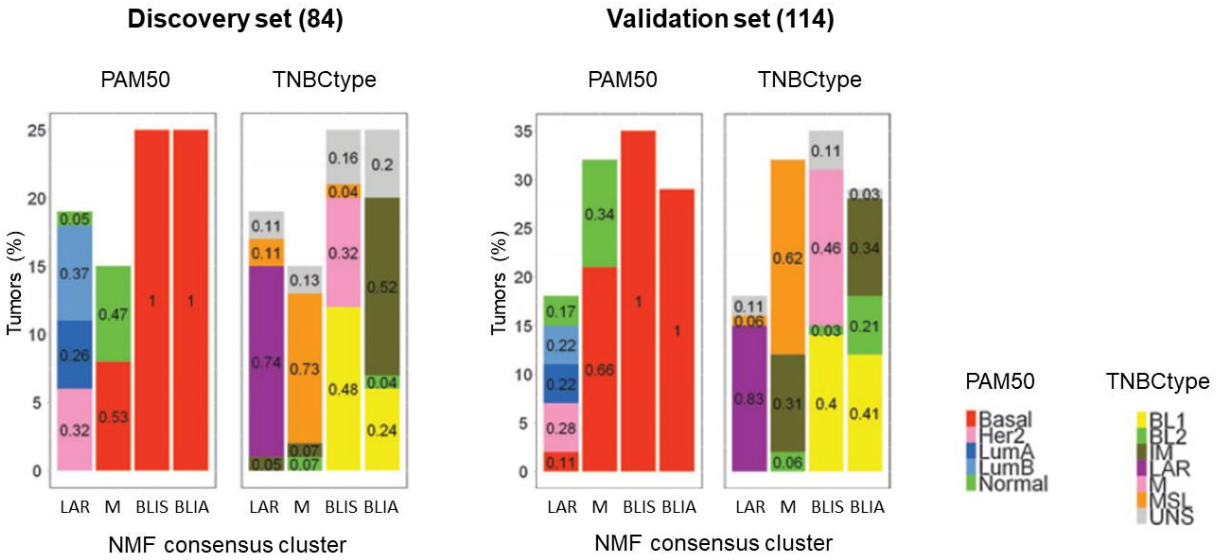


Figure 5 TNBC classification according to Burstein *et al.* (2015) with 4 stable molecular phenotypes, LAR, M, BLIS and BLIA, comparing with PAM50 (Prat et al., 2012) and TNBCtype (Lehmann et al., 2011) subtyping. The discovery set had 84 TNBC patients and the validation set had 114 TNBC patients. *NMF* Non-negative matrix factorization was the method used for TNBC clustering. *LAR* Luminal-androgen receptor. *M* Mesenchymal. *BLIS* Basal-like immune-suppressed. *BLIA* Basal-like immune activated. *LumA* Luminal A. *LumB* Luminal B. *BL1* Basal-like 1. *BL2* Basal-like 2. *IM* Immunomodulatory. *MSL* Mesenchymal stem-like. *UNS* Unstable. Adapted from Burstein et al., 2015.

Later, Lehman et al. refined their six TNBC molecular subtypes, TNBCtype subtyping, into the following four (TNBCtype-4) tumor-specific subtypes: basal-like 1 (BL1), basal-like 2 (BL2), mesenchymal (M) and LAR (Lehmann et al., 2016). With this study, their objective was to ascertain the complexity of the varying histological landscape of tumor specimens. This time, through histopathological quantification and laser-capture microdissection, authors observed that transcripts in the previously described IM and MSL subtypes were enriched with infiltrating lymphocytes and tumor-associated stromal cells, respectively. Therefore, the new classification takes into account the contribution of transcripts from normal stromal and immune cells in the tumor environment (Lehmann et al., 2016). Figure 6 shows the molecular subtype distribution of 767 TNBC samples stratified by PAM50, TNBCtype or refined TNBCtype-4. TNBCtype-4 significantly differ from TNBCtype not only in

prognosis and response to CT, but also in initial presentation and patterns of recurrence, where regional nodal involvement is more common in LAR TNBC and metastatic recurrences have tropism to the lung in M subtypes and to the bone in LAR subtypes. Similar to the TNBCtype, response to NACT, such as platinum- and taxane-based regimens, is significantly associated with TNBCtype-4 subtypes ($p = 0.027$), with the highest and lowest pCR rates reported in BL1 (65.6%) and LAR (21.4%), respectively (Echavarria et al., 2018).

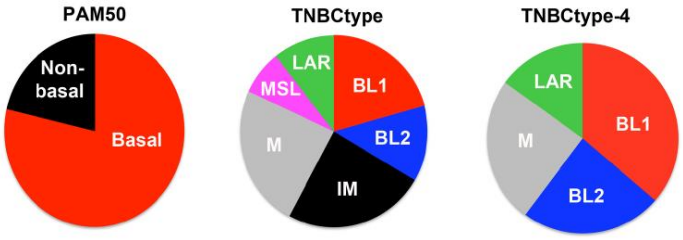


Figure 6 Molecular subtype distribution of TNBC samples stratified by PAM50 (Prat et al., 2012), TNBCtype (Lehmann et al., 2011) or refined TNBCtype-4 (Lehmann et al., 2016). Adapted from Lehmann et al., 2016.

Afterwards, Ring et al. (2016) designed an expression algorithm which reduced Lehmann's 2188-gene expression gene collection into 101 genes, equally able to subtype TNBCs similarly to the original TNBCtype, as well as to predict patient's outcomes (Ring et al., 2016). This newly refined assay also displayed the ability to identify more than one subtype – called a dual subtype, with a primary and potential secondary subtype – in any given patient, which should better reflect tumor heterogeneity. The assay also separately classifies each patient as either IM-negative or IM-positive to provide possible insight into the tumor immune-microenvironment (Ring et al., 2016).

In 2019, Jézéquel et al., with the aim of defining robust TNBC subtypes with clinical relevance, used 54 gene-expression signature DNA chips to perform gene expression profiling in a cohort composed of 238 patients (Jézéquel et al., 2019). In addition, external data (n = 257), obtained by using the same DNA chip, was used for validation. Authors identified three molecular clusters within TNBC: one molecular apocrine (C1) and two basal-like enriched (C2 and C3). C1 presents a luminal phenotype, with AR receptor. C2 presents a more pronounced mesenchymal phenotype, has a pro-tumorigenic immune response (i.e., immune suppressive profile and high neurogenesis, with nerve infiltration), and high biological aggressiveness. In contrast, C3 exhibited adaptive immune response associated with complete B cell differentiation that occurs in tertiary lymphoid structures, and immune checkpoint upregulation (Jézéquel et al., 2019). As shown in Figure 7, C2 and C3 overlap with PAM50's basal-like subtype, while C1 is a mixture of the other PAM50 subtypes. Furthermore, C1 is indeed most comprised by TNBCtype and Burstein's 4-TNBC clustering LAR subtype, as well as TNBCtype's MSL and

Burstein's M subtypes. C2 comprises TNBCtype's BL1 and M subtypes and both Burstein's BLIS and BLIA. Finally, C3 mostly overlaps with TNBCtype's IM and Burstein's BLIA.

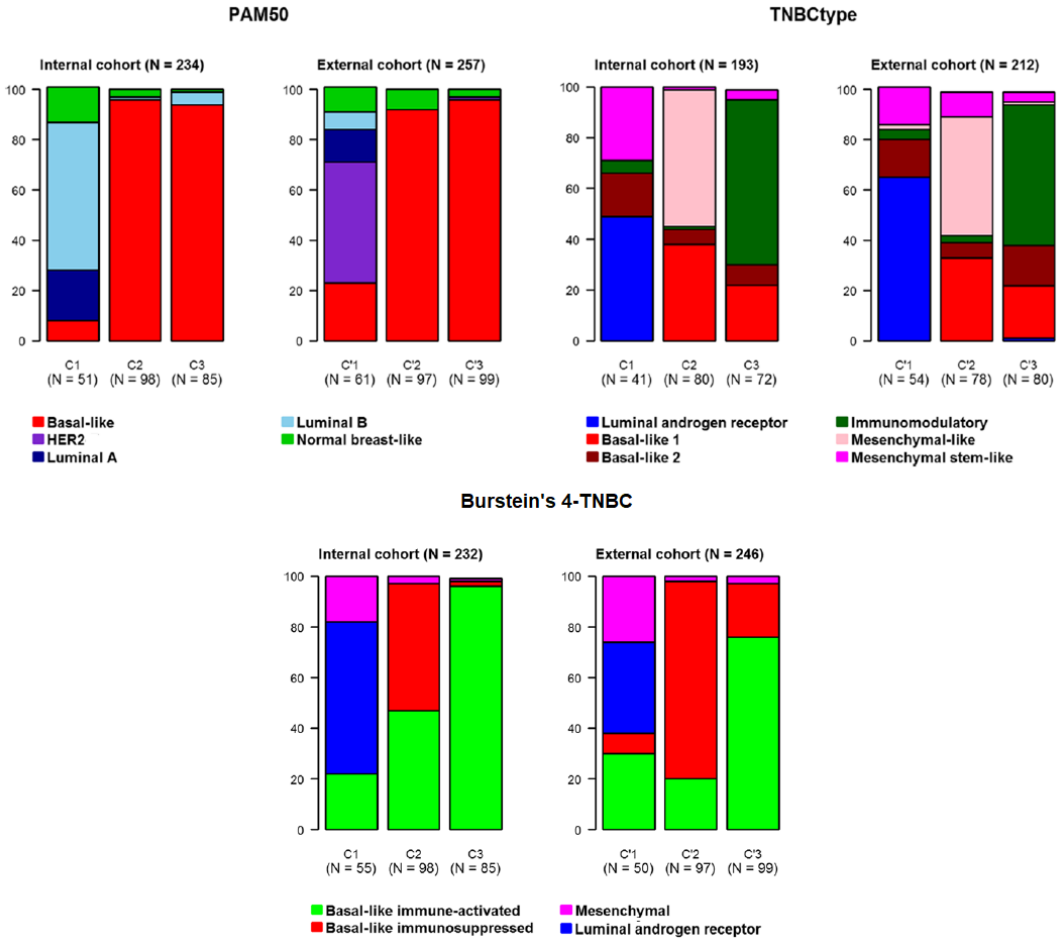


Figure 7 Subtype distributions of patients between the three clusters, PAM50 (Prat et al., 2012), TNBCtype (Lehmann et al., 2011) and Burstein's 4-TNBC (Burstein et al., 2015) by means of categorical 54 gene-expression signatures, for the internal (left) and external (right) TNBC cohorts. Adapted from Jézéquel et al. 2019.

Description of the TNBC subtypes is presented below.

3.1.4.1. Balsal-like subtypes

TNBCtype and TNBCtype-4 (Lehmann et al., 2016, 2011) BL1 and BL2 subtypes share a highly proliferative phenotype, being enriched in *EGFR*, *MET* and *EPHA2*, which correlated with improved pCR with mitotic inhibitors, such as taxanes. BL1 is enriched in genes involved in DNA-damage response and cell-cycle regulation, including the highest rate of *TP53* mutations (92%), high gain/amplifications of *MYC, CDK6*, or *CCNE1*, and deletions in *BRCA2, PTEN, MDM2*, and *RB1* (Bareche et al., 2018). On the other hand, BL2 subtype has high levels of growth factor signaling, namely EGF pathway, NGF pathway, MET pathway, Wnt/β-catenin, and IGF1R pathway, as well as increased metabolic pathway activity, such as glycolysis and gluconeogenesis (Lehmann et al., 2011). Even though BL1 tumors are more likely to be of higher grade, they are more responsive in general to genotoxic chemotherapies. The latter is likely due, in part, to aberrant DNA signaling and repair functions in the BL1 subtype tumors. On the

other hand, BL2 patients may not benefit of CT but still experience unnecessary and sometimes harmful side effects (Lehmann et al., 2016).

Burstein also identified two basal-like clusters (Burstein et al., 2015), however, their distinction was according to the positive or negative expression of immune regulators. BLIS subtype exhibits downregulation of B cell, T cell, and natural killer (NK) cell immune-regulating pathways, and cytokine pathways, and patients have the worst disease-free survival (DFS) and disease-specific survival (DSS). On contrary, BLIA immunoregulation pathways are upregulated, including genes controlling B cell, T cell, and NK cell functions, and patients have the best prognosis (Burstein et al., 2015).

A high correlation has been described between PAM50 basal-like, TNBCtype and TNBCtype-4 BL1/BL2, and Burstein's BLIA/BLIS subtypes, emphasizing the high stability of the basal TNBC subtype (Bianchini et al., 2016; Burstein et al., 2015; Curtis et al., 2012b).

3.1.4.2. Immunomodulatory subtype and the immune regulation in TNBC

According to Lehmann's TNBCtype classification (Lehmann et al., 2011), IM subtype has increased expression of genes involved in immune cell processes. These processes include immune cell signaling, namely TH1/TH2 pathway, NK cell pathway, B cell receptor signaling pathway, and T cell receptor signaling; cytokine signaling pathways, such as IL-12 pathway, and IL-7 pathway; antigen processing and presentation; and signaling through core immune signal transduction pathways such as NFKB, TNF, and JAK/STAT signaling (Lehmann et al., 2011).

Only Lehmann's TNBCtype considers an independent immunomodulatory TNBC subtype (Lehmann et al., 2011). However, later, authors hypothesized that tumor infiltrating lymphocytes (TIL) levels in a tumor specimen would influence the IM subtype classification in a given TNBC. To test this hypothesis, authors scored levels of infiltrating lymphocytes relative to total nuclei in TNBC sections and compared the results relative to the TNBCtype call generated from correspondent RNA-sequencing (RNA-seq) data. Tumors classified as IM had indeed the highest average percentage of lymphocytes (38%). However, all of the other subtypes also had TILs, specifically BL2 (23%), MSL (21%), LAR (17%), BL1 (15%) and M (9%) (Lehmann et al., 2016). These data provide evidence that TILs contribute significantly to the gene expression profiles and that correlations to this signature should be considered as a descriptor of the tumor immune status rather than an independent subtype (Lehmann et al., 2016).

Interestingly, a published case report from 2017 used Ring's 101-gene assay (Ring et al., 2016) as a potential immuno-oncology diagnostic to identify a patient who tested negative for programmed death-ligand 1 (PD-L1), an immune checkpoint inhibitor, by IHC, but as IM-positive by TNBCtype. The patient, who had already received exhaustive CT, had few other treatment options. Partially based on the positive IM result, the patient was approved for treatment with pembrolizumab, a mAb that binds to programmed cell death (PD-1). After four treatment cycles, the patient experienced a complete radiologic response (Bhatti et al., 2017), showing that, although somewhat obsolete, the IM subtype might still be relevant in the clinical context.

3.1.4.3. Mesenchymal and mesenchymal-like subtypes

According to TNBCtype (Lehmann et al., 2011) mesenchymal-like TNBC subtypes, M and MSL, display similar expression profiles regarding pathways involved in cell motility, such as regulation of actin by Rho, extracellular matrix (ECM) receptor interaction, and cell differentiation pathways, including Wnt pathway, anaplastic lymphoma kinase [ALK] pathways, and TGF-β signaling. However, unique to the MSL subtype, there is the upregulation of genes involved in processes linked to growth factor signaling pathways, including inositol phosphate metabolism, EGFR, PDGF, calcium signaling, G-protein coupled receptor, and ERK1/2 signaling, as well as ABC transporter and adipocytokine signaling. Another interesting difference between the M and MSL subtypes is the fact that the MSL subtype expresses low levels of proliferation genes and shows enrichment in the expression of genes associated with stem cells, namely *BCL2*, *ALDHA1*, *BMP2*, or *THY1* (Lehmann et al., 2011). Furthermore, the MSL subtype also displays low expression of claudins 3, 4, and 7, consistent with claudin-low subtype of BC (Prat et al., 2010).

Burstein's M subtype is characterized by pathways involved in cell cycle, mismatch repair, and DNA damage networks, as well as hereditary breast cancer signaling pathways. Furthermore, genes normally exclusive to osteocytes (*OGN*) and adipocytes (*ADIPOQ*, *PLIN1*) and growth factors such as insulin growth factor (*IGF1*) are highly expressed in this subtype (Burstein et al., 2015). M subtype overlaps with TNBCtype's MSL (Lehmann et al., 2011) and present similarities with the claudin-low intrinsic subtype (Prat et al., 2010).

In Lehmann's TNBCtype refinement study (Lehmann et al., 2016) analysis of TNBC subtypes from matched tumor epithelium and stroma revealed that, in fact, the MSL component was significantly higher in the adjacent stromal cells compared to the tumor epithelium for each of the pairs ($p = 0.002$). Since MSL subtype classification is strongly weighted by stromal cell gene expression, besides removing IM subtype from TNBCtype, authors also removed the MSL subtype (Lehmann et al., 2016).

3.1.4.4. Luminal androgen receptor

The LAR TNBC subtype is the most concise among different studies (Burstein et al., 2015; Jézéquel et al., 2019; Lehmann et al., 2016, 2011; Ring et al., 2016). LAR group is also the most differential among TNBC subtypes, being enriched for hormonally regulated pathways and being dependent on androgen receptor (AR) signaling. Additionally, this subtype displays a luminal pattern of gene expression, namely high levels of *FOXA1, GATA3, SPDEF*, and *XBP1*. LAR tumors are enriched in mutations in *PIK3CA* (40-55%), *KMT2C* (19%), *CDH1* (13%), together with a higher prevalence of invasive lobular histology, *NF1* (13%), and *AKT1* (13%) (Bareche et al., 2018).

Although AR can be expressed in multiple molecular subtypes of TNBC, overlapping in 82% of cases with luminal-A- or luminal-B intrinsic subtypes (Bareche et al., 2018), the LAR subtype has the highest level of AR expression. It is predominantly subclassified in the non-basal TNBC subgroup, having a distinct prognosis that offers an opportunity to develop targeted therapeutics.

LAR tumors appear to respond poorly to conventional CT (Masuda et al., 2013). However, since AR is a potent mitogenic driver of the LAR subtype (Fioretti et al., 2014), and because previous data indicate that LAR cell lines and xenografts are sensitive to AR antagonists, these patients may benefit from simultaneous targeting of AR and the PI3K/mTOR pathway, a combination known to be synergistic in AR-dependent prostate cancer cells (Lehmann et al., 2011; Lehmann and Pietenpol,

2014). This and other corroborating data should prompt clinical trials to confirm the efficacy of AR antagonist therapy in LAR TNBC.

3.1.5. The Future of TNBC classification

To sum up, although different numbers of clusters were identified, four TNBC subtypes seem to be present in each of these works: a molecular apocrine/LAR, a mesenchymal and two basal-like enriched clusters with opposite immune status (pro-tumorigenic and anti-tumorigenic). Furthermore, the mesenchymal subtype is also greatly influenced by its microenvironment and immune regulation. Also, some discrepant results between the TNBC subtyping studies (Burstein et al., 2015; Jézéquel et al., 2019; Lehmann et al., 2016, 2011; Ring et al., 2016) suggest that TNBC heterogeneity is so marked that for each patient at least a secondary subtype should be taken into account for a more precise prognosis and treatment (Ring et al., 2016). Figure 8 displays a summary of the current TNBC subtypes.

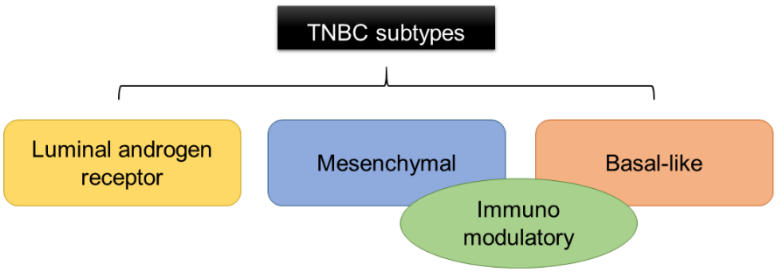


Figure 8 Summary of the current TNBC subtypes.

These findings in TNBC subtyping highlight a major limitation of classifiers defined based on the profiling of bulk tumors that cannot distinguish between tumor and stromal cells. Thus, increasing use of single-cell techniques to improve the characterization of the very heterogeneous tumor and its microenvironment is urgently needed. In fact, single-cell RNA-seq has demonstrated the presence of multiple subtypes within most primary TNBC tumors, suggesting that the dominant signature identified through bulk sequencing may not accurately inform underlying biological processes, including interactions between malignant and normal stromal cell types (Karaayvaz et al., 2018). Differences in the prevalence of intratumoral heterogeneity between TNBC and ER-positive BC could partly explain the challenges to date to apply commercially available gene-expression assays in routine clinical practice to provide prognostic and predictive information in TNBC (Garrido-Castro et al., 2019).

These studies also highlight the inherent problems associated with the TNBC definition, because it does not reflect a clear molecular entity. What seems clear is that luminal (AR-positive) and nonluminal (basal and mesenchymal) tumors have very different evolutionary paths, and this is in part likely driven by their normal cell-of-origin reflected in distinct epigenetic profiles. Thus, improved classifications based on epigenetic profiles and quantitative measures of intratumoral heterogeneity may lead to a better definition of clinically relevant TNBC subtypes (Garrido-Castro et al., 2019).

3.2. TNBC treatment

TNBCs lack ER and PR and do not express HER2. Since the cancer cells lack these proteins, treatment options for TNBC patients are limited. Since HT and drugs that target HER2 are not helpful,

CT is the main systemic treatment choice. Interestingly, although being more chemosensitive than other types of BC, TNBC is characterized to harbor the most aggressive behavior with the front-loaded risk of relapse within the first 3-5 years after completion of adjuvant CT, especially in younger women (Boyle, 2012; Carey et al., 2007; Masuda and Masuda, 2016).

In the last decade, extensive efforts were undertaken to unravel newer therapeutic targets for TNBC based on its molecular landscape (Burstein et al., 2015; Jézéquel et al., 2019; Koboldt et al., 2012; Lehmann et al., 2016, 2011). With the advances of multiomics (Banerji et al., 2012; Burstein et al., 2015; Curtis et al., 2012a; Jézéquel et al., 2019; Karaayvaz et al., 2018; Koboldt et al., 2012; Lehmann et al., 2011; Ring et al., 2016), recent experimental investigations revealed a variety of potentially actionable targets, including immune signatures (Kang and Syed, 2020; Saraiva et al., 2017). In fact, in 2020, the first targeted therapy was approved for TNBC, consisting in the combination of the checkpoint inhibitor atezolizumab with nab-paclitaxel (Kang and Syed, 2020; Schmid et al., 2020). However, this immunochemotherapy option is specifically for patients with unresectable locally advanced or metastatic TNBC, in which the tumors have a PD-L1 expression ≥ 1% (Kang and Syed, 2020). Thus, most patients with TNBC are still heavily dependent on conventional CT.

Despite the molecular heterogeneity, the standard of systemic treatment for TNBC follows the same general principle than with other types of BC (Bianchini et al., 2016; Moran, 2014). Hence, neoadjuvant or adjuvant CT remains a key component of systemic treatment in early TNBC, which is determined primarily by its clinical or pathologic stage (Moran, 2014).

3.2.1. Neoadjuvant and adjuvant chemotherapy

Neoadjuvant and adjuvant CT are the standard systemic treatment for early TNBC, and doxorubicin/docetaxel combination, anthracycline and taxane-based CT regimens comprise some of the standard of care (Albain et al., 2012; Cortazar et al., 2014), with pCR rates of approximately 40% (Liedtke et al., 2008; Von Minckwitz et al., 2012). Although the guideline for adjuvant CT is generally similar for all BC subtypes, adjuvant CT in TNBC is recommended for primary tumors larger than 0.5 cm due to their aggressive behavior (Wang et al., 2011).

Regarding the paradox that TNBC carries both higher chemosensitivity and the risk of early relapse (Carey et al., 2007), efforts have been continued to develop more effective chemotherapeutic regimens for both responders and non-responders. In the CREATE-X trial (Masuda et al., 2017; Zujewski and Rubinstein, 2017), patients with an initially high tumor burden or residual disease after NACT were selected for intensive systemic treatment, as they carry higher chance of relapse and metastatic spread. Results demonstrated the potential survival benefit of adding capecitabine to the standard adjuvant CT regimen in early TNBC with a residual tumor burden after neoadjuvant treatment (Masuda et al., 2017; Zujewski and Rubinstein, 2017). Despite its substantial toxicity, additional capecitabine might be a reasonable option for patients with TNBC carrying a higher risk of relapse. Additionally, in such patients, post-neoadjuvant trials based on actionable molecular targets identified from residual tumor tissues should also be considered, as well as molecular profiling of residual TNBC, which might contribute to find genetic alterations involved in drug resistance in the neoadjuvant setting and to further guide adjuvant targeted therapy to decrease the chance of clinically silent micrometastases formation (Park et al., 2018).

The additive benefit of platinum-based agent has also been consolidated in the neoadjuvant and adjuvant setting in the subset of TNBC patients presenting "BRCA-ness" phenotype (Section 3.1.2.). An unusual sensitivity to platinum-based agents was observed in these TNBC cases, particularly due genomic instability caused by DNA damage repair impairment (Kennedy et al., 2004; Yoshida and Miki, 2004). Subsequent clinical trials of metastatic TNBC showed a modest efficacy of platinum-based monotherapy (Isakoff et al., 2015; Tovey et al., 2015), consistently suggesting a greater benefit in BRCA1/2 mutation carriers, which present markedly higher pCR rates, from 61% to 90%, after neoadjuvant cisplatin monotherapy (Byrski et al., 2014, 2009; Silver et al., 2010). In later phase II neoadjuvant trials, paclitaxel combined either with cisplatin or carboplatin yielded encouraging efficacy outcomes with improved pCR and survival outcomes, which also suggested a strong relationship between clinical benefit and genetic alterations of DNA damage repair-related pathways (Jovanovic et al., 2017; Zhang et al., 2016).

Nonetheless, we should give a particular concern about the absence of statistically valid long-term survival benefits in the previous studies, since pCR do not necessarily translates into a long-term survival advantage, particularly in the presence of undetectable micrometastasis (Park et al., 2018). Furthermore, with drug combination, additional safety concerns will have to be taken into account. Thus, platinum-based combination CT should be selectively applied in specific patients, such as those carrying the "BRCA-ness". In addition, patients who need enriched locoregional control, namely those having an enhanced risk of relapse due to extensive cancer burden or of a younger age, might also be considered for combined treatment.

3.2.2. PARP inhibitors

Because of their critical role in the process of DNA damage repair, PARP inhibitors (PARPi) have been thought to be a potential game changer in TNBC treatment. PARPs, specifically PARP1 and 2, are enzymes that facilitate DNA damage repair at sites of single-strand breaks by activation of intracellular signaling pathways through auto-poly(ADP)-ribosylation (Audebert et al., 2004; Bürkle, 2001; Shall and de Murcia, 2000). An expressive link between PARP inhibition and "BRCA-ness" is confirmed by the concept of synthetic lethality. As BRCA1/2 deficiency leads to homologous recombination deficiency (HRD), a dysfunction of the cell's intrinsic DNA repair mechanism, repair of DNA damage solely depends on the action of PARP1 in BRCA-deficient BC (Bryant et al., 2005; Farmer et al., 2005). Thus, inhibiting PARP1 in patients with "BRCA-ness" might induce accumulation of double-strand breaks and eventually result in synthetic lethality, which could in turn enhance the sensitivity to PARPi (Bryant et al., 2005; Farmer et al., 2005; Turner et al., 2008).

Clinical data suggested that a subset of TNBC patients having BRCA deficiency might benefit from PARPi treatment (Tutt et al., 2010). The pivotal phase III OlympiAD trial of metastatic BC (Robson et al., 2017) showed olaparib effect in metastatic HER2-negative BC with germline BRCA1/2 mutations. The results were so dramatic comparing with standard single CT that these patients had almost double the response rate (59.5%) as well as a longer median progression-free survival (7.0 months) and a better toxicity profile, making PARPi approved for BC with BRCA1/2 mutations, which mainly constitutes TNBC (Robson et al., 2017). Other PARPi than olaparib have been also investigated and subjected to phase II/III clinical trials with metastatic BC and TNBC patients, namely veliparib (Anampa et al., 2018; Isakoff et al., 2017; Kummar et al., 2016; Rodler et al., 2016; Somlo et al., 2017), talazoparib (de Bono et al., 2017; Litton et al., 2015) and rucaparib (Miller et al., 2015).

Nonetheless, the possibility of a combination therapy in which PARPi are added to platinum-based CT is still controversial (Hope S. Rugo et al., 2016), for platinum itself already showed its efficacy either as monotherapy or in combination (Sikov et al., 2015; Von Minckwitz et al., 2014).

3.2.3. Inhibitors of the PI3K-Akt-mTOR pathway

TCGA data showed that the major genetic aberrations observed in TNBC occurred within the PI3K-Akt-mTOR (PAM) pathway (Koboldt et al., 2012), and accumulating data from NGS studies further confirmed PAM pathway as an appealing actionable target in TNBC (Marotti et al., 2017; Weisman et al., 2016). Particularly, *PIK3CA* hotspot mutations and aberrations in *PTEN* are the two most frequent but mutually exclusive alterations, accounting for approximately 10% and 35%-50% of TNBC, respectively (Bianchini et al., 2016).

The PAM pathway comprehensively controls the cell cycle from survival to apoptosis and, consequently, is crucially involved in tumorigenesis and cancer progression. PAM signaling is modulated by the key molecule PI3K but is also regulated by active communication with other growth factor tyrosine kinase receptors, including *EGFR* and *IGF1R*. In TNBC, PI3K was able to enhance the effects of BRCA1/2 mutations by interacting with the homologous recombination machinery, stabilizing DNA DSBs. Thus, inhibition of the PI3K pathway with buparlisib, an oral pan-PI3K inhibitor, produced promising antitumor cytotoxicity in TNBC cell lines. Experimentally, it also enhanced sensitivity to PARPi in both BRCA1/2-deficient and BRCA1/2-sufficient TNBC cell lines by activation of *ERK* and MAPK kinase (*MEK1*), which induced downregulation of BRCA1/2 (Ibrahim et al., 2012; Kimbung et al., 2012; Yi et al., 2015).

AKT, which is activated by PI3K, serves as another central node in the PAM signaling pathway. Preclinical studies demonstrated the antitumor activity of the AKT inhibitors in TNBC, such as ipataserib (Davies et al., 2012; Hudis et al., 2013; Kim et al., 2017; Sangai et al., 2012). Interestingly, LOTUS trial (Kim et al., 2017), involving a combination of paclitaxel and ipatasertib in phase II trial, was the first study showing a significant progression-free survival benefit of anti-PAM pathway targeted agents in IHC-defined TNBC population, although it was not maintained in the PTEN-low subgroup of patients, which was initially considered as an appealing candidate for the compound. However, when PIK3CA/AKT1/PTEN alterations were refined based on NGS, the benefit in these patients increased, suggesting that NGS-derived genomic biomarkers might better define the target population.

3.2.4. EGFR inhibitors

EGFR amplification is, in general, rare among patients with BC. However, EGFR is reportedly remarkably overexpressed in TNBC, ranging from 13% to 76% of all cases, and is considered a poor prognostic factor in TNBC patients (Nakai et al., 2016).

Based on the dynamic molecular network between EGFR and PAM signaling (Guerrab et al., 2016), preclinical studies revealed a therapeutic synergism between anti-EGFR targeted treatment and DNA damaging agents such as platinum or PARPi, by increasing chemosensitivity in TNBC cell lines with *BRCA1* deficiency but intact PTEN (Al-Ejeh et al., 2013; El Guerrab et al., 2017; Guerrab et al., 2016). In two previous phase II trials of metastatic TNBC, adding the anti-EGFR compound cetuximab to cisplatin or carboplatin improved clinical outcomes compared with platinum-based monotherapy regardless of

statistical insignificance (Baselga et al., 2013; Carey et al., 2012). Given that these studies did not account for biomarker-driven selection of target populations, further studies of EGFR inhibitors, particularly anti-EGFR mAbs, should be performed in a molecularly predefined subset of TNBC cases.

3.2.5. MEK inhibitors

The mechanism of MAPK activation in TNBC and basal-like BC was firstly investigated in vitro. These cell lines rarely have activating mutations of canonical oncoproteins such as Ras or Myc, however, they often harbor copy-number aberrations in these oncogenes or show overexpression of other growth factor receptors such as *EGFR*, *IGF1R*, *VEGFR* or *FGFR1*. Based on these molecular characteristics, MEK (mitogen-activated protein kinase kinase enzymes MEK1 and/or MEK2) inhibitors were suggested as an interesting therapeutic possibility in TNBC cases, particularly with wild-type *PTEN*, since *PTEN* was considered a potential negative predictor of response (Hoeflich et al., 2009). Nonetheless, in the presence of abundant signaling crosstalk, MEK inhibitors alone could not efficiently suppress activation of the MAPK pathway. Thus, combination treatments with MEK inhibitors and other targeted therapies or chemotherapeutic agents have been proposed to overcome chemoresistance. Among numerous ongoing combination trials with MEK inhibitors, a phase Ib trial combining the MEK inhibitor trametinib with gemcitabine in advanced solid tumors showed a case of complete response in the patient with metastatic TNBC (Infante et al., 2013). However, other phase I trials evaluating combination treatment of MEK inhibitors either with mTOR or PI3K inhibitors revealed unacceptable fatal toxicities, prohibiting subsequent phase II studies (A W Tolcher et al., 2015; Anthony W Tolcher et al., 2015). These studies showed the potential lethality of co-blockade of master signaling pathways, such as MAPK and PAM, which should be taken into consideration when designing future combinatorial trials with targeted agents.

3.2.6. Immunotherapy for TNBC

Immunotherapy is a promising new direction for the management of TNBC, with immune checkpoint blockade demonstrating strong anti-tumor activity and prolonged patient survival (Kang and Syed, 2020; Khosravi-Shahi et al., 2018; Mittendorf et al., 2014).

TILs are a well-known prognostic factor in TNBC, particularly in early-stage disease, and is positively correlated to both patient survival and pathological complete response after NACT. In addition, TILs have shown a predictive value in patients with TNBC who were treated with immune checkpoint inhibitors monotherapy, and their assessment is being implemented as a stratification factor in BC immunotherapy trials (Criscitiello et al., 2016; Dieci et al., 2015; Kang and Syed, 2020; Loi et al., 2019, 2013). In TNBC treated with neoadjuvant treatment, TILs were identified as a robust predictive biomarker of long-term survival and its significance in remnant disease was subsequently validated (Adams et al., 2014; Denkert et al., 2015; Gu-trantien et al., 2013; Ibrahim et al., 2014; Issa-Nummer et al., 2013; Loi et al., 2014, 2013), revealing an active communication between immune system and cytotoxic agents (Adams et al., 2014; Bates et al., 2006; Denkert et al., 2015, 2010; Dieci et al., 2015; Liu et al., 2011; Loi et al., 2013; Oda et al., 2012). The significance of TILs in the adjuvant setting has also been studied (Adams et al., 2014; Dieci et al., 2015; Gu-trantien et al., 2013; Loi et al., 2014, 2013), suggesting that TILs have prognostic significance in systemically untreated early TNBC. Also, lower level of TILs was significantly associated with enhanced activation of Ras-MAPK signaling,

which, in turn, can trigger immune evasion in TNBC [207]. Thus, the presence of TILs may help in candidate refinement for adjuvant CT or immunotherapy (Leon-Ferre et al., 2018).

Compared with other BC subtypes, TNBC is not only more likely to harbor TILs but also to express PD-L1, mainly on TILs rather than on tumor cells (Mittendorf et al., 2014; Molinero et al., 2019; Schmid et al., 2018). TILs and PD-L1 expression in TNBC has been shown to usually range from 20 to 65% in tumor cells and/or immune infiltrate cells (Beckers et al., 2016; Cimino-Mathews et al., 2016; Kwa and Adams, 2018; Li et al., 2016; Mittendorf et al., 2014). Retrospective studies with early TNBC cases demonstrated significantly worse survival outcomes in patients harboring high PD-L1 expression and a low level of TILs or a high ratio of PD-L1/CD8 expression (Mori et al., 2017; Okabe et al., 2017; Tomioka et al., 2018). Remarkably, the only approved targeted therapy for TNBC consists in Atezolizumab, an immune checkpoint inhibitor (ICI) against PD-L1, in combination with nanoparticle albumin-bound paclitaxel (nab-paclitaxel) (Kang and Syed, 2020; Schmid et al., 2020). Approval was based on the results of the phase III IMpassion130 trial in patients with unresectable locally advanced or metastatic TNBC, in which atezolizumab plus nab-paclitaxel significantly prolonged progression-free survival (PFS) when compared to placebo plus nab-paclitaxel in the intent-to-treat population and the PD-L1+ subgroup (PD-L1 expression ≥ 1%) (Schmid et al., 2020, 2018).

All things considered, TILs have a major role in orchestrating the immune microenvironment and dynamically interact with cytotoxic signals from both CT and immunotherapy (Pruneri et al., 2016; Wein et al., 2017). In fact, CT can induce multiple immunomodulatory changes in the tumor microenvironment, such as increased antigen release by tumor cells, PD-L1 upregulation, which may be behind the good results of the IMpassion130 trial (Schmid et al., 2020), and overexpression of immunogenic cell surface markers (e.g., MHC class I) (Pruneri et al., 2016; Wein et al., 2017).

Similarly, studies have been assessing the combination of targeted agents with ICIs. In a study, PARPi modulated cancer-associated immunosuppression by upregulating PD-L1 in BC cell lines, suggesting that blockade of PD-L1 could restore their sensitivity to PARPi. TNBC and basal-like BC cell lines showed broad sensitivity to MEK inhibition, and the generation of effector T cells was enriched by the treatment. These results suggested the MEK inhibitor as another weapon to harness immune surveillance as well as a potential synergy with ICIs (Loi et al., 2016).

Currently, a selection of potential immune biomarkers is being evaluated to predict chemo and immunotherapeutic efficacy in BC beyond PD-L1 expression, including gene signatures, TILs and/or their particular subsets (Saraiva et al., 2018), tumor mutational burden, microsatellite instability, and mismatch repair (MMR) deficiency (Marra et al., 2019).

3.2.7. The Future of TNBC treatment

Recent advances in integrative omics introduced a variety of actionable targets and increased the new horizon of targeted therapies, which could be particularly successful in subgroups of TNBC patients that are expected to present more chemoresistance. However, increasing the treatment spectrum does not necessarily ensure consequent therapeutic benefits. Therefore, it is important to recognize that molecular heterogeneity is a double-edged sword, which could be the hope or the hype for TNBC. Furthermore, although plenty of novel immune-molecular targets have been experimentally validated, so far only one study transformed these preclinical results in clinical benefits in the daily practice and current standard treatment (Kang and Syed, 2020; Schmid et al., 2020). Additionally,

customized clinical trials based on individualized genotyping and transcriptomics seems increasingly inevitable, in order to set precise personalized medicine against cancer wide evolutionary mutational spectrum. In parallel, optimization of conventional CT in the landscape of immune-molecular heterogeneity should also be continued to establish the most effective regimens. Juggling with these efforts, the next chapter of TNBC treatment should focused on the novel combinatorial strategies, which demand refinement of target populations, more robust biomarkers, and drug compounds. Figure 9 depicts the interconnection of the possible therapeutic strategies in patients with TNBC based on its chemosensitivity and immune molecular heterogeneity.

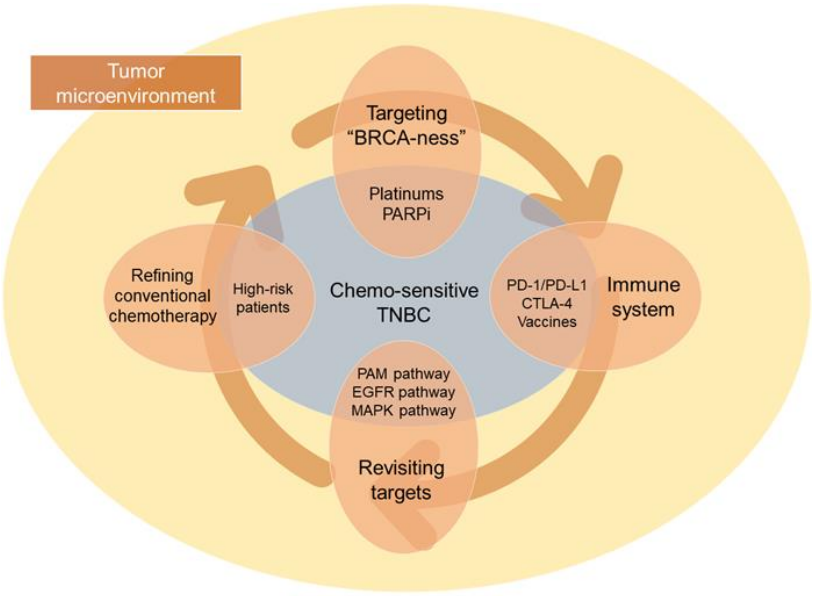


Figure 9 Future aspects of therapeutic strategies in patients with TNBC based on its chemosensitivity and immune molecular heterogeneity. Future challenge in TNBC is fundamentally to enrich the therapeutic efficacy to the optimal level both for chemosensitive and chemo-resistant population. Thus, conventional chemotherapy and these four key entities constitute the main domain of upcoming treatment strategies. Targeting "BRCA-ness", revisiting targets including PAM, EGFR and MAPK pathways and emerging immunotherapy can be the master molecular regulators of TNBC tumor microenvironment. Smart refining of conventional chemotherapy should be accompanied with these molecular targeting. Finally, arrows represent combinatorial chains between these four independent domains and these would be the key of future therapeutics for TNBC. *CTLA-4* cytotoxic T-lymphocyte-associated protein 4. *EGFR* epidermal growth factor receptor. *MAPK* mitogen-activated protein kinase. *PAM* PI3K-Akt-mTOR pathway. *PARPi* poly(ADP-ribose) polymerase inhibitors. *PD-1* programmed cell death protein 1. *PD-L1* programmed cell death ligand 1. Adapted from Park et al., 2018.

4. THE ANCESTRY-ASSOCIATED DISCREPANCY OF TNBC

The association between women of African-ancestry, frequently referred to as Black or Black-origin women, by both World Health Organization (WHO) and TCGA, and an increased susceptibility to more aggressive BC goes back to the 1980s, before BC subtypes were part of our thought process. At that time, BC in African-ancestry women was known as being more frequently ER-negative, affecting women of younger age and, when stage and age matched, had worse prognosis than women from other population (Coates et al., 1990; Crowe et al., 1986; Nachimuthu Natarajan et al., 1985; Swanson and Lin, 1994).

Most African countries have higher age standardized BC mortality rates than what is reported in other continents (Sung et al., 2021). Overall, women from the African continent, Caribbean, Melanesia and Polynesia have a higher cumulative risk of mortality by BC than other regions of the globe (Sung et al., 2021). Nonetheless, the annual incidence of BC varies among African-ancestry populations and accurate reporting of the data may be affected by the inadequacy of cancer registries in low-income countries (Birnbaum et al., 2018).

Ancestry-associated disparities in BC-related mortality have also been reported in high-income countries with mixed populations. For instance, in 2018, BC-related mortality in African-American (AA) women, with 28.4 deaths per 100 000 (DeSantis et al., 2019), was quite similar by comparison with the age-standardized cancer-related mortality of women living in the non-Latin Caribbean, who are predominantly of African-ancestry, with 25.0 deaths per 100 000, despite decreased health-care capacity (Razzaghi et al., 2016).

In the USA, AA women have a 40% higher BC-related death rate than women of European/Caucasian descent, hereafter denominated as White patients, as stipulated in TCGA race list (DeSantis et al., 2019). The same ancestry-associated discrepancy is observed in the UK, where a 5-year distant BC relapse-free survival is 62.8% for young African-ancestry women, compared with 77% for young White women with equal access to health care ($p = 0.0053$) (Copson et al., 2014). Although differences in the biological characteristics of BC play a major role in these ancestry-associated discrepancies, availability of early detection, access to diagnosis and treatment, cultural differences in lifestyle behaviors and socioeconomic factors also influence prognosis and treatment (Newman et al., 2006).

In the main scope of this project, TNBC is particularly disproportionately prevalent in African-ancestry women (Brewster et al., 2014; Danforth, 2013; Keenan et al., 2015; Newman and Kaljee, 2017; Sørlie et al., 2001; Stark et al., 2010). Overall, TNBC accounts for about 10-20% of all BC cases but such percentage increases in African-ancestry women when compared with other ethnic origins, particularly in women who develop BC before the age of 50 years (Bauer et al., 2007; Shoemaker et al., 2018). Overall, while White and Asian women tend to have TNBC at a later age of onset and with less aggressive clinical course (Kwong et al., 2009; Lim et al., 2009; Saraiva et al., 2017), African-ancestry women are often premenopausal (Carey et al., 2006; Huo et al., 2009; Lund et al., 2009; Zaky et al., 2009), with a faster and more aggressive development of TNBC, independent of other risk factors for BC (Stead et al., 2009). Accumulating evidence suggest that African-ancestry TNBC patients present more unfavorable clinicopathological characteristics such as larger tumor size, higher proliferation and more extensive lymph node involvement than White TNBC patients (Dietze et al., 2015a; Lund et al., 2009; Sullivan et al., 2014), even when socioeconomic factors are accounted for (Dietze et al., 2015a; Sachdev et al., 2010). Additionally, AA patients have been reported to harbor more aggressive TNBC subtypes such as BL1 and MSL, according to TNBCtype classification (Lehmann et al., 2011), as well as greater intratumoral heterogeneity than White patients (Keenan et al., 2015; Lindner et al., 2013). Also, African-ancestry women are less responsive to NACT and have worse prognosis when diagnosed with locally advanced TNBC (Frasci et al., 2009).

Although there is evidence that disparities in healthcare access, co-morbid diseases and income affect the stage of presentation and survival of African-ancestry women with TNBC (Danforth, 2013; Vona-Davis and Rose, 2009), emergent data strongly suggests that biological disparities may drive aggressive BC and, specifically, TNBC development in African-ancestry patients (Grunda et al., 2012;

Keenan et al., 2015; Siddharth and Sharma, 2018). Thus, it is crucial to dissect the potential regulators and molecular mechanisms involved in the higher incidence and development of aggressive TNBC in these women. Such knowledge may lead to the development of models that specifically capture the risk of African-ancestry women for TNBC, increase access to effective early detection and lead to the implementation of health policies that improve the outcome and life quality of these patients.

4.1 Demographics

TNBC incidence is globally higher in African-ancestry women than in women from other ancestries. In USA-based studies, the high incidence of TNBC in AA women was observed in BC cohorts from North Carolina (Carey et al., 2006), California (Amirikia et al., 2011; Clarke et al., 2012), Philadelphia (Morris et al., 2007), Boston (Stead et al., 2009), Georgia (Lund et al., 2009), and Michigan (Stark et al., 2010). Studies also reported that TNBC is particularly prevalent in premenopausal AA women when comparing with White women (Amirikia et al., 2011; Carey et al., 2006). In fact, the group of non-Hispanic African-ancestry women younger than 44 years had the highest lifetime TNBC incidence rates and higher incidence rates of stage III and IV disease (Amirikia et al., 2011).

In African countries, although the incidence of BC is lower than in higher income countries, several studies showed that TNBC occurs in sub-Saharan and West African women at a higher frequency and at a younger age than in AA women (P. Boyle, 2012; Fregene and Newman, 2005; Newman et al., 2019; Stark et al., 2010). In general, BC incidence rates among sub-Saharan African women are low (10-40 per 100 000), however, these women experience higher mortality rates (5-20 per 100 000) and are more often diagnosed at a younger age than either AA or White women, with a peak incidence at 35-45 years. In 2009, in one of the fist BC-related epidemiolocal studies with data from African countries, from six locations in Nigeria and Senegal, Huo et al. observed that TNBCs were indeed predominant, making 55% of all cases (Huo et al., 2009). In a Mali-based study, in which BC patients had a mean age of 46 years, the TNBC subgroup represented 46% of the cohort and displayed a particularly aggressive pattern, with 90% of cases being grade III and 88% highly proliferative (Ki67 < 20%) (Ly et al., 2012). Furthermore, in a study comparing Detroit-based patients and Ghanian patients diagnosed with invasive BC, 76% of Ghanians, 7.9% of AAs and 2.8% White-Americans were diagnosed with poorly differentiated and advanced stage (III/IV) BC. Of those patients, 83% of Ghanaian women had TNBC, while 41.9% of AA women and 15.4% of White women had TNBC (Stark et al., 2010). Later, the same group included Ethiopian/east Africans BC patients (Jiagge et al., 2016). A total of 234 Ghanaian (mean age 49 years), 94 Ethiopian (mean age 43 years), 272 AA (mean age 60 years), and 321 White-American (mean age 62 years, $p = 0.001$) patients were compared. Again, TNBC was more common among Ghanaian (53.2%) and AA (29.8%) compared with White (15.5%) and Ethiopian (15.0%) cases and, regarding patients younger than 50 years, prevalence of TNBC remained highest among Ghanaians (50.8 %) and AA (34.3 %) compared with White and Ethiopian patients (approximately 16 % in each, $p = 0.0002$). Thus, this study shows an association between TNBC and West African ancestry (Jiagge et al., 2016). Figure 10 shows the prevalence of TNBC among White and Black-origin patients from the USA and among African countries, according to epidemiological studies involving patients from those countries, compiled by Siddhart et al. (Siddharth and Sharma, 2018).

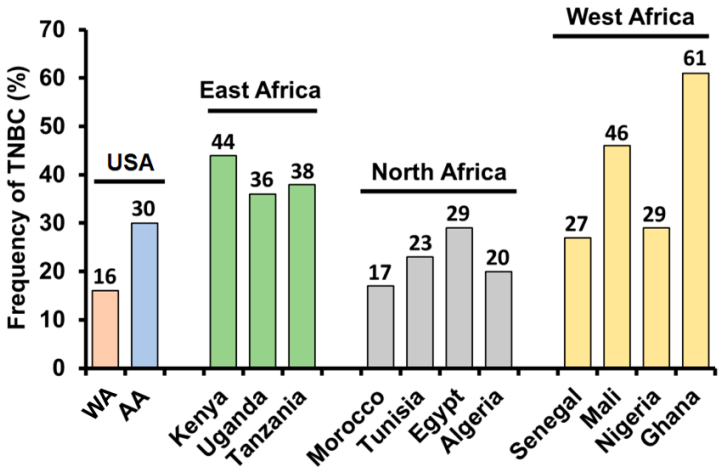


Figure 10 Prevalence of TNBC among USA and African countries. *WA* White-American. *AA* African-America. Adapted from Siddhart et al. (2018).

To our best knowledge, the only Portuguese-based study comparing the ancestry-associated discrepancy of TNBC incidence and aggressiveness in African-ancestry women was performed with data collected from the public Hospital Professor Doutor Fernando Fonseca (HFF), which is located in a catchment area where a large part of the population is from the former Portuguese colonies (Honório et al., 2016). In this study, from a total of 1552 BC patients, 144 had TNBC (9.3% of the whole sample), most of which do not had information regarding ancestry. Among the TNBC patients, 17 patients were identified as being from African-ancestry (12%). Authors also observed that the prevalence of initially metastatic BC was significantly higher among the African population, comparing with patients from other populations or without indication of ancestry (41.2% vs 11%, $p < 0,005$), and the outcome was also worse for African-ancestry patients (median survival: 62 vs 15 months, $p < 0.005$) (Honório et al., 2016).

4.2 Prognosis

African-ancestry women have higher overall BC mortality compared with White women (Carey et al., 2006; Chlebowski et al., 2005; Shen et al., 2007), with a mortality rate in AA 40% higher than in White women (28.4 vs 20.3 deaths per 100 000). This disparity is magnified among AA women aged <50 years, who have a death rate double than White women (DeSantis et al., 2019). Furthermore, African-ancestry women also have a higher rate of tumor recurrence compared to other populations across all BC subtypes (Keenan et al., 2015; Kroenke et al., 2014).

Contributing factors include disparities in income, barriers to screening, differences in treatment, higher stage of disease at diagnosis and increased incidence of TNBC (Ayanian et al., 1993; Birnbaum et al., 2018; Bradley, 2002; Furberg et al., 2001; McWhorter and Mayer, 1987; Servick, 2014). Although Western nations have a higher incidence of BC in comparison to developing nations, mortality is highest in developing nations. The factors influencing patients' treatment and survival are affected by cultural and economic barriers and differ significantly throughout the world (Birnbaum et al., 2018; Dietze et al., 2015b; Servick, 2014).

As with general BC cases, AA patients with TNBC have also been reported to experience shorter OS and progression-free survival than their White counterparts (Dietze et al., 2015a; Lund et al., 2009; Sullivan et al., 2014). Although not taking in account BC subtypes, Newman et al. (Newman et al., 2006) conducted a large meta-analysis of over 13 000 AA women with BC compared with 75 000 White women, observing that AA women had nearly 30% higher mortality rate when compared with White patients. In other study, although African ancestry was associated with poor prognosis, it was unclear whether poor survival was due to differences at the molecular level of TNBC, beyond an increased frequency of this type of BC (Carey et al., 2006).

Overall, most studies suggest that African-ancestry women diagnosed with TNBC have lower survival rates, worse response to treatment and faster relapse (Albain et al., 2009; Bauer et al., 2007; Lund et al., 2009; Shen et al., 2007; Woodward et al., 2006). Particularly, Woodward et al. (Woodward et al., 2006) explored the effect of ancestry on survival outcome in TNBC patients after doxorubicin-based NACT or CT and reported that, despite a uniform distribution of treatment, African ancestry was independently associated with poorer survival in both treatment regimes. Albain's group (Albain et al., 2009) found that AA women with early-stage premenopausal BC had a significantly worse prognosis than age- and prognosis-matched White women. Bauer et al. (Bauer et al., 2007) and Lund et al. (Lund et al., 2009) also reported a worse survival for AA women with TNBC after controlling for socioeconomic factors and treatment delay. Taken together, these studies provide evidence that even after controlling for treatment disparities, biological differences may contribute to the poor survival of women with TNBC. However, studies investigating TNBC at molecular level are needed to better define potential biological differences.

Currently, there are no reliable methods to identify African-ancestry TNBC patients at high risk of poor outcomes, which would indicate the need to offer more tailored treatments. Identification of biomarkers that can risk-stratify African-ancestry TNBC patients, particularly the younger, and predict responsiveness to targeted and cytotoxic agents could improve prognosis of this high-risk patient population.

4.3. Biology of BC and TNBC in African-ancestry patients

Explanations for the biological and hereditary racial disparities of cancer are complex and studies often include factors such as socioeconomic and cultural (Newman, 2017; Newman and Kaljee, 2017). However, even when accounting for these factors, the observed differences in the incidence, age of onset and disease aggressiveness strongly suggest a biological role for differences regarding oncogenic signaling pathways of breast tumorigenesis between African-ancestry and White women, particularly in the TNBC subtype (Chen and Li, 2015; Chlebowski et al., 2016; Coates et al., 1990; Daly and Olopade, 2015; Iqbal et al., 2015).

4.3.1. Risk factors

White women diagnosed with BC experienced increased survival rates in 1999-2005 relative to 1991-1998 in the USA. By contrast, AA women did not experience improved BC survival during this time period (Silber et al., 2013). As previously mentioned, lower survival rates for AA women are attributed to advanced disease presentation, co-morbid diseases, disparities in income, lack of access to BC screening and oncology care, and delays in treatment (Vona-Davis and Rose, 2009). Particularly,

AA women experience an unequal burden of co-morbid disease, including obesity and diabetes (Black, 2002).

4.3.1.1. Obesity, diabetes and inflammation

According to the Center for Disease Control and Prevention (CDC), obesity is defined as a body mass index (BMI) ≥ 30 kg/m^2 and these individuals are considered as metabolically unhealthy (https://www.cdc.gov/obesity/adult/defining.html, last consulted on 24 March 2021). Waist/hip ratio (WHR) is also used to measure the abdominal obesity. According to the WHO, a WHR of ≥ 0.85 is also considered to be an increased risk for metabolic disorders (World Health Organization, 2008).

Several studies have shown that obesity is a general risk factor for TNBC (Kwan et al., 2009; Lee et al., 2011; Pierobon and Frankenfeld, 2013), particularly in premenopausal AA women (Bandera et al., 2015; Carey et al., 2006; Millikan et al., 2008). Because there is a higher incidence of obesity in AA women and obesity predicts poor survival, it is hypothesized that obesity is a potential driver of aggressive TNBC biology in AA women, however no clinical trial has directly evaluated TNBC progression in obese African-ancestry and White women. There are many potential biological mechanisms by which obesity might increase the incidence of aggressive subtypes of TNBC in African-ancestry women. Obese women have tissue inflammation associated with increased circulating levels of insulin and inflammatory cytokines, including interleukin-6 (IL-6), IL-8, tumor necrosis factor (TNF) and leptin. IL-6 and IL-8 activate signal transducer and activator of STAT3, NFKB and EZH2 signaling and predict poor prognosis in women with TNBC (Creighton et al., 2009; Dietze et al., 2018; Hartman et al., 2013; Kahn et al., 2006; Sharma and Davidson, 2013).

Type 2 diabetes (T2D) is another co-morbidity which may increase the risk for BC development. Meta-analyses of T2D and risk of overall BC have shown an approximately 20% increased risk associated with T2D (Boyle et al., 2012; Larsson et al., 2007). In 2017, Palmer et al. published a study involving more than 54 000 AA women who were cancer-free at the start of the study, being followed during a period of 18 years (Palmer et al., 2017). Analysis of this cohort of AA BC patients suggested that that AA women with T2D have a 40% increased risk of developing ER- BC. Additionally, T2D was not associated with incidence of ER+ BC in AA women (Palmer et al., 2017). Multiple mechanisms have been proposed to explain associations of T2D with BC risk (Ferroni et al., 2015; Giovannucci et al., 2010), such as the dysregulation of glucose metabolism experienced by diabetics that can lead to a chronic proinflammatory condition with associated oxidative stress and promotion of tumor initiation and progression (E. Goldberg and L. Schwertfeger, 2010; Gunter et al., 2015). In a chronic inflammatory state, there are typically high levels of proinflammatory cytokines (Kahn et al., 2006) and infiltration of adipose depots with immune cells, notably proinflammatory macrophages (Weisberg et al., 2003). These adipose tissue macrophages secrete cytokines that promote insulin resistance in adipocytes (Stephens et al., 1997). Evidence is accumulating that chemokines and cytokines of the breast adipose microenvironment, such as IL-6 (Walter et al., 2009), promote carcinogenic processes in epithelial cells, including increased cell proliferation and survival, and EMT in early stage BC cells (Grivennikov et al., 2010; Iyengar et al., 2003; Quail and Joyce, 2013; Sullivan et al., 2009) (Figure 11).

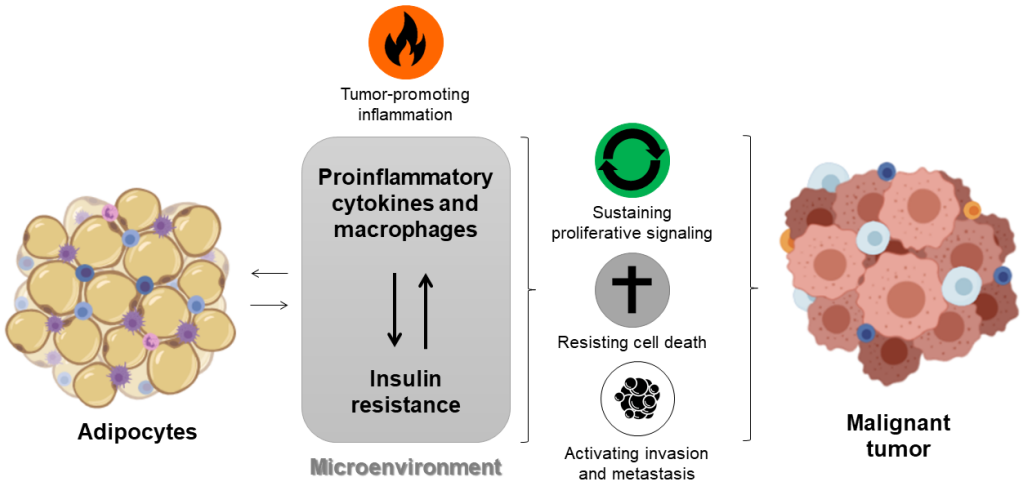


Figure 11 Adipocytes and insulin resistance promote secretion proinflammatory cytokines in the tumor microenvironment and induce the major hallmarks of cancer development, such as proliferation, survival, invasion and metastasis. Created with BioRender.

4.3.1.2. Breast density

Breast density is higher in African-ancestry women. Not only the rigidity of the extracellular matrix can be a factor that drives the malignancy (Boyd et al., 2014; Collett et al., 2005; McCormack and Dos Santos Silva, 2006), as cancer in dense breasts is harder to diagnose by mammography (Barlow et al., 2006; Boyd et al., 2014), raising the possibility that the worse prognosis in these cases might be due to a delay in diagnosis. Additionally, denser mammary glands may have distinct growth factor profiles that predispose women to more aggressive cancer with estrogen independence. Nevertheless, breast density as a risk factor in African-ancestry women is controversial, since stage matching in epidemiological studies contradict this hypothesis (Albain et al., 2009; Chlebowski et al., 2016).

4.3.1.3. Breastfeeding, involution and inflammation

Approximately 21% of AA women breastfeed in comparison to 37% of White women (Hall et al., 2005). The lack of breastfeeding reportedly increases the risk of women developing TNBC (Hall et al., 2005; Palmer et al., 2011; Shinde et al., 2010). After lactation, or parturition in the absence of nursing, the mammary gland undergoes involution (Lyons et al., 2011). Using mice, Lyons et al. have shown that breast tissue involution results in tissue inflammation and activation of cyclooxygenase 2 (*COX2*, also known as prostaglandin G/H synthase 2), and wound-healing programs that promote the deposition of high-risk cross-linked, fibrillar collagen (Lyons et al., 2011). Although not confirmed in African-ancestry BC patients, aligned fibrillar collagen predicts poor survival in White women with invasive BC (Conklin et al., 2011). Nonetheless, no breastfeeding as a risk factor for BC and TNBC development in African-ancestry women is also controversial, since studies performed in African countries, where the fertility rate is among the highest in the world, are still lacking. Thus, overall, further studies are needed to better understand the molecular link between pregnancy, parity and TNBC in African-ancestry women.

4.3.2. TNBC profile in African-ancestry patients

The fact that basal-like BCs/TNBCs account for a preeminent percentage of BC in African-ancestry patients were first outlined by the Carolina Breast Cancer Study, in 2006 (Carey et al., 2006). In 2010, it was published the first study using massively parallel DNA sequencing technologies, which provided an unprecedented ability to screen entire genomes for genetic changes associated with tumor progression, focused in an 44-year-old AA patient with inflammatory basal-like BC (Ding et al., 2010). In this study, authors performed genomic analyses of four DNA samples from that patient, including peripheral blood, the primary tumor, a brain metastasis and a xenograft derived from the primary tumor. Results showed that the metastasis contained two de novo mutations and a large deletion not present in the primary tumor and was significantly enriched for 20 shared mutations. The xenograft retained all primary tumor mutations and displayed a mutation enrichment pattern that resembled the metastasis. Also, overlapping large deletions, encompassing *CTNNA1*, were present in all three tumor samples. Thus, the study of the primary tumor-metastasis-xenograft trio suggested that, although additional somatic mutations, copy number alterations and structural variations do occur during the clinical course of the disease, most of the original mutations and structural variants present in the primary tumor are propagated (Ding et al., 2010).

Although not comparing the obtained results with those from patients with other ancestries, this study (Ding et al., 2010) opened a path for the investigation of molecular mechanisms that potentially drive the higher incidence and aggressiveness of TNBC in African-ancestry women. Since most translational studies and clinical trials are performed in Europe, North America and East Asia, African-ancestry patients included in those studies are scarce or even inexistent. Thus, if cancer in general, and BC in particular, still present more questions than answers regarding the main regulators, molecular mechanisms and development, African-ancestry patients are in an even more hazardous position, especially in relation to the massively heterogeneous TNBC disease. Hence, the study of the driving molecules and pathways and respective effects in TNBC incidence and progress in African-ancestry patients is crucial. Some of the most comprehensive discoveries at the level of molecular biology of TNBC in African-ancestry patients from the past years are highlighted in the next subsections.

4.3.2.1. BRCA1 incidence

Most BCs with *BRCA1* mutations are TNBCs (69%) (Mavaddat et al., 2012). Although the incidence of TNBC is high in African-ancestry women, several studies show that the incidence of germline *BRCA1* mutations is low relatively to the incidence in White women. In a study of 155 high-risk families, AA had a lower rate of deleterious germline *BRCA1* mutations compared with non-Hispanic, non-Jewish White patients (27.9% vs 46.2%, respectively), but a higher rate of sequence variations (44.2% vs 11.5%, respectively, $p < 0.001$) (Nanda et al., 2005; Olopade et al., 2003). In fact, deleterious mutations in *BRCA1* are frequently find in Ashkenazi-Jewish women (69.0%) (Nanda et al., 2005). Additionally, in a TNBC cohort, while 50% of White non-Ashkenazi-Jewish women had germline *BRCA1* mutations, fewer than 20% of AA women had germline *BRCA1* mutations (Greenup et al., 2013). This suggests that other genetic mechanisms beyond germline mutation of *BRCA1* may promote TNBC in AA women.

In fact, a diverse spectrum of *BRCA1* mutations and sequence variations unique to African-ancestry women have been reported (Fackenthal et al., 2005; Gao et al., 2000; Oluwagbemiga et al.,

2015; Szabo and King, 1997; Zhang et al., 2009, 2010), with only a small fraction of the mutations resulting in known structural defects that block BRCA1 function (Fackenthal et al., 2005; Gao et al., 2000; Szabo and King, 1997; Zhang et al., 2009, 2010). In 2017, Polak et al. (Polak et al., 2017), using TCGA data, evaluated differences in the frequencies of genetic (germline and somatic) and epigenetic events. Authors reported that epigenetic silencing of *BRCA1* and *RAD51C* - a member of RAD51 family which is involved in DNA repair by homologous recombination, promoting DNA-strand invasion and homology search (Nielsen et al., 2016) - due to promoter-methylation events, was highly enriched in basal-like BCs in young AA patients. This suggests that the mechanism of HRD differs between White patients, in whom genetic mutations predominate, and AA patients, in whom promoter methylation is more frequent ($p = 0.0009$) (Polak et al., 2017).

Furthermore, *EZH2*, a member of Polycomb group family that blocks *BRCA1* function by inducing protein kinase B (PKB or AKT) dependent genomic instability (Gonzalez et al., 2011), was linked to aggressive TNBC in AA women. Overexpression of nuclear EZH2 was significantly associated with basal-like TNBC in African-ancestry women in a joint study by Kleer and Newman, which included 100 invasive BC cases from Ghanaian women (Pang et al., 2012).

4.3.2.2. Germline and somatic alterations

In a multigene hereditary cancer panel testing study (Shimelis et al., 2018), two panels with 21 and 17 genes were tested in 8752 and 2148 TNBC patients, respectively. Overall, germline pathogenic variants in *BARD1*, *BRCA1*, *BRCA2*, *PALB2*, and *RAD51D* were associated with high risk (odds ratio [OR] > 5.0) of TNBC and greater than 20% lifetime risk for overall BC and pathogenic variants in *BRIP1*, *RAD51C*, and *TP53* were associated with moderate risk (OR > 2) of TNBC. These pathogenic variants were detected in 12.0% (3.7% non-BRCA1/2) of all TNBC participants. No statistically significant differences were observed in TNBC patients from AA and White patients (Shimelis et al., 2018).

Regarding genome-wide association studies (GWASs), most of them have been carried out in populations mainly composed by White patients. As the Black Women's Health Study matures, it is anticipated that GWASs focusing on African-ancestry women will provide important insights into the genetic basis of TNBC in this population. GWASs of AA women identified a common risk variant at the telomerase reverse transcriptase (TERT)-CLPTM1-like (CLPTM1L) locus on chromosome 5p15 (OR = 1.25; $p < 0.0001$), that was present at a greater frequency in AA women than in White women and was significantly associated with TNBC in women with less than 50 years old (OR = 1.48; $p < 0.0001$) (Haiman et al., 2011). A genetic variant in the *LOC643714* gene has also been identified as predicting a 23% increased risk for BC in AA women but not in White women (OR = 1.23; 95% confidence interval, 1.05-1.44) (Ruiz-Narváez et al., 2010).

Using TCGA's exome sequencing data from 105 AA women and 663 White women, Keenan et. al. (Keenan et al., 2015) showed that BCs from AA patients presented a greater intratumor genetic heterogeneity, having more *TP53* mutations (42.9% vs 27.6%, $p = 0.003$) and fewer *PIK3CA* mutations (20.0% vs 33.9%; $p = 0.008$). Furthermore, AA patients had a higher risk of tumor relapse than White patients (hazard ratio, 2.22; 95% CI, 1.05 to 4.67) (Keenan et al., 2015).

In 2017, Ademuyiwa et al. (Ademuyiwa et al., 2017) reported a mutational analysis of AA and White TNBC patients from TCGA, where tumor somatic mutations and microarray data were analyzed in order to compare the mutational landscape between these populations. Interestingly, they found

no compelling differences in the panorama of mutated genes between AA and White TNBC patients. Also, they found no ancestry-associated differences in high prevalence genes, such as *TP53*, *PI3CA* and *MLL3*, between AA and White TNBC patients, attributed to the fact that there is no significant difference in somatic mutations of these genes. This work is one of the few in which authors argue that the higher frequency of TNBC in AA women is not associated with a different genetic profile of commonly established tumor regulatory pathway genes and that other modifiable factors may exist that contribute to the ancestry-associated disparity in TNBC (Ademuyiwa et al., 2017).

Also in 2017, Huo et al. (Huo et al., 2017) reported the increased frequency of basal-like BC subtypes among AA compared with White cases from TCGA. Also, their TCGA analysis suggested that > 40% of the differences in frequencies of BC subtypes may be explained by inherited germline variants, suggesting that a significant proportion of ancestry-associated differences in subtype frequencies are due to genetic factors. Furthermore, as Keenan et al. (Keenan et al., 2015), authors also reported that AA patients have more *TP53* and fewer *PIK3CA* mutations compared with White patients (52% vs 31%, $p < 0.0001$, and 24% vs 36%, $p = 0.012$, respectively) (Huo et al., 2017).

More recently, to investigate subtype-specific risk of germline alleles associated with TNBC in African-ancestry populations, Newman et al. (Newman et al., 2019), through the International Center for the Study of Breast Cancer Subtypes (ICSBCS) biorepository, used a surgically maintained biospecimen cohort of 2884 BC cases. Different groups of patients, (760 AAs, 962 Whites, 910 West African/Ghanaians, 252 East African/Ethiopians), as well as a subset of 417 healthy controls, were analyzed for genotypes of candidate alleles in order to measure, among the groups of African-ancestry patients, the associations between the overall BC risk and TNBC. Authors observed that TNBC frequency was highest in Ghanaian and AA patients (49% and 44% respectively, $p < 0.0001$) and lowest in Ethiopian and White patients (17% and 24% respectively, $p < 0.0001$). Among the obtained results, authors highlight the high frequency of the Duffy-null allele, rs2814778 or DARC/ACKR1 variant, an African ancestral variant adopted under selective pressure as protection against malaria, which was associated with TNBC-specific risk ($p < 0.0001$) and West African ancestry ($p < 0.0001$) (Newman et al., 2019). Interestingly, previous TCGA analyses revealed that low DARC/ACKR1 expression was an adverse prognostic factor in AA patients (Davis et al., 2015). Likewise, correlations between DARC/ACKR1 mutations specific to African-ancestry patients and cancer-related chemokines were confirmed among separate patient cohorts from Michigan as well as Georgia (Jenkins et al., 2019). Thus, Duffy-null variant of *DARC* is independently associated with TNBC risk in African-ancestry women (Newman et al., 2019).

4.3.2.3. Gene expression

Few studies have provided limited but important findings concerning gene expression profiles of African-ancestry BC patients, mostly composed by USA-based studies.

The first study using TCGA's BC data was developed by TCGA consortium itself. In this study, TCGA group analyzed primary BCs from 627 women, 53 of whom (8.5%) were AA (Koboldt et al., 2012). Comparatively to BCs in White women, BCs from AA women showed increased expression of members of the p53, BRCA1, Aurora A, Aurora B and polo-like kinase (PLK) signaling networks (Koboldt et al., 2012).

In a 2013 study, Lindner et al. compared a total of 136 TNBC tumor samples from White (54%) and AA women (39%) (the remaining 7% were of Hispanic ethnicity), from the Yale TNBC cohort, through gene expression profiling and IHC. They not only verified that the basal-like subtypes were more common in AA cases, as they also reported that the transcriptional profiles from AA women demonstrated a gene expression signature that was consistent with increased loss of *BRCA1* expression, increased activation of *IGF1R* and increased expression of VEGF-activated genes, as compared with the transcriptional profiles of White women (Lindner et al., 2013).

Also published in 2013, and although not being specific to TNBC subtype, Stewart et al. (Stewart et al., 2013) performed a comprehensive differential gene expression analysis, including subtype- and stage-specific analysis, using the BC data from TCGA. Differentially expressed genes between 53 AA and 574 White BC patients increased in number in each stage of tumor progression, being 26 in stage I, 161 in stage II, and 223 in stage III. Resistin, a gene that is linked to obesity, insulin resistance, and BC, was expressed more than four times higher in AA tumors. Also, network analysis showed increased expression of a majority of components in TP53 and BRCA1 subnetworks in AA patients, and members of the aurora B and polo-like kinase signaling pathways were also highly expressed. Authors suggest that the higher gene expression diversity observed in more advanced stages may be due to increased genomic instability during tumor progression.

As the previous study, the work of Keenan et. al (Keenan et al., 2015) was not TNBC specific. Microarray gene expression data, obtained from TCGA, from 159 AA women and 711 White women was assessed in this study. Among TNBC cases, according to TNBCtype classification, AA patients had more BL1 and MSL tumors than White patients. Also, AA patients had a higher risk of tumor relapse than White patients (hazard ratio, 2.22; 95% CI, 1.05 to 4.67). Taken together, these observations suggest a more aggressive biology in AA patients than in White patients.

In 2017, Huo et al. (Huo et al., 2017) also reported on the increased frequency of PAM50 basal subtype tumors among 154 AAs compared with 776 White patients from TCGA, using upper quartile normalized RNA-seq data. 142 differentially expressed genes (FDR < 0.05) between AA patients and White patients were identified by linear regression, adjusting for basal intrinsic subtype (56 AA vs 114 White patients). Among the differentially expressed genes, authors highlight *CRYBB2*. This gene has been consistently found to be overexpressed in BCs from AA comparing with White patients (Field et al., 2012; Grunda et al., 2012; Martin et al., 2009; Sturtz et al., 2014). In fact, *CRYBB2* expression was also higher in AA patients within each BC subtype and in normal breast tissue from AA, indicating that *CRYBB2* overexpression is a ancestry-specific feature and thus, not necessarily involved in BC and TNBC progression or aggressiveness in this population (D'Arcy et al., 2015; Field et al., 2012). Authors also performed Kyoto Encyclopedia of Genes and Genomes (KEGG) gene enrichment analysis of the 142 genes differentially expressed genes, observing an enrichment in ether lipid ($p = 0.023$) and alpha-linolenic acid ($p = 0.049$) metabolic pathways. Such ancestry-associated differential expression in lipid metabolizing genes could be due to quantitative trait loci regulation or may reflect differences in the nutritional status or obesity rates between AA and White women (Dietze et al., 2018; Wang and Beydoun, 2007).

4.3.2.4. Metabolism

Tayyari et al. (Tayyari et al., 2018) performed metabolic analysis of BCs to identify how TNBC differs from luminal A subtype within AA and White patients. Authors used High-Resolution Magic

Angle Spinning (HR-MAS) 1H Nuclear magnetic resonance (NMR) to perform the metabolomic analysis of BC and adjacent normal tissues (n = 82 samples). As expected, TNBC and luminal A subtypes in AA patients exhibited different metabolic profiles, showing also a distinct profile from those revealed in White patients. TNBC in AA women showed reduced ATP levels, as well as profound metabolic alterations characterized by decreased mitochondrial respiration and increased glycolysis. Authors suggest that rapidly proliferating cancer cells undergo metabolic reprograming to meet their high rates of growth and proliferation. Thus, these cells upregulate the glycolytic flux to lactate in the presence of oxygen (i.e., the Warburg effect). In this situation, ATP is preferentially generated through aerobic glycolysis instead of oxidative phosphorylation, which leads to the rapid, yet inefficient, production of ATP per unit of glucose consumed (Zheng, 2012). Furthermore, previous studies also reported that TNBC cells exhibited profound metabolic alterations characterized by a decrease in mitochondrial respiration and increased glycolysis (Pelicano et al., 2014). It is documented that rapid ATP consumption, resulting in low levels of ATP, and its degradation product, adenosine, increased BC cell migration. Also, an adenosine receptor antagonist was found to attenuate the ATP stimulation of tumor cell migration and metastases in vitro and in vivo (Jiang et al., 2015). Hence, these metabolic alterations could potentially be exploited as novel treatment targets for TNBC in AA patients.

4.3.2.5. Cancer stem cell-like cells

Cancer stem cells (CSC), a rare population of cancer cells harboring the potential for self-renewal and differentiation, have been shown to exist in BC. Theoretically, any cell produced along the pathway from the parent mammary stem cell through a multipotent progenitor, to a committed progenitor, and ultimately to a differentiated mammary gland cell can experience some aberrant oncogenic activity and transform into a breast CSC (BCSC). Essentially, differentiated epithelial cells are expected to interact with their microenvironment in a well-regulated fashion, but they will occasionally be triggered to assume mesenchymal properties featuring the capacity to migrate and avoid apoptosis (in the so called EMT process), and plasticity to transition between these patterns. BCSCs represent the subpopulation of tumor cells that can self-renew and recapitulate the parent tumor, while the cancer cells lacking the BCSC properties do not have this ability. The latter non-BCSCs, more proliferative, comprise the bulk of the detected tumor and its metastases, while the minority-subpopulation BCSCs account for the metastatic potential. Conventional CT regimens typically focus on exterminating the hyperproliferative and abundant non-BCSC population. Thus, efforts to identify the quiescent BCSCs and to develop therapies that disrupt the BCSC metastatic progression and cancer relapse represent exciting prospects in precision medicine research (Creighton et al., 2009; Jiagge et al., 2018; Lin et al., 2016; Liu and Wicha, 2010). BCSCs are generally characterized with a CD44+/CD24low and/or ALDH1+ (ALDH1A1+) population (Al-Hajj et al., 2003; Fillmore and Kuperwasser, 2007; Ginestier et al., 2007), which are usually detected by flow cytometry and fluorescence-activated cell sorting. Although the prognostic value of these markers has not been definitively determined, findings from several studies have shown that tumors with higher proportions of CSCs indicated by any of these markers are associated with worse prognosis (Ginestier et al., 2007; Lin et al., 2016; Liu et al., 2014).

Preliminary data suggest that TNBCs are enriched with populations of cells that feature BCSC markers, and this feature may account for the metastatic potential of this aggressive BC subtype. However, very little is known regarding BCSC patterns African-ancestry women, as most stem cell studies in humans have been based on specimens from White and East Asian patient populations

(Jiagge et al., 2018). A study by Nalwoga et al. (Nalwoga et al., 2010) tested for ALDH1 in 192 BCs from African-ancestry women. ALDH1 expression was detected in 88 BC specimens from Uganda (48%) and was associated with p53 (p = 0.034) and basal-type markers (p = 0.008) (Nalwoga et al., 2010), while only 19% and 30% of BCs from non-African-ancestry women in Michigan and France, respectively, overexpressed ALDH1+ population (Ginestier et al., 2007). Schwartz et al. reported overexpression of ALDH1 in 42% and 17% of the stromal and epithelial compartments, respectively, from 104 Ghanaian BC cases. These authors also found notably higher expression of ALDH1 in TNBCs as well as in normal breast tissue from Ghanaian patients. An updated and expanded but unpublished series from this group found ALDH1 expression to be increased among specimens from AA and Ghanaian BC patients (32% and 36%, respectively) compared with White and Ethiopian patients (23% and 17%, respectively, p = 0.007) (Jiagge et al., 2018; Schwartz et al., 2013). Finally, Nakshatri et al., studied the distribution of mammary stem cell phenotype in the healthy breast tissue samples from AA and White women from the Susan G Komen Tissue Bank (Indianapolis, USA) and found that CD44+/CD24low multipotent stem cells were significantly more frequent in AA in comparison to White patients (Nakshatri et al., 2015).

In addition to these markers, a number of studies suggest the activation of Wnt signaling pathway in TNBC (Barker and Clevers, 2006). Specifically, upregulation of Wnt signaling leads to increased stemness in TNBC cells (Ibrahim et al., 2013). A study conducted by Getz et al. revealed that Wnt signaling pathway represents a major functional pathway in a cohort of AA patients with TNBC and that genes and proteins associated with Wnt signaling pathway, like TNC, Cav1, FOXO3A and TCF4, are significantly induced in this subset of patients (Getz et al., 2015). Also, co-activation of Wnt and Hedgehog signaling in TNBC samples correlated with shorter recurrence and poor survival (Arnold et al., 2017).

Overall, these studies show that the abundance of CSCs can potentially explain the inherently aggressive biology of TNBCs in African-ancestry women, particularly regarding treatment resistance and faster relapse. Nonetheless, although this notion is gaining traction and is supported by multiple studies, additional population-based studies are required to prove this concept.

4.3.2.6. Immune microenvironment

The degree of immune infiltration in the tumor microenvironment is a consistently reported prognostic and CT treatment response marker in TNBC. Both a high TIL count and high immune-related gene expression predict higher pCR rates and better prognosis (Section 3.2.6.) (Adams et al., 2014; Bianchini et al., 2010; Denkert et al., 2015; Loi et al., 2019).

In order to test for immunological differences between AA and White patients, using RNA-seq data from TNBC patients from TCGA, immune metagene analysis demonstrated marginal immune attenuation in AA patients relatively to their White counterparts, not reaching statistical significance, with no consistent ancestry-associated differences in immune gene expression or TIL counts in TNBC according to patient's ancestry (O'Meara et al., 2019). Nonetheless, such study is not able to rule out small differences in immune cell subtype distribution and activity status that may not be apparent in bulk RNA analysis.

Additionally, Mremi et al. examined cases of BC in Tanzanian (n = 83), AA (n = 120), and White women (n = 120). Not only BC from the Tanzanian and AA groups were more likely to be high grade (p

= 0.008), to have a high mitotic rate ($p < 0.0001$), and to be ER-negative ($p < 0.001$), as higher levels of TILs were observed in Tanzanian and AA patients compared to White patients ($p = 0.0001$). Furthermore, interestingly, among all subjects TIL levels were higher in tumors with a high mitotic rate and ER- cases (Mremi et al., 2019).

4.4. Future directions

Despite significant improvements in BC therapeutics, TNBC remains a challenge owing to aggressive progression and lack of targeted therapies. In recent years, it has become evident that younger African-ancestry women are disproportionately affected by TNBC. Although many epidemiological and clinical studies have shown a higher prevalence of TNBC in this population and put forth various genetic and environmental factors that may influence aggressive progression of TNBC, we still lack studies focusing entirely on ancestry-associated disparity in TNBC with emphasis on underprivileged African-ancestry women.

Future challenges include defining the molecular drivers and signaling pathways that promote TNBC incidence and development in African-ancestry women. Figure 12 displays a summary of the most comprehensive aspects involving aggressive behavior of BC and TNBC in African-ancestry women. Despite the efforts in exploring the biology of TNBC development and aggressiveness in African-ancestry women in the past decade, this subject needs to be deepened urgently. Also, it will be important to distinguish whether women of African-ancestry may have a population-specific TNBC subtype, with unique biological factors that promote a more aggressive biology when comparing with other populations. Also, defining the molecular link between environmental exposures, socioeconomic disparities and TNBC in African-ancestry patients, both from higher income and from African countries, will be a complex undertaking.

Multiple new combinational or sequential therapies are currently under investigation in prospective clinical trials to counter TNBC, based on their unique biology in this era of precision medicine studies. Even with considerable advancement in understanding the unique biology of TNBC in African-ancestry women, owing to epidemiological, clinical and preclinical studies, very few clinical trials have undertaken the task to directly compare the efficacy of therapeutic agents in African-ancestry vs White TNBC patients. In fact, only one such trial is being conducted (ClinicalTrials.gov identifier: NCT03057600). This trial consisted in a phase 2 study of the combination of CB-839 with paclitaxel in AA patients and non-AA patients with advanced TNBC. On February 8, 2021, results were not yet available.

TNBC disparity in African-ancestry and White women stems from a complex interaction of socioeconomic factors and biology. In conclusion, improving survival in African-ancestry TNBC cohort would require a transdisciplinary approach involving a cohesive interplay encompassing biology, genetics, and socioeconomic factors.

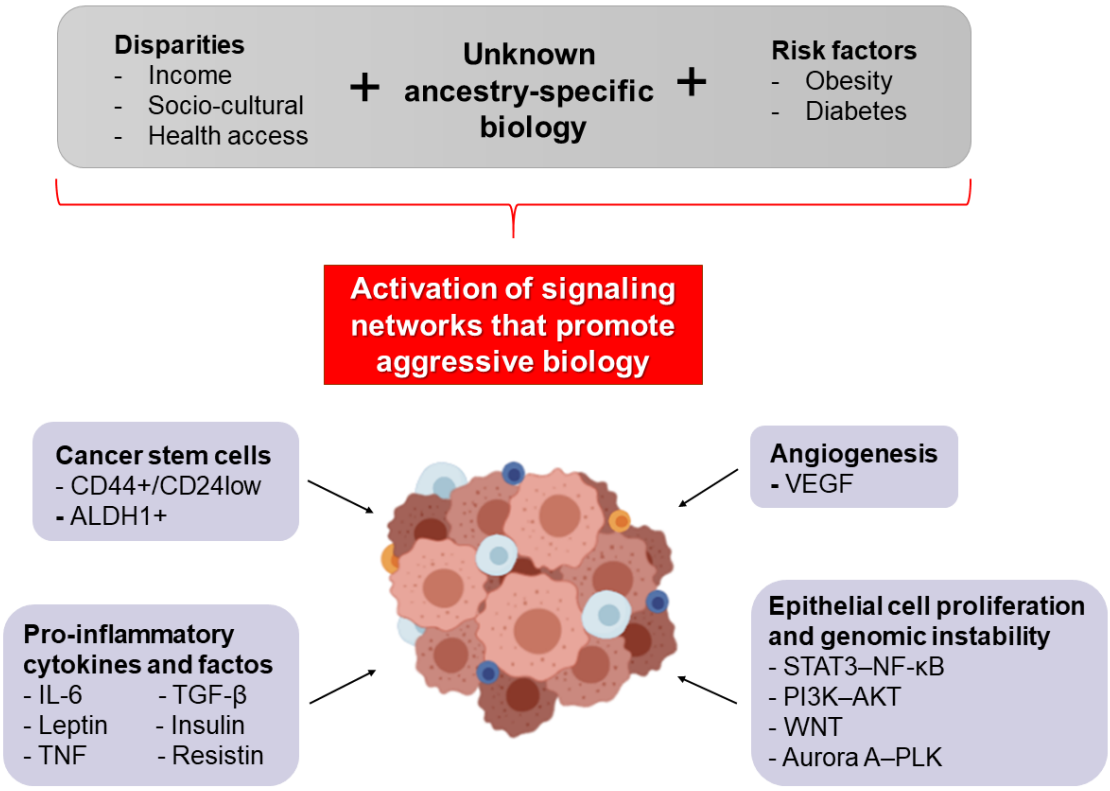


Figure 12 Comprehensive summary of the current known factors involved in the activation of signaling networks that promote aggressive biology of breast cancer and TNBC in African-ancestry women. Such factors include socioeconomic disparities, co-morbidities, such as obesity and diabetes, and unknown ancestry-specific biology. Aggressive development of breast cancer and TNBC in African-ancestry women is suggested to be caused by an increased population of cancer stem cells, an inflammatory environment, increased angiogenesis, epithelial cell proliferation and genomic instability.

Chapter II

Rationale and specific aims

Breast cancer (BC) is the most common cancer worldwide and the fifth most deadly (Sung et al., 2021). BCs form a heterogeneous group of tumors, which are characterized by a wide range of clinical, pathologic, and molecular features. This wide spectrum of factors accounts for variations in response to therapy and outcomes among patients. The expression of estrogen receptor (ER), progesterone receptor (PR) and Human Epidermal Growth Factor Receptor 2 (HER2), the current conventional BC biomarkers which are assessed through immunohistochemistry (IHC), is a major determinant of patients prognosis and therapy (Henry et al., 2020). BCs presenting hormone receptor expression (ER and PR) or HER2 expression or amplification have specific treatments, hence the better prognosis of such patients. Nonetheless, triple-negative breast cancer (TNBC), lacking the conventional molecular targets ER, PR, and HER2 expression, is "negatively defined". Being a remarkably inter- and intra-heterogeneous BC subtype and due to the absence of specific treatments, TNBC is considered one of the most relevant unmet medical problems in BC care and women with this BC subtype have worse outcome compared to other BC patients (Masuda and Masuda, 2016; Schmadeka et al., 2014). Epidemiologically, TNBC is estimated to account for 10-20% of BC cases worldwide (P Boyle, 2012). However, there is evidence that TNBC prevalence is higher in African-ancestry women, particularly in younger patients (Carey et al., 2006; Huo et al., 2017; Lund et al., 2009). Furthermore, African-ancestry patients present a faster and more aggressive clinical course of TNBC, independent of other risk factors for BC (Stead et al., 2009) and having into account socioeconomic factors when comparing with other ethnicities (Keenan et al., 2015). Thus, it is crucial to identify molecular regulators and altered signaling pathways that are potentially involved in the higher incidence and development of aggressive TNBC in African-ancestry women. Such knowledge may lead to the development of models that specifically capture the risk of African-ancestry women for TNBC, increase the access to effective early detection and leading to the implementation of health policies that will improve the outcome and life quality of these patients

Thus, this PhD thesis aims to contribute to unveiling ancestry-associated molecular regulators and signaling pathways involved in the observed discrepancies regarding the higher incidence and aggressiveness of TNBC in African-ancestry women. Specifically, we intended to:

1) Identify differentially expressed genes in TNBC cases from African-ancestry patients, comparing with matching White patients.

Here we performed differential gene expression analysis (DGEA), using RNA-sequencing (RNA-seq) data available in the public American data base The Cancer Genome Atlas (TCGA), between TNBCs from African-ancestry and White patients.

DGEA was performed comparing all TNBC patients from both ancestries. Furthermore, given the reportedly premenopausal status and younger age of incidence of TNBC in African-ancestry patients (Carey et al., 2006; Huo et al., 2009; Lund et al., 2009; Zaky et al., 2009), and the fact that the expression profile between African-ancestry and White patients is described as being more distant throughout the increased staging of BC disease (Stewart et al., 2013), we also grouped TNBC patients to normalize the DGEA according to menopause status and/or pathological stage of the disease.

Additionally, we also performed DGEA using RNA-seq data from normal breast tissue between African-ancestry and White cases, in order to exclude differentially expressed genes (DEG) that are normally associated with patient's ancestry. Also, we also compared the expression profile of TNBCs

and normal breast tissues from African-ancestry women, to identify gene expression alterations that the breast tissue acquires in order to develop TNBC in this group of patients.

2) Identification of differential signaling pathways and key molecular regulators in African-ancestry patients with TNBC

To shed some light on interesting altered signaling pathways and molecular mechanisms, as well as to identify key molecules involved in the regulation of altered networks, with potential prognostic and/or therapeutic impact specific of TNBC in African-ancestry patients comparing with White patients, the following analyses were performed using as input the complete gene lists obtained from DGEA results:

- Gene set enrichment analysis (GSEA) was performed using collections of diseases and pathways from Kyoto Encyclopedia of Genes and Genomes (KEGG), a curated database of molecular pathways and disease signatures (Kanehisa et al., 2017, 2016; Kanehisa and Goto, 2000), in order to identify which signaling pathways and cellular mechanisms are more altered in TNBCs from African-ancestry patients;
- Ingenuity Pathway Analysis (IPA)'s core analysis, in which we focused on the network design to aid visualize in what way the DEG interact and which genes have central roles in the most altered networks identified in African-ancestry patients diagnosed with TNBC.

Together with DGEA results, the essential clues given by GSEA and IPA's network design contributed to the selection of candidates for further validation in TNBC samples from African-ancestry patients.

3) Validation of candidate genes in TNBC samples

To validate the selected candidate genes in TNBC samples, we stablished a collaboration with Hospital Professor Doutor Fernando Fonseca (HFF), a suburban hospital in Lisbon and whose catchment area has one of the higher percentages of migrants from the former Portuguese colonies in Africa. Specifically, we obtained a collection of formalin-fixed paraffin-embed (FFPE) TNBC tissue, from both African-ancestry and White patients, as well as the respective clinicopathological data.

Selected candidate genes were then validated through protein expression analysis performed by IHC. Protein expression was quantified and compared between the two groups of patients, as well as with its respective gene expression obtained from the DGEA with TCGA patients.

Altogether, we identified potential ancestry-associated candidates in TNBCs from African-ancestry patients, as will be shown and discussed in Chapters IV and V, which may have important implications in improving the prognosis, treatment and overall life quality of African-ancestry patients diagnosed with TNBC.

Chapter III

Material & Methods

In order to find specific differentially expressed molecules in African-ancestry patients with TNBC, comparing with matching White patients, that may justify the aggressiveness and early onset of this disease in the African-ancestry women, this work will be divided in the following two main analysis (Figure 1):

- Computational analysis, in which the main study involved data from RNA-sequencing (RNA-seq) and respective clinical information from 52 African-American (AA) and 90 White patients with TNBC. To ascertain if the observed findings are not due to normal differences associated with patient's ancestry, we also analyzed the RNA-seq data from normal tissue adjacent to breast cancer from 6 AA patients and 105 White patients. The data was retrieved from the American data base The Cancer Genome Atlas (TCGA). Patients were diagnosed between 1998 and 2013, and clinical information was submitted to the TCGA between August 2010 and March 2015.

- Validation of results using clinical samples. This analysis included samples of formalin-fixed paraffin embed (FFPE) TNBC tissue from 12 African-ancestry patients and 11 White patients, diagnosed between January 2012 and August 2018, collected in a Portuguese hospital.

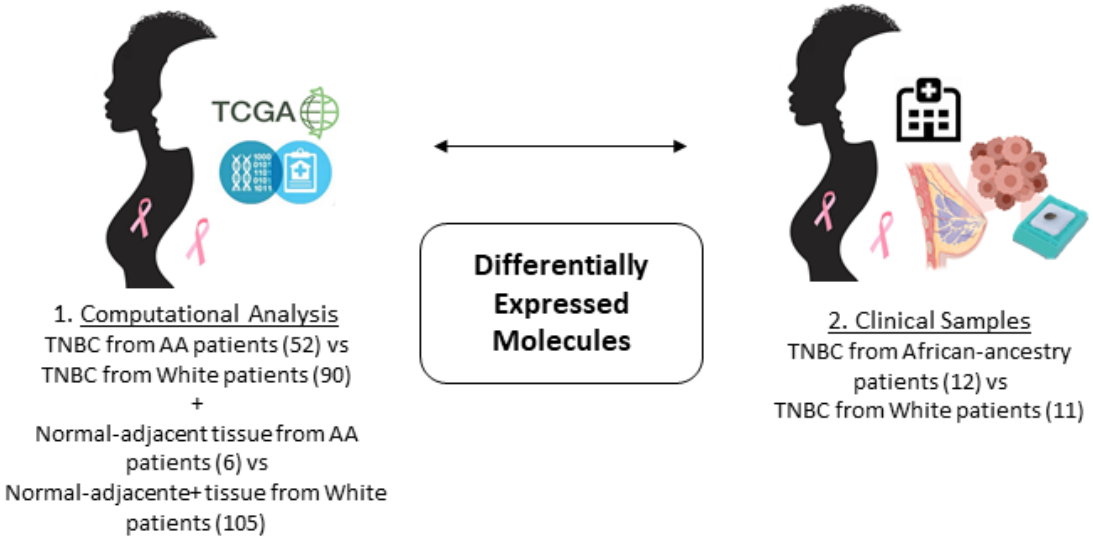


Figure 13 Simplified design of the methodology of the project, involving two main analysis from two distinct groups of patients, in order to identify differentially expressed molecules specific of TNBC of African-origin patients comparing with White patients. 1. Computational analysis – includes RNA-sequencing and clinic-pathological information of 52 AA patients and 90 White patients with TNBC retrieved from TCGA. Additionally, to ascertain if the findings observed in the group of TNBC patients are not due to normal differences associated with patient's ancestry, we also analyzed RNA-sequencing data from normal tissue adjacent to breast cancer of 6 AA patients and 105 White patients. 2. Clinical samples – includes samples from formalin-fixed paraffin embed TNBC tissue from 12 African-origin patients and 11 White patients. TCGA The Cancer Genome Atlas. AA African-American.

1. COMPUTATIONAL ANALYSIS

Breast cancer (BC) data, including patients' clinic and transcriptomic information, was collected from The Cancer Genome Atlas (TCGA, https://www.cancer.gov/tcga) and used for the computational analysis.

R version 3.6.0 was used for all the bioinformatic analysis. *TCGAbiolinks* (Colaprico et al., 2016; Mounir et al., 2019), a package from Bioconductor (Silva et al., 2016), was used to query and download harmonized data from TCGA (as of March 6, 2019). Harmonized database consists in all available data being harmonized against GRCh38 (hg38) (Schneider et al., 2017) using Genomic Data Commons (GDC) Bioinformatics Pipelines, which provide methods to the standardization of biospecimen and clinical data.

1.1. The Cancer Genome Atlas

TCGA is a public funded project that aims to catalogue and discover major cancer-causing genomic alterations, being a comprehensive "atlas" of cancer genomic profiles. TCGA molecularly characterized over 20 000 primary cancer and matched normal samples spanning 33 human cancer types through large-scale genome sequencing and integrated multi-dimensional analyses. Patients' genomic profiles include mutations, copy number variations, gene expression, DNA methylation and proteins profiling, as well as pathology slides and full clinical information (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga, last consulted on March 20, 2021) (Tomczak et al., 2015). To our best knowledge, there is currently no equivalent to this resource with so many "omics" data for so many patients and with so much clinical information.

TCGA Research Network involves several cooperating centers responsible for the collection and sample processing, followed by high-throughput sequencing and sophisticated bioinformatics data analyses. Firstly, biospecimens such as blood or tissue are collected from eligible cancer patients in different USA-based Tissue Source Sites (TSSs) which then deliver the samples to the Biospecimen Core Resource, which is the laboratory and biorepository responsible for cataloging, process, and verify the quality and quantity of the samples. Then, the Biospecimen Core Resource lab submits the clinical data and metadata to the Data Coordinating Center, which standardizes data formats, validates submitted data and provides the molecular analytes to Genome Characterization Centers and to Genome Sequencing Centers for further genomic characterization and high-throughput sequencing. The sequence-related data are then deposited in the Data Coordinating Center and the generated genomic data is made available to the research community (Tomczak et al., 2015).

Remarkably, TCGA data is completely free and is available in the National Cancer Institute (NCI) research program Genomic Data Commons (GDC) public data repository (https://portal.gdc.cancer.gov/) (Grossman et al., 2016). GDC is an oncology data sharing platform in which the different data sets and genomic information can be submitted by different institutes and entities in the USA and harmonized so that the data can be directly compared.

Systematic advances in cancer genomics provided by TCGA have revealed a new comprehensive picture of the molecular biology of cancer. The application of sophisticated high-throughput technology together with well-developed bioinformatics tools has contributed to highlight the similarities and differences in the genomic architecture of each cancer and across multiple types. This huge source of knowledge about cancer genetic and epigenetic profiles, has allowed the highlight of candidate cancer biomarkers and drug targets, with impact in patients prognosis and life quality.

1.2. RNA-sequencing and clinical data

1.2.1. RNA-sequencing data

TCGA data, including RNA-seq data, must be "harmonized" according to the current human reference genome (hg38). Essentially, ''data harmonization'' is the process of bringing together data from multiples sources and ensuring uniform and consistent processes, including mapping the data to a common data model, such as the current human reference genome (hg38) (Lee et al., 2018). This process is performed by GDC's bioinformatic team. RNA-seq data is submitted by TCGA Data Coordinating Center to GDC in BAM or FASTQ format. Briefly, the GDC pipeline regarding the RNA-seq alignment workflow consists in quantifying gene expression based on the number of reads that align to each gene, in which a "two-pass" sequencing method is used. Firstly, the RNA-seq reads are aligned to the current human reference genome (hg38) to detect splice junctions and, secondly, a second alignment is made using the information from splice junctions to increase the quality of the alignment. Finally, read counts are measured on a gene level using the Python package *HTSeq* (Anders et al., 2015) (https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/, last consulted on March 20, 2021), and the "harmonized" data is made available for download.

RNA-seq data from "Primary solid Tumor" was retrieved from "TCGA-BRCA" project, the BC project, being the data category "Transcriptome Profiling" and the data type and workflow type "Gene Expression Quantification" and "HTSeq - Counts", respectively.

```
> library(TCGAbiolinks)
> Query <- GDCquery(project = "TCGA-BRCA",
                    data.category = "Transcriptome Profiling",
                    data.type = "Gene Expression Quantification",
                    workflow.type = "HTSeq - Counts",
                    sample.type = "Primary solid Tumor")
> GDCdownload(Query)
> Expression <- GDCprepare(Query, summarizedExperiment = FALSE)
```

The RNA-seq data, in a matrix format, consisted in 1102 patients' samples and 60483 genomic features with Ensembl (Hunt et al., 2018) identifier. Thereafter, the RNA-seq data matrix will be referred as the Expression matrix.

1.1.2. Clinical data

"TCGA-BRCA" project's "Clinical" data category includes the list of patients and their respective metadata, such as patient's ancestry/race, hormone receptor status, age at initial diagnosis, menopause status, pathological stage of the disease and patients' follow up.

```
> Clinical <- GDCquery(project = "TCGA-BRCA",
                       data.category = "Clinical",
                       file.type = "xml")
```

```
> GDCdownload(Clinical)
> Patients <- GDCprepare_clinic(Clinical,  clinical.info = "patient")
```

Patients were separated by ancestry, “Black or African American” or “White”, and according to hormone receptor status and HER2 amplification. Patients were considered to have TNBC if they met the following conditions: “Negative” receptor statuses for estrogen and progesterone; “Negative” or “Equivocal” HER2 status determined by immunohistochemistry (IHC); in the case of HER2 status determined by IHC being “Equivocal”, the in situ hybridization must be “Negative”; IHC level of HER2 is not “+3”. Patients were considered hormone receptor positive if they are “Positive” for estrogen receptor or progesterone receptor, and considered HER2 positive if they were “Positive” in IHC or in situ hybridization and “Negative” for estrogen and progesterone receptors.

Patients appearing in the Clinical data matrix in duplicate, i.e. patients with more than one sample, and patients without corresponding RNA-seq data were removed from the analysis.

Other downloaded clinical information includes patients’ follow up, treatments (drugs and radiation), staging and new tumor event information.

```
> Follow_Up <- GDCprepare_clinic(Clinical,  clinical.info = "follow_up")
```

1.1.3. TCGA barcode

The TCGA barcode is the primary identifier of a biospecimen data. The TCGA barcodes in the Expression and Clinical data matrices are different and is important to create matching TCGA barcodes that will serve as a link between both matrices for further analysis.

The barcodes in the Clinical data matrix only include information regarding the TCGA project, the tissue source site and the patient code, whereas the barcodes in the Expression matrix also contain bench working information, such as sample, vial, portion, analyte, plate and center (Figure 14). Thus, we removed part of the TCGA barcodes of the Expression matrix, in order to create a matching patient barcode with the Clinical data matrix.

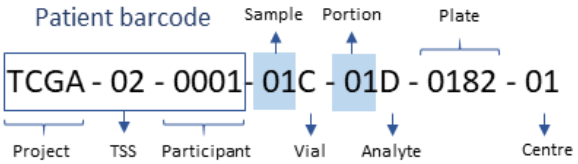


Figure 14 Example of a TCGA barcode. Project - Project name. TSS - Tissue source site. Participant - Study participant. Sample - Sample type, e.g. solid tumor. Vial - Order of sample in a sequence of samples in alphabetical order. Portion - Order of portion in a sequence of 100 - 120 mg sample portions. Analyte - Molecular type of analyte for analysis, e.g. D for DNA sample. Plate - Order of plate in a sequence of 96-well plates. Centre - Sequencing or characterization center that will receive the aliquot for analysis.

1.1.4. Survival curves and differences between the populations

Wilcoxon Rank-Sum (Mann-Whitney U or Wilcoxon-Mann-Whitney) test with continuity correction was used to compare age of diagnosis and months of survival after diagnosis between AA and White TNBC patients. This procedure was performed using *wilcox.test()* function in R environment.

Pearson's Chi-squared test with Yates' continuity correction was used to test for statistically significant differences between the percentage of patients of each ancestry regarding BC subtype, clinical stage and menopause status. These tests were performed in R environment using *chisq.test()* function.

The packages *survival* (Therneau, 1999) and *survminer* (Kassambara et al., 2019) were used to draw and test for statistically significant differences regarding the time of survival after diagnosis between AA and White TNBC patients according to the clinical stage and menopause status.

P-values < 0.05 were considered statistically significant.

1.3. Differential Gene Expression Analysis

Differential gene expression analysis (DGEA) consists in taking the normalized read count data and performing statistical analysis to identify quantitative changes in expression levels between experimental groups. For a given gene, a statistical test is used to decide whether an observed difference in read counts is significant, that is, if it is greater than what would be expected just due to natural random variation. In other words, the null hypothesis is that, for a given gene, the mean expression between two groups is the same.

Although the procedure of choosing a random location in the genome to produce a read is a Poisson process, in the sense that a plot of the depth of the sequence along the genome would have a Poisson distribution, data from real genomes are not so simple. Not only genomes have biases in their composition, such as chemistry of sequencing, assembly errors in the reference genome or inability to map to some repetitive regions, as, in the case of RNA-seq, the reads come from the transcriptome. Since not only some transcripts are present at higher abundance than others, as well as the fact that there is variation between the replicates (different patients, in this case), the Poisson assumption will tend to underestimate the variance and any differentially expressed gene (DEG) will be overstated. For that reason, most procedures for identification of DEG choose a negative binomial model over Poisson since, in real biological applications, there is more variability than the Poisson model can explain.

1.3.1. edgeR package for differential gene expression analysis

edgeR (Robinson et al., 2009a) is a Bioconductor software package that performs differential abundance analysis for pre-defined genomic features. Thereafter, for simplicity, those genomic features will be referred as “genes” and its abundance as gene expression, although they could also be exons, genomic intervals or some other type of feature.

RNA-seq data input for *edgeR* must be in counts format, i.e., without any normalization after the alignment of the raw sequence reads to a reference genome, specifically, hg38. This was previously performed by the GDC team (Section 1.2.1.).

edgeR models RNA-seq count data using a negative binomial model, to account for both biological and technical variability. Also, *edgeR uses empirical* Bayes methods to moderate the degree of overdispersion across genes, allowing the estimation of gene-specific biological variation even for experiments with minimal levels of biological replication (Casella, 1992; Robinson and Smyth, 2007).

edgeR also provides functions to plot results throughout DGEA. This allows for quality control, checking model hypotheses, and track of the whole analysis process.

1.3.1.1. glm edgeR for multigroup experiments

One of the aims of this project is not only to disclose differences in gene expression levels between African-ancestry and White TNBC patients, but also to identify differences according to patient's menopause status/age and pathological stage of the disease. Those differences may not only help in understanding the aggressiveness and early onset of TNBC in African-ancestry patients, as it may also have potential impact in patients' treatment and outcome. For that reason, different groups of patients from both ancestries were compared.

Generalized linear model (GLM) methods *edgeR* (McCarthy et al., 2012), or "glm *edgeR*", uses glm-based statistical methods and it is suitable for multifactor analysis, as it offers more flexibility. On the other hand, "classic *edgeR*" uses exact tests and it is more adequate to two group only comparisons. GLM methods are an extension of classical linear models to non-normally distributed response data (Nelder and Wedderburn, 1972) and specify probability distributions according to their mean-variance relationship.

There are two testing methods under the glm framework: likelihood ratio test and quasi-likelihood F-test. Whereas the likelihood ratio test can be useful in some special cases such as single cell RNA-seq and datasets with no replicates, the quasi-likelihood method, used in the DGEA of this project, is highly recommended for a greater amount of RNA-seq data as it gives stricter error rate control by accounting for the uncertainty in dispersion estimation.

1.3.1.1.1. Contrasts

In the context of a glm *edgeR* analysis, "contrast" is a pairwise comparison between groups. For the study of ancestry-associated gene expression differences between African-American (AA) TNBC patients and White TNBC patients, the following groups of patients were compared: 1) all AA patients vs all White patients; 2) Menopause status-matched comparison, not having into account disease stage (Table 6); 3) Pathological stage-matched comparison, independently of patients' menopause status (Table 7); and 4) Both menopause status and pathological stage-matched comparison (Table 8).

Table 6 Table of contrasts involving groups of patients of different ancestry discriminated by menopause status, independently if pathological stage. BC Breast cancer. AA African-American.

Ancestry	BC	Menopause	# Patients
AA	TNBC	Pre-	9
		Post-	32
White	TNBC	Pre-	28
		Post-	55

Table 7 Table of contrasts involving groups of patients of different ancestry discriminated by pathological stage of the disease, independently of menopause status. BC Breast cancer. AA African-American.

Ancestry	BC	Stage	# Patients
AA	TNBC	I	9
		II	32
		III	10
White	TNBC	I	19
		II	55
		III	11

Table 8 Table of contrasts involving groups of patients of different ancestry discriminated by both menopause status and pathological stage of the disease. BC Breast cancer. AA African-American.

Ancestry	BC	Menopause	Stage	# Patients
AA	TNBC	Pre-	I	2
			II	5
			III	2
AA	TNBC	Post-	I	3
			II	22
			III	7
White	TNBC	Pre-	I	7
			II	17
			III	4
White	TNBC	Post-	I	11
			II	36
			III	6

1.3.1.1.2. Gene identifier

TCGA's genomic features in the Expression matrix are identified through an Ensembl Gene identifier (ID). *edgeR*'s guidelines (Chen et al., 2018) recommend using Entrez Gene IDs for each gene as row names of the input Expression matrix. Therefore, *biomaRt* package (Durinck et al., 2009, 2005) was used to convert human Ensembl IDs into Entrez IDs. Genes without annotation of Entrez Gene ID or in duplicate were removed from the analysis.

```
> library(biomaRt)
> mart <- useMart(biomart = "ENSEMBL_MART_ENSEMBL",
```

```
                    dataset = "hsapiens_gene_ensembl",
                    host = "uswest.ensembl.org")
> ens1 <- as.character(Expression$EnsemblID_column)
[1] "ENSG00000100601.2" "ENSG00000178826.13" "ENSG00000243663"
# Remove the version of the Ensembl IDs
> ens <- gsub("\\.\\d+$", "", ens1)
[1] "ENSG00000100601" "ENSG00000178826" "ENSG00000243663"
> annot <- getBM(mart = mart,
          attributes = c("ensembl_gene_id", "entrezgene_id", "external_gene_name"),
          filter = "ensembl_gene_id", values = ens, uniqueRows = TRUE)
> annot <- data.frame(
          ens[match(annot$ensembl_gene_id, ens)], annot)
> colnames(annot) <- c("original_id",
                    c("ensembl_gene_id", "entrezgene_id", "external_gene_name"))
> Expression1 <- Expression
colnames(expression1)[1] <- "original_id"
Expression <- merge(annot, Expression1, by = "original_id")
```

1.3.1.1.3. DGEList data

The first step to perform a DGEA using glm *edgeR* is to create a list-based data object called a *DGEList.* The *DGEList* object contains 1) the Expression matrix, with the integer counts, in which each row is a gene and each column corresponds to a library/patient and 2) the data frame with the patients' list and variables in study, namely ancestry, menopause status and pathological stage of the disease, where to each group of patients with the same variables is assigned a letter (Table 9). List of patients in both counts and target matrices must be the same and have the same order. For each of the four analyzed groups of patients (section 1.3.1.1.1.) a different *DGEList* was created, keeping the same Expression matrix.

```
> library(edgeR)
> group <- factor(targets$Group)
> DGEList <- DGEList(counts = counts, group = targets$Group)
# Creating the barplot with the library sizes:
> barplot(DGEList$samples$lib.size*1e-6, names=1:122, ylab="Library size (millions)",      xlab =
"Patients")
```

Table 9 Example of the design of a target data frame in a DGEList. BC Breast cancer. AA African-American.

Patient	Ancestry	BC	Menopause	Stage	Group
TCGA-XX-0001	AA	TNBC	Pre-	Stage I	A
TCGA-XX-0002	AA	TNBC	Pre-	Stage II	B
TCGA-XX-0003	AA	TNBC	Pre-	Stage III	C
TCGA-XX-0004	AA	TNBC	Post-	Stage I	D
TCGA-XX-0005	AA	TNBC	Post-	Stage II	E
TCGA-XX-0006	AA	TNBC	Post-	Stage III	F
TCGA-XX-0007	White	TNBC	Pre-	Stage I	G
TCGA-XX-0008	White	TNBC	Pre-	Stage II	H
TCGA-XX-0009	White	TNBC	Pre-	Stage III	I
TCGA-XX-0010	White	TNBC	Post-	Stage I	J
TCGA-XX-0011	White	TNBC	Post-	Stage II	K
TCGA-XX-0012	White	TNBC	Post-	Stage III	L

1.3.1.1.4. Filtering

Genes with very low counts across all libraries were filtered out prior to further analysis. Genes with very low counts across libraries not only provide few evidences for differential expression as they may impair some statistical approximations.

Filtering was performed using counts-per-million (CPM) as unit. As a rule of thumb, edgeR authors consider that a gene should have at least 10 CPM in at least some libraries before it is considered to be expressed in the study (Chen et al., 2016). Also, to ensure that a gene is retained if it is expressed in the libraries belonging to the group with less patients (Chen et al., 2016), which, in this case, are the groups of premenopausal stage I and stage III AA patients (Table 8), with 2 patients each, we applied a filtering threshold of at least 10 CPM in at least two or more libraries to ensure the retention of a gene in the analysis. The library sizes of the *DGEList* were then recalculated.

```
> keep <- rowSums(cpm(DGEListw) > 10) >= 2
> DGEList <- DGEList[keep, , keep.lib.sizes = FALSE]
```

To standardize the procedure, this filtering requirement was used in all DGEA.

1.3.1.1.5. Normalization

edgeR performs differential expression analysis rather than quantification of expression levels. In other words, edgeR focuses on relative changes in expression levels between conditions, but not directly in estimating absolute expression levels. This rationale takes out some weight of the technical influences, since technical factors that are unrelated to the experimental conditions should cancel out of any differential expression analysis. Thus, technical factors such as the number of read counts, which are influenced by gene length, or GC-content of each gene, do not need to be adjusted by edgeR, because gene length has the same relative influence on the read counts in each RNA sample. Therefore, normalization issues arise only to the extent that technical factors have sample-specific effects.

One of such technical factors is the sequencing depth of the RNA samples, which affects the number of read counts. However, *edgeR*'s modelling procedure can adjust any differential expression result that may arise due to different sequencing depths, as represented by differing library sizes, with automatic impact in fold-change (FC) and p-value calculations. This normalization does not require any user intervention.

Another important sample-specific technical factor to be adjusted is the RNA abundance in each sample. RNA-seq measures the relative abundance of each gene in each RNA sample but does not supply any information of the total amount of RNA on a per-cell basis. This is particularly important when few genes are highly expressed in one sample but not in the others. In this situation, the highly expressed genes may consume a considerable proportion of the total library size, causing the remaining genes to be under-sampled and making them appear as falsely downregulated. For that reason, in *edgeR*, normalization takes the form of correction factors that take part of the statistical model. *calcNormFactors()* function was used to normalize RNA composition between each pair of samples. This normalization function uses a Trimmed mean of M-values (TMM) (Robinson and Oshlack, 2010), between each pair of samples, normalizing the RNA composition by finding a set of normalization factors for the library sizes in order to minimize the log-FC between samples for most genes. The TMM normalization is recommended when the majority of the genes are believed not differentially expressed between any pair of samples.

```
> DGEList <- calcNormFactors(DGEList)
```

A normalization factor below one shows that a small number of high counts' genes were monopolizing the sequencing, causing the counts for other genes to be lower than would be usual given that library size. Thus, such library sizes were scaled down, analogous to scaling up the counts in that library. Similarly, a factor above one scales up the library sizes, analogous to downscaling the counts. For example, in Table 10, the group A patients have low normalization factors. This is a sign that these samples contain a number of very highly upregulated genes.

Accordingly, the effective library size, which is the product of the original library size and the normalization factor, replaced the original library size for each patient in all further analyses. Thus, the expected size of each count is the product of the effective library size and the relative abundance of that gene in that patient.

Table 10 Example of normalizing factors generated to each sample by the calcNormFactors function. lib.size library size. norm.factors normalization factors.

Samples	Group	lib.size	norm.factors
TCGA-XX-0001	A	54424136	0.8016286
TCGA-XX-0002	A	44791118	0.9828032
TCGA-XX-0003	B	55903336	1.1615748
TCGA-XX-0004	B	65019567	1.0636690
TCGA-XX-0005	B	70349592	1.0947037
…	…	…	…

1.3.1.1.6. Data Exploration

Differences between groups were explored by generating multi-dimensional scaling (MDS) plots. MDS plots of the RNA-seq samples represent the distances correspond to leading log-fold-changes between each pair of samples. Thus, MDS plots allow the visualization of the differences between the expression profiles of different samples in two dimensions. In fact, this plot can be viewed as a type of unsupervised clustering. Hence, if patients from the same group cluster together in the plot, and patients from different groups form separate clusters, it shows that the differences between groups are larger than those within groups, i.e., differential expression is greater than the variance and can be detected.

```
> group <- targets$Group
> points <- c(15, 16)
> colors <- c("blue", "red"))
> plotMDS(DGEList, col = colors[group],
          pch = points[group],
          xlim = c(-2.5, 4.5), ylim = c(-3, 3))
> legend("topright", legend = c("AA", "White"),
        pch = points, col = colors, ncol = 1, cex = 1)
```

1.3.1.1.7. Design matrix

The glm approach needs a design matrix to describe groups of samples in the analysis. In order to define a coefficient for the expression level of each group, the following arguments were used with the *model.matrix* function:

```
> design <- model.matrix(~ 0 + group, data = DGEList$samples)
> colnames(design) <- levels(group)
```

The 0+ in the model formula is an instruction not to include an intercept column and instead to include a column for each group (Table 11). With this model, each group of patients may be compared with any other different group of patients from the same analysis.

Table 11 Example of the design of a model matrix in the glm edgeR approach. In this example, there are 5 groups of patients and each patient belongs to a different group.

	A	B	C	D	E
TCGA-XX-0001	1	0	0	0	0
TCGA-XX-0002	0	1	0	0	0
TCGA-XX-0003	0	0	1	0	0
TCGA-XX-0004	0	0	0	1	0
TCGA-XX-0005	0	0	0	0	1
…	…	…	…	…	…

1.3.1.1.8. Technical and Biological Coefficient of Variation

Two levels of variation can be distinguished in any RNA-seq experiment. One of them is the measurement error, which is the uncertainty with which the abundance of each gene in each sample is estimated by the sequencing technology. The other is the relative abundance of each gene, which will vary between RNA samples due to biological causes. With these two main sources of variation, technical and biological, there are two coefficients of variation (CV).

Technical CV (TCV) is denoted as (the standard deviation divided by the mean of) the fraction of all cDNA fragments generated from each gene from each sample from the same group. Biological CV (BCV) is the CV with which the unknown true abundance of a gene varies between RNA samples from the same group. If sequencing dept could be increased indefinitely, BCV would remain between replicates, while the TCV would decrease since the size of the counts would increase. As consequence, BCV is the main source of uncertainty for high-count genes. Authors assume that, although *edgeR*'s pipeline focus on BCV, strictly speaking, *edgeR*'s BCV is not just covering for the true biological variation between samples as, ultimately, also includes other sources of the inter-library variation, namely technical issues such as library assemble (Chen et al., 2018).

edgeR authors also recognize that the magnitude of BCV is more important than the exact probabilistic law followed by the true gene abundances. However, for mathematical convenience, *edgeR* assumes that the true gene abundances follow a gamma distributional law between RNA samples from the same group. This implies that the read counts follow a negative binomial probability law, as previously mentioned (Section 1.3.)(Chen et al., 2018).

BCV is the square root of the dispersion parameter under the negative binomial model. All genes have the same mean-variance relationship when considering the parallel nature of sequencing data, including the ensemble of genes. In other words, under this more simplistic model, the dispersion is the same for all the genes and it is defined as "common dispersion" (Chen et al., 2018; Robinson and Smyth, 2008). Thus, since the common dispersion is the "squared BCV", the higher the estimate of the common dispersion, the higher will be the variability between replicates. This is what happens in a DGEA of human samples, which have a huge genetic variability compared to other animal models and cell lines.

However, gene expression levels have a dependent and non-identical distribution between genes. Thus, *edgeR* also includes a more general approach which allows gene-wise variance functions with empirical Bayes moderation (Robinson and Smyth, 2007). With the gene-wise dispersion, the more consistent genes between samples of the same group will be ranked more highly than genes that are not. Thus, the DGEA will not be driven by outliers.

Particularly, in the case of *glm edgeR,* which is the most adequate approach for multi-factor testing (Section 1.3.1.1.), *edgeR* uses the Cox-Reid profile-adjusted likelihood (CR)(Cox and Reid, 1987) method to estimate dispersions. The CR method deals with multiple factors by fitting GLMs within the design matrix (Section 1.3.1.1.6.). All systematic sources of variation are accounted for, allowing a valid estimation of the dispersion. The CR method can be used to calculate a common dispersion for all the genes, trended dispersion depending on the gene expression abundance, or separate dispersions for individual genes.

The common dispersion for all genes and the trended and gene-wise dispersion for each gene were estimated in one run, with *estimateDisp()* function. In this project, we focused on the gene-wise dispersion approach, since it is strongly recommended in multi-factor experiment cases (Chen et al., 2018).

```
> DGEList <- estimateDisp(DGEList, design, robust = TRUE)
> plotBCV(DGEList)
```

1.3.1.1.9. Testing for differentially expressed genes

After estimating the dispersion, the next step was to fit the negative binomial GLM, before proceeding with the DGEA. Given raw counts, negative binomial dispersion(s) and a design matrix, *glmQLFit()* fitted the negative binomial GLM for each gene and produced an object of class *DGEGLM* with some new components.

```
> fit <- glmQLFit(DGEList, design, robust = TRUE)
> plotQLDisp(fit)
```

Quasi-likelihood F (QLF)-test for determining differential expression is recommended in the context of glm framework, since it reflects the uncertainty in estimating the dispersion for each gene. It supplies more robust and reliable error rate control when the number of samples in a group is small.

QLF-test identifies differential expression based on statistical significance regardless of how small the difference might be and, although a combination of choosing a p-value and FC cut-off are customary, it is an *ad hoc* procedure that may be favoring lowly expressed but highly variable genes and destroying the control of false discovery rate (FDR). For that reason, instead of *glmQLFTest(), glmTreat()* function was used to test for differential expression. This approach is similar to t-tests relative to a threshold method (TREAT) (Mccarthy and Smyth, 2009). Under the *glm edgeR* framework, this function implements a statistical test for thresholded hypotheses through a given FC (or $\log_2$FC). As an argument, a $\log_2(1.2)$ was used for all DGEA. The DGEA results can be visualized using an MD plot, in which the log-FC change for each gene is plotted against the average abundance, i.e., logCPM.

```
> contrast <- makeContrasts(A - B, levels = design)
> treat <- glmTreat(fit, contrast = contrast, lfc = log2(1.2))
> plotMD(treat, cex = 0.5, main ="AA TNBC vs White TNBC")
> abline(h = c(-1, 0, 1), lty = 2)
```

This FC threshold, called in the *glmTreat()* function, is not the minimum value of FC expected to see in results. Instead, genes will need to exceed this threshold before being considered statistically significant. Hence, this threshold should be considered as “the FC below which the gene is not interesting" rather than “the FC above which the gene is interesting".

For each contrast (Section 1.3.1.1.1.), genes were then ranked in order of evidence for differential expression, based on the FDR estimated for each gene.

1.3.1.1.10. Converting Entrez Gene IDs in gene symbols

The lists of differentially expressed genes have their identifiers in Entrez Gene ID format, since that was the format used in the Expression matrix. However, since it is routinely more intuitive to identify genes through a gene symbol, the Entrez Gene IDs were converted using the Bioconductor package *org.Hs.eg.db* (Carlson M, 2019), which contains genome wide annotation for Human genes, primarily based on mapping using Entrez Gene IDs.

1.3.1.1.10. Converting Entrez Gene IDs in gene symbols

The lists of differentially expressed genes have their identifiers in Entrez Gene ID format, since that was the format used in the Expression matrix. However, since it is routinely more intuitive to identify genes through a gene symbol, the Entrez Gene IDs were converted using the Bioconductor package *org.Hs.eg.db* (Carlson M, 2019), which contains genome wide annotation for Human genes, primarily based on mapping using Entrez Gene IDs.

1.3.1.1.11. Heat maps

Heat maps were design using *heatmap.2()* function from *gplots* package (Warnes et al., 2020). Heat maps help displaying the expression pattern of the genes across all patients either in a specific order of genes or through gene or patient clustering. This procedure was performed only for contrasts with 10 or more significantly DEG.

1.3.1.1.12. UpSet plot

UpSetR package (Conway et al., 2017) was used to help identifying the number of DEG in common between different contrasts and to design an UpSet plot representing those results. Essentially, the UpSet plot provides a way to visualize intersections of multiple sets compared to the traditional approaches, such as the Venn Diagram.

1.3.1.1.13. Differential gene expression analysis using RNA-sequencing data from normal breast tissue

To understand if the ancestry-associated differences observed regarding gene expression in TNBC are due to the disease or due to expected populational differences, glm *edgeR* analysis was also performed using RNA-seq data from normal-adjacent breast cancer tissue from TCGA patients. This analysis involved 6 RNA-seq samples from AA patients and 105 RNA-seq samples from White patients. The analysis was performed using the same parameters as in with the TNBC patients (Section 1.3.1.1). Information about the normal-adjacent breast cancer cohort can be consulted in Supplementary - Patients eTable 2.

Furthermore, a *glm edgeR* analysis was also performed to compare RNA-seq data from normal-adjacent breast cancer tissues of AA patients (6) and the RNA-seq data from TNBC of AA patients (52) with the aim of finding potential DEG involved in TNBC progression in African-ancestry patients.

1.4. Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) was used to interpret DGEA results in a biological context. Specifically, GSEA was performed to identify differential signaling pathways and molecular and cellular mechanisms between AA and White TNBC patients.

edgeR offers different types of gene set tests for RNA-seq data, where the previously created fitted *DGEList* objects can be directly tested. Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg/) was chosen to help getting biologically meaningful interpretations of the DGEA results. KEGG is a curated resource for understanding the high-level functions and utilities of the biological system, such as molecular pathways, cell mechanisms, organism-ecosystem interactions, disease and health signatures and large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (Kanehisa et al., 2017, 2016; Kanehisa and Goto, 2000). Particularly, KEGG Pathway is a collection of gene sets manually drawn into pathway maps representing our knowledge of the molecular interaction, reaction and relation networks for metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases and drug development.

Each one of the gene sets of the KEGG Pathway database was statistically tested for over-representation in each gene list generated in the DGEA relative to what is expected by chance. These statistical tests take into consideration the number of genes detected in the DGEA (FDR ≤ 0.05), their relative ranking and the number of genes annotated to the collection.

GAGE "Generally Applicable Gene-set Enrichment" Bioconductor package (Luo et al., 2009) was used to bring KEGG gene sets into the R environment. Entrez Gene ID was supplied for each gene as row names of the fitted *DGEList* objects and the species was set as human.

edgeR's *fry()* function was used to perform GSEA. *fry()* is based on the ROAST method (Wu et al., 2010), which attributes p-values using an infinite number of rotations of the residual space in a linear model. fry() is able to perform tests for multiple sets, and includes adjustment for multiple testing. Given a gene set, *fry()* tests whether the majority of the genes in the set are differentially expressed in the contrast of interest.

```
> library(gage)
> human.kegg <- kegg.gsets(species = "human", id.type = "entrez", check.new = TRUE)
> human.gs <- human.kegg$kg.sets
> indices <- ids2indices(human.gs, row.names(fit))
> contrast_AB <- makeContrasts(A_B = A - B, levels = design)
> fry_AB <- fry(DGEList, index = indices, design = design, contrast = contrast_AB)
```

The resulted enriched gene sets for each contrast are then presented in ranked order of FDR, taking into consideration both up- and downregulated genes in the set ("FDR.mixed"). The output results also have indication of the net direction, Up or Down, of the gene set in that contrast. However, more upregulated genes do not necessarily mean that the gene set has a positive net direction, since downregulated genes, although in small number in that contrast, might have a higher FDR in the DGEA, having more impact in the FDR of that gene set.

Gene pleiotropy is expected to influence GSEA. Some gene sets were enriched (FDR ≤ 0.05) even though they are not relevant in the context of finding pathways potentially involved in ancestry-associated differences in TNBC incidence and aggressiveness. For that reason, I focused in the enriched gene sets from specific collections related to pathways that may be altered in cancer or more expressed in women. Such collections included signal transduction, cellular community - eukaryotes and cell motility, cell growth and death, cancer: overview, where the BC gene set was included, as well as the drug resistance - antineoplastic collection, women-specific endocrine system collection gene sets, and gene sets related to immuno oncology, from the immune system collection.

GAGE was used to call the list of genes in each of the enriched gene sets in order to identify which genes were differentially expressed in each of those gene sets. *org.Hs.eg.db* package was used to transform the Entrez Gene IDs of the DEG of each enriched gene set in gene symbols. *ggplot2* (Wickham, 2016) and *gridExtra* (Auguie, 2015) were used to design the barplots showing the DEG of the enriched gene sets according to their log_2FC.

1.5. Ingenuity Pathway Analysis

Ingenuity Pathway Analysis (IPA, QIAGEN Inc., https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/) software was used to perform a core analysis of the sets of DEG. Briefly, IPA is an all-in-one, web-based software application that enables analysis, integration, and understanding of data from gene expression, RNA-seq experiments, miRNA, and SNP microarrays, as well as metabolomics and proteomics. IPA's data analysis and search capabilities help in the biological contextualization and significance of data, including identification of candidate biomarkers in the context of larger biological systems. Also, the software is backed by the Ingenuity Knowledge Base, which is a repository of biological interactions and functional annotations created from millions of individually modeled relationships between proteins, genes, complexes, cells, tissues, metabolites, drugs, and diseases. All information in the Ingenuity Knowledge Base is curated from the published literature, being that every gene interaction is supported by evidence extracted from the underlying publications.

To perform IPA's core analysis, Ingenuity Knowledge Base (Genes only) was used as reference to analyze DGEA results, having into account both direct and indirect relationships between genes. As input, the complete lists of genes and respective FC and FDR obtained from all DGEA with TNBC patients were used. Since the TREAT method was previously imposed in the differential expression test (Chapter III – Section 1.3.1.1.9.), no additional cut-off was applied to FC. However, a threshold of FDR ≤ 0.05 was applied to highlight the DEG in a network. The expression analysis was set as the core analysis type and the experience log ratio as the measurement type. We also set a size constraint of 35 focus molecules per network. In the network analysis, two genes are considered to be connected if there is a path (a line called edge) in the network between them. The gene or gene products are referred to as nodes and the intensity of the node color indicates the degree of up- (red) or down- (green) regulation of a given gene/complex. Nodes are represented with various shapes to distinguish the functional class of the molecule. Labels on the edges describe the nature of the relationship between the nodes and genes (Gopurappilly and Bhonde, 2015).

Highly interconnected networks are more likely to represent significant biological function. Networks are assembled based on gene/molecule connectivity with other gene/molecules and the

more connected a gene/molecule is, the more influence it has in that network. Such genes/molecules are called central or core network regulators. Also, the resulting networks are scored and then sorted based on the score, which is influenced by the number of interactions and FDR and FC of the focus genes/molecules involved (https://chhe.research.ncsu.edu/wordpress/wp-content/uploads/2015/10/IPA-Data-Analysis-training-slides-2016_04.pdf, last consulted on March 20, 2021). Thus, in the context of this project, central regulators from highly scored networks are potential drivers for ancestry-associated biological differences.

2. VALIDATION OF RESULTS USING CLINICAL SAMPLES

To experimentally validate the results obtained through DGEA using a bioinformatic approach, BC samples were collected from patients with TNBC from Hospital Professor Doutor Fernando Fonseca (HFF), also known as Hospital Amadora-Sintra. HFF is a suburban hospital of Lisbon whose catchment area has one of the higher percentages of migrants from the former Portuguese colonies in Africa.

Previously, our group performed a retrospective review of TNBC using data from patients diagnosed or being followed in HFF (Honório et al., 2016). The main aim of that study was to compare clinical and pathological features and survival between African-ancestry and non-African patients. In that cohort, with patients diagnosed with BC between 2005 and 2014, 9.3% had TNBC (144 of all patients diagnosed with BC) and 17 of those patients (12%) were from African-ancestry. Nonetheless, ancestry is a feature which is not always designated in patients' clinical files, especially in older files. Thus, quite possibly, the real number of African-ancestry patients diagnosed with TNBC during that period is higher than 12%. The clinical observations from that study showed that African-ancestry patients with TNBC had a very poor or null response to neo-adjuvant chemotherapy (NACT). The relapse was very fast, within 6 months after finishing NACT, and usually with very aggressive disease pattern, with metastization predominantly to liver and brain. African-ancestry patients had a significantly higher prevalence of initially metastatic disease (41.2% vs 11% in non-Blacks, $p < 0.005$) and a worse outcome (median survival: 15 months vs 62 non-Blacks, $p < 0.005$). Thus, although studies concerning TNBC development and ancestry-associated discrepancies are still scarce in African-ancestry populations, particularly those living in Europe, these previous results obtained with patients from HFF are in line with the claims of most American and African-based studies (Brewster et al., 2014; Danforth, 2013; DeSantis et al., 2019; Keenan et al., 2015; Newman, 2017; Newman et al., 2019; Stark et al., 2010). Information about the validation cohort can be consulted in Supplementary Data – Patients, eTable 3.

2.1. Ethics statement

This study was conducted at CEDOC. This project was approved by both the ethics committee of HFF, from where the samples come from, and NMS|FCM-UNL. Patients' informed consent was gathered. Patient's information includes medical records and personal data. Data was anonymized and stored securely.

2.2. Samples

TNBC patients diagnosed from 2005 to 2014 were select based on the curated data base created by Honório et al. (Honório et al., 2016), where White patients had yet to be identified. TNBC patients diagnosed from January 2015 to August 2018 were selected among all BC patients from HFF. Furthermore, only patients diagnosed or subjected to surgery in HFF were included in this study, since patients which started being followed in HFF but that performed those procedures in other hospitals do not have available samples for this study.

Formalin-fixed paraffin-embed (FFPE) TNBC tissues were prepared and stored in HFF's Pathological Anatomy Service (SAP). FFPE tissues include biopsies, in which part of them are used in the diagnosis, or surgical pieces, removed during the mastectomies. Whenever possible, surgical pieces were preferred over biopsies since not only the tissue slice area is expected to be larger, as a portion of normal-adjacent BC tissue is also removed during surgery. The normal-adjacent breast tissue samples would be used to validate the results obtained with the TNBC samples.

12 African-ancestry patients and 11 White patients with TNBC, diagnosed between January 2012 and August 2018, were included in the validation cohort. Per patient, 6 to 8 TNBC FFPE tissue sections with 2.5 µm thick were sliced with a disposable microtome blade (S-35, Feather) and placed in an adhesion histological slide (Superfrost Plus, Thermo Scientific). Tumoral and normal-adjacent breast tissue were placed in different slides and stored at RT.

2.3. Immunohistochemistry

IHC, a method for demonstrating the presence and location of proteins in tissue sections, was used to validate candidate molecules potentially involved in the ancestry associated discrepancy of TNBC incidence and aggressiveness.

As result of the DGEA, GSEA and IPA, as well as literature review, the selected molecules were the following: EGFR, Myc, Bcl2 and β-Catenin.

Slides were firstly deparaffinized through 10 min incubation at 90°C followed by other 10 min in xylene. Tissue was then rehydrated through an ethanol series at 100%, 96% and 70%, respectively, 2 min each, finishing with the slides in ddH_2O. Since the tissues were fixed in formalin, an antigen retrieval step was crucial due to the formation of methylene bridges during fixation, which cross-link proteins and therefore mask antigenic sites. The retrieval was necessary to break the methylene bridges and expose the antigenic sites to allow the antibodies to bind. Slides were placed in a metal rack and a heat-induced antigen retrieval was performed using a pressure cooker with sodium citrate buffer (10mM Sodium Citrate, 0.05% Tween 20, pH 6.0). 2 min after the cooker reached its full pressure, the hotplate was turned off. As soon as the cooker and buffer's temperature allow the protocol to safely continue, slides were rapidly transferred from the buffer to ddH_2O. It is important that the slides do not dry since that would cause non-specific antibody binding and therefore high background staining.

Since the secondary antibody used was Horseradish Peroxidase (HRP)-conjugated (Peroxidase IgG Fraction Monoclonal Mouse Anti-Rabbit IgG, light chain specific, code 211-032-171, Jackson ImmunoResearch), endogenous peroxidase was blocked through a 5 min incubation of the slides in 3% H_2O_2. If not blocked, endogenous peroxidase in the presence of the HRP-conjugated antibody may lead

to high, non-specific background staining. H_2O_2 was removed and the slides were then washed in ddH_2O for 5 min.

A hydrophobic pen was used to draw a barrier circle around the tissue on each slide. Slides were incubated for 1h in a humidified chamber with blocking solution (0.5% BSA), in order to avoid non-specific antibody binding. After 3 washes, 5 min each, with washing solution (PBS 1x, 0.5% Tween 20, pH 7.0), slides from each patient were incubated for 1h with their respective primary antibodies (Table 7) in a humidified chamber. All antibodies were diluted in blocking solution. Control slides were incubated with blocking solution.

Table 12 List of primary antibodies used in immunohistochemistry assays.

Antibody	Reference	Host	Dilution	Localization	Company
Anti-EGFR [EP38Y]	ab52894	Rabbit	1:100	Cell membrane, cytoplasm and nucleus (Brand et al., 2013; Rong et al., 2019)	abcam
Anti-c-Myc [Y69]	ab32072	Rabbit	1:100	Cytoplasm and nucleus (Gong et al., 2017)	abcam
Anti-Bcl-2 [E17]	ab32124	Rabbit	1:50	Between the internal and the external mitochondrial membranes, endoplasmic reticulum, nuclear outer membrane and nucleus (Popgeorgiev et al., 2018)	abcam
Anti-beta Catenin	ab16051	Rabbit	1:100	Cell membrane, cytoplasm and nucleus (Chen et al., 2014)	abcam

After other 3 times cycle of slide washing, slides were incubated for 1h with the secondary antibody in a humidified chamber. After a last 3 times washing cycle, slides were quickly washed with ddH_2O to remove eventual residues of Tween 20, which may impair even dispersal and uniform infiltration of DAB (Bright-DAB Kit, VWR). DAB (3,3'-Diaminobenzidine) is used in IHC staining as a chromogen. DAB is oxidized by hydrogen peroxide in a reaction catalyzed by HRP (conjugated with the secondary antibody). The oxidized DAB forms a brown precipitate, at the location of the HRP, which

can be visualized using light microscopy. Slides were incubated with DAB in a humidified chamber for 8 min and rapidly washed in running water.

Slides were then incubated for 1 min with Mayer's Hematoxylin to counterstain. After washing excess Hematoxylin, tissues in the slides were dehydrated in a series of 70%, 96% and 100% ethanol, 30 seconds incubation each, and then 3 min incubation in xylene. Lastly, slides were mounted using Entellan mounting medium and covered with microscope cover glasses.

2.4. Imaging analysis and quantification

Slides were observed in bright field in an Axio Imager Z2 microscope (Zeiss). Each slide had 4-10 pictures of areas with epithelial cells acquired with a 10x objective and ZEN 2.1 microscope software (Zeiss).

Fiji ImageJ (Schindelin et al., 2012) software was used for image analyses and processing. From the total pictures taken for each studied protein in each patient, converted in RGB format, a minimum of 7 crops with 132 μm^2 each were quantified. That specific area and the 7 crops, which made up a total quantified area of 924 μm^2 per patient, were stipulated based on the maximum area of epithelial cells of the patient with both the most damaged tissue and less area of epithelial cells. Thus, the quantified area for each protein in each patient was the same. Each crop was subjected to color deconvolution, where “H DAB” stain (Hematoxilin + DAB) was chosen as vector. DAB intensity was then quantified in its corresponding channel generated by color deconvolution.

Protein expression was evaluated through the semiquantitative approach used to assign an H-score (or “histo” score) (Ishibashi et al., 2003; John et al., 2009). First, an H-score was calculated for the area of cells in a crop corresponding to an intensity level. Low positive cells have a 1+ score, positive cells have 2+, and high positive cells have 3+. The sum of individual H-scores for each intensity level was performed for each crop, according to the following formula:

H-score = [1 × (% area 1+) + 2 × (% area 2+) + 3 × (% area 3+)]

Based in the threshold method for H-score used by the IHC profiler plugin (Varghese et al., 2014), using the threshold feature of Image menu of the ImageJ program, the thresholds stablished for each pixel intensity level were the following: 0-55 level 3+ or high positive, 56-110 level 2+ or positive and 111-165 level 1+ or low positive area. Pixels with an intensity equal or above the 166 threshold were considered as background staining and were not included in the H-score. The area for each threshold interval was then measured and included in the H-score calculus for each crop.

Mean H-score for each protein of each patient was estimated on basis of the 7 H-scores. Wilcoxon Rank-Sum (Mann-Whitney U) test was used to compare H-score of each protein between African-ancestry and White patients. This procedure was performed using *wilcox.test()* function in R environment and GraphPad Prism 6.01 software. p-values ≤ 0.05 were considered statistically significant.

Chapter IV

Results

1. CHARACTERIZATION OF TCGA PATIENTS

Metadata from African-American (AA) and White patients diagnosed with Triple-Negative Breast Cancer (TNBC), including their respective RNA-sequencing (RNA-seq) data and clinical information, retrieved on March 6, 2019 from The Cancer Genome Atlas (TCGA), was used for the computational analyses. 142 TNBC patients with clinical information were identified, of which 52 (36.62%) are AA and 90 (63.38%) are White. Table 13 describes TCGA's AA and White breast cancer (BC) patients

Table 13 Description of TCGA's African-American and White breast cancer patients.[a]

	African- American (n = 183)	White (n = 757)	p-value
Age at **BC** diagnosis, mean (IQR)	56.29 (47 - 67)	58.61 (49 - 67)	0.037[b]
Tumor subtype, n (%)			
TNBC	52 (28.42)	90 (11.89)	< 0.0001
HER2	7 (3.83)	22 (2.91)	0.684
Hormone receptor	112 (61.20)	589 (77.81)	< 0.0001
Indeterminate or not evaluated	12 (6.56)	56 (7.40)	
Age at **TNBC** diagnosis, mean (IQR)	55.87 (48 - 62.75)	54.23 (46.25 - 62)	0.780[b]
Stage at **TNBC** diagnosis, n (%)			
I	9 (17.31)	19 (21.11)	0.741
II	32 (61.54)	56 (62.22)	> 0.999
III	10 (19.23)	11 (12.22)	0.375
IV	1 (1.92)	1 (1.11)	NA
Indeterminate or not evaluated	-	3 (3.33)	NA
Menopause status at **TNBC** diagnosis, n (%)			
Pre (<6 months since LMP, no prior bilateral ovariectomy and not on estrogen replacement)	9 (17.31)	29 (32.22)	0.082
Peri (6-12 months since last menstrual period)	3 (5.77)	1 (1.11)	NA
Post (prior bilateral ovariectomy or >12 months since LMP with no prior hysterectomy)	32 (61.54)	55 (61.11)	> 0.999
Indeterminate or not evaluated	8 (15.48)	5 (5.56)	NA
Follow-up **TNBC** patients, months (IQR)[c]	36.83 (14.70 - 50.77)	28.79 (5.78 – 44.92)	0.932[b]
Vital Status of **TNBC** patients, n (%)			0.883
Alive	45 (86.54)	80 (88.89)	
Dead	7 (13.46)	10 (11.11)	
Time to death of **TNBC** patients, months (IQR)	30.75 (25.13 - 36.80)	42.73 (12.43 - 58.94)	0.743[b]

[a] Some values do not sum to heading totals because of missing data.
[b] Wilcoxon-Mann-Whitney test. Otherwise, the comparisons are by Fisher exact test.
[c] Months of follow-up after initial diagnosis.
NA – Not available
IQR – Interquartile range
LMP – Last menstrual period

AA patients were diagnosed with BC at earlier age, comparing with White patients (56.29 vs 58.61 years old, p = 0.037), and are more likely to develop TNBC than their White counterparts (28.42 % of all AA patients vs 11.89 % of all White patients, p = 0.0001). There were no statistically significant differences between AA and White patients with TNBC (thereafter, TNBC-AA and TNBC-White, respectively), concerning their age at diagnosis, pathological stage of the disease, menopause status and follow-up, although those differences were previously observed in other studies (Ademuyiwa et al., 2017; Huo et al., 2009; Keenan et al., 2015; Stewart et al., 2013). Additional information regarding the TCGA patients can be consulted in the Supplementary - Patients file, eTable 1.

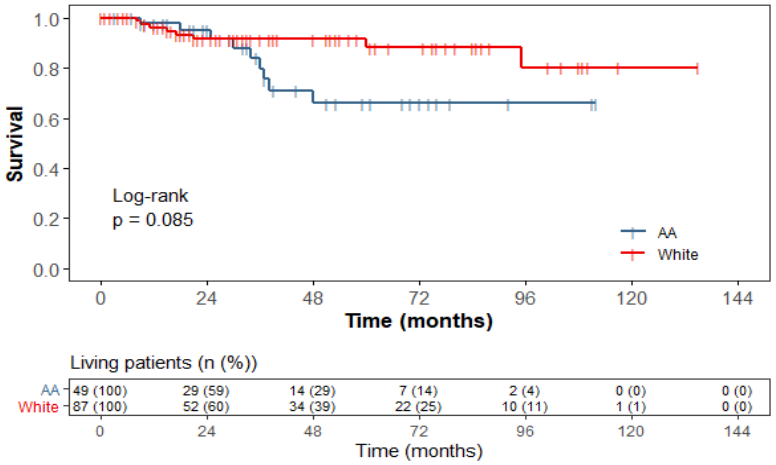
1.1. TNBC has different survival impact in AA and White patients when considering pre-menopause condition and the disease at stage II

Figure 15 shows the Kaplan-Meier survival curves from the TNBC cases from TCGA (Table 13) up to 12 years follow-up after diagnosis. Patients were divided according to their ancestry and the survival rate of both populations was analyzed in general or taking into account patient's menopause status and the different pathological stages of the disease at diagnosis.

Taking into consideration all TNBC cases, the survival curves between the two populations is not statistically significant (Figure 15a). Nonetheless, there is a clear tendency for a faster and higher mortality rate of AA patients. A similar non-statistically significant tendency can be observed when comparing patients presenting postmenopausal status (Figure 15c).

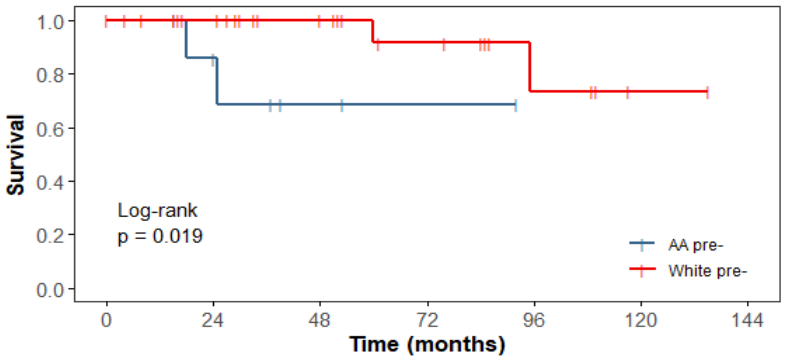
Notably, AA patients have a significantly lower survival probability comparing with their White counterparts within the group presenting pre-menopause condition and the group diagnosed at stage II TNBC (Figure 15b and e, log-rank p = 0.019 and 0.0038, respectively). To our best knowledge, this is the first study reporting such significantly lower survival probability in African-ancestry patients specifically associated to their premenopausal condition or pathological stage II at initial TNBC diagnosis, when comparing with matching White patients. Nonetheless, AA diagnosed with BC and TNBC have reportedly higher mortality than White patients at all ages, including the premenopausal ages (DeSantis et al., 2019). Furthermore, the low mortality among patients diagnosed at stage I (Figure 15d), and the low survival probability in the groups of patients from both ancestries diagnosed at stage III (Figure 15f) were somewhat expected, given the early and advanced stage of the disease, respectively.

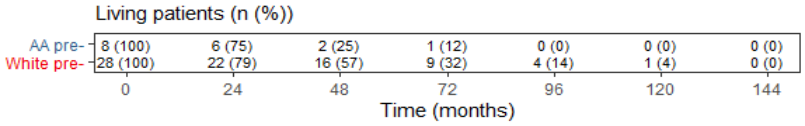
a


Survival
1.0
0.8
0.6
0.4
0.2
0.0
Log-rank
p = 0.085
AA
White
0
24
48
72
96
120
144
Time (months)

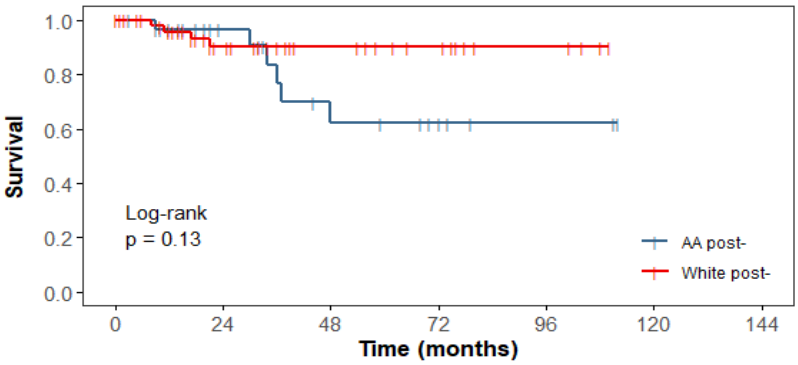
Living patients (n (%))
AA 49 (100) 29 (59) 14 (29) 7 (14) 2 (4) 0 (0) 0 (0)
White 87 (100) 52 (60) 34 (39) 22 (25) 10 (11) 1 (1) 0 (0)
0 24 48 72 96 120 144
Time (months)

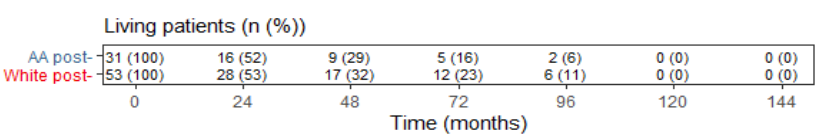
b


Survival
1.0
0.8
0.6
0.4
0.2
0.0
Log-rank
p = 0.019
AA pre-
White pre-
0
24
48
72
96
120
144
Time (months)

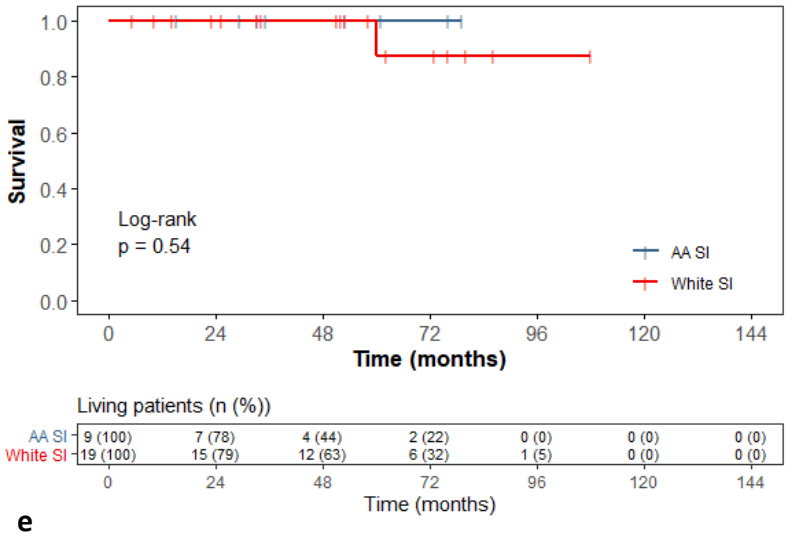

Living patients (n (%))
AA pre- 8 (100) 6 (75) 2 (25) 1 (12) 0 (0) 0 (0) 0 (0)
White pre- 28 (100) 22 (79) 16 (57) 9 (32) 4 (14) 1 (4) 0 (0)
0 24 48 72 96 120 144
Time (months)

c


Survival
1.0
0.8
0.6
0.4
0.2
0.0
Log-rank
p = 0.13
AA post-
White post-
0
24
48
72
96
120
144
Time (months)

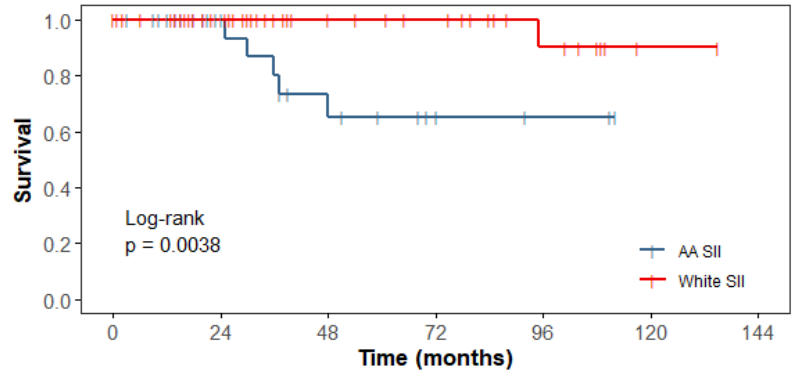

Living patients (n (%))
AA post- 31 (100) 16 (52) 9 (29) 5 (16) 2 (6) 0 (0) 0 (0)
White post- 53 (100) 28 (53) 17 (32) 12 (23) 6 (11) 0 (0) 0 (0)
0 24 48 72 96 120 144
Time (months)

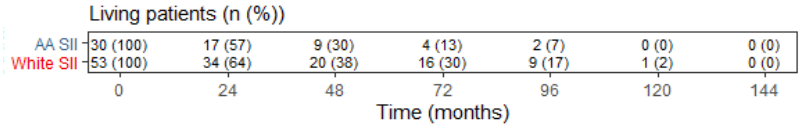
d


Survival
1.0
0.8
0.6
0.4
0.2
0.0
Log-rank
p = 0.54
AA SI
White SI
0
24
48
72
96
120
144
Time (months)

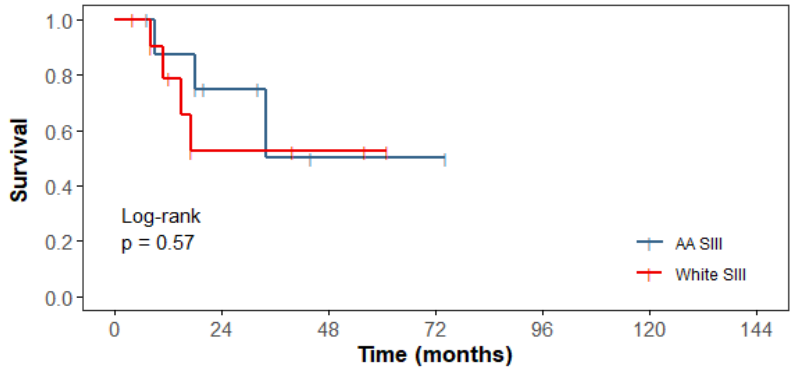
Living patients (n (%))
AA SI 9 (100) 7 (78) 4 (44) 2 (22) 0 (0) 0 (0) 0 (0)
White SI 19 (100) 15 (79) 12 (63) 6 (32) 1 (5) 0 (0) 0 (0)
0
24
48
72
96
120
144
Time (months)

e


Survival
1.0
0.8
0.6
0.4
0.2
0.0
Log-rank
p = 0.0038
AA SII
White SII
0
24
48
72
96
120
144
Time (months)


Living patients (n (%))
AA SII 30 (100) 17 (57) 9 (30) 4 (13) 2 (7) 0 (0) 0 (0)
White SII 53 (100) 34 (64) 20 (38) 16 (30) 9 (17) 1 (2) 0 (0)
0
24
48
72
96
120
144
Time (months)

f


Survival
1.0
0.8
0.6
0.4
0.2
0.0
Log-rank
p = 0.57
AA SIII
White SIII
0
24
48
72
96
120
144
Time (months)

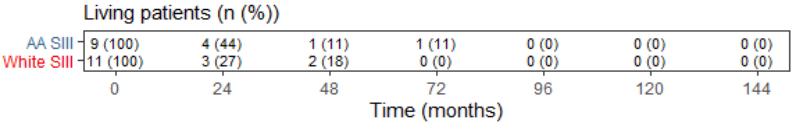

Living patients (n (%))
AA SIII 9 (100) 4 (44) 1 (11) 1 (11) 0 (0) 0 (0) 0 (0)
White SIII 11 (100) 3 (27) 2 (18) 0 (0) 0 (0) 0 (0) 0 (0)
0
24
48
72
96
120
144
Time (months)

Figure 15 Kaplan-Meier survival curves of TNBC cases stratified by ancestry up to 12 years follow-up after initial diagnosis. **a** All TNBC patients. **b** Premenopausal (pre-) TNBC patients. **c** Postmenopausal (post-) TNBC patients. **d** Stage I (SI) TNBC patients. **e** Stage II (SII) TNBC patients. **f** Stage III (SIII) TNBC patients. Blue survival curves correspond to African-American patients (AA), red curves show White patients. Values do not sum to heading totals because of missing follow-up information.

2. DIFFERENTIAL GENE EXPRESSION ANALYSIS

Before exploring gene expression data, RNA-seq data was filtered and normalized. Filtering of genes with low number of counts across libraries, each corresponding to a patient, must be performed prior to Differential Gene Expression Analysis (DGEA) in order to remove genes with very low expression since not only they have no power to infer differential expression between groups of patients, as they may also interfere with the real expression and biological relevance of other genes in a library and across libraries. To set the filtering criteria, we must have into account 1) the minimum number of libraries included in a group of patients and 2) the minimum number of counts a gene must have in a limiting number of libraries. For more information about this procedure, consult Chapter III – Section 1.3.1.1.4.

Since the groups of patients with smaller libraries, corresponding to premenopausal stage I and stage III disease AA patients, have two libraries each (Chapter III – Section 1.3.1.1.1.), we set this value as the minimum number of libraries in which a gene must be expressing so as not to be filtered. Figure 16 shows the total size of each library in counts-per-million (CPM). Having into consideration the libraries with a smaller total count size, we choose 10 CPM as a filtering criterion regarding the minimum number of counts a gene must have. Thus, a requirement of at least 10 CPM of expression in two or more libraries was used to ensure the retention of a gene in the analysis.

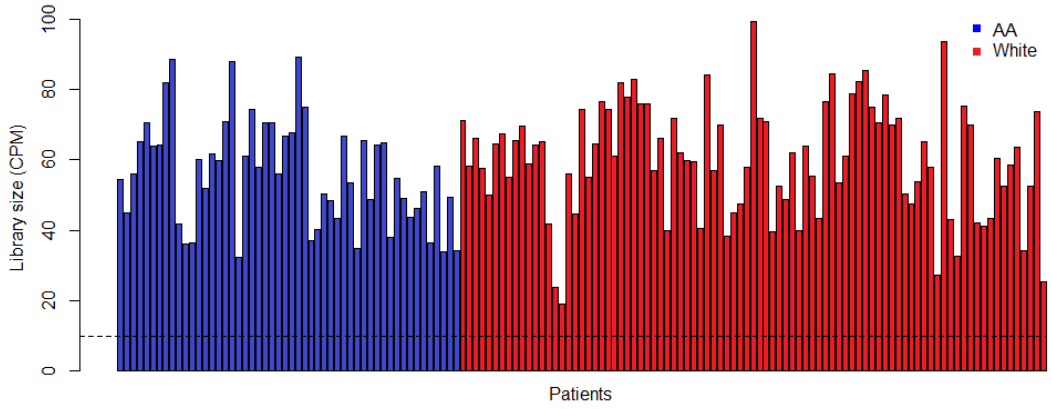


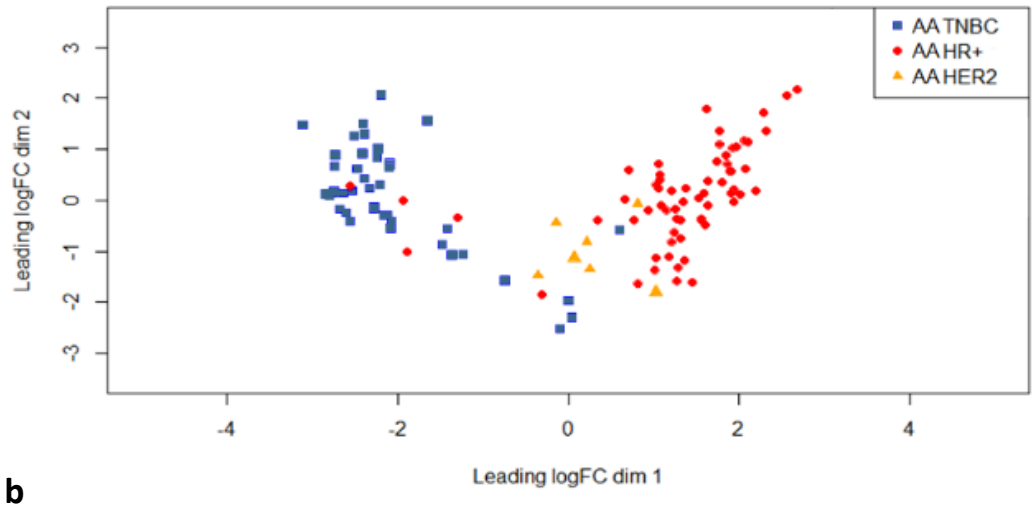
Figure 16 Library sizes of TNBC patients. Dashed line corresponds to 10 CPM filtering threshold. Blue bars correspond to libraries of AA patients and red bars correspond to White patients' libraries. *CPM* Counts per Million. *AA* African-American.

After filtering and normalization, and before performing the DGEA, differential expression between groups of patients was explored through multi-dimensional scaling (MDS) plots (Figure 17).

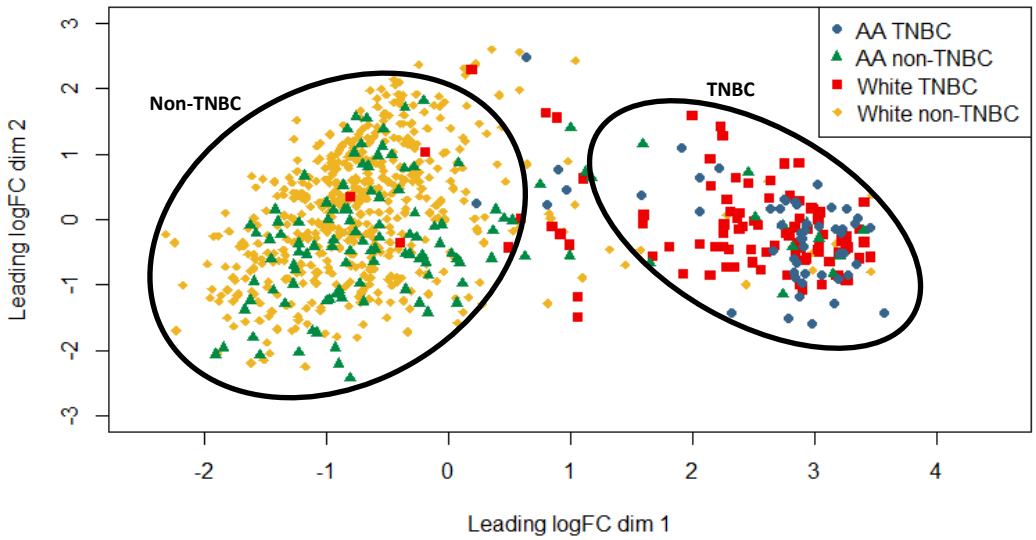
MDS plots of the RNA-seq samples represent the distances correspond to leading log-fold-changes between each pair of samples. In this case, MDS plots show the gene expression profile distance between patients in two dimensions, in which each dot represents a patient. In accordance

with previous studies (Huo et al., 2017; Stewart et al., 2013), and despite some outlier results, differences in gene expression profiles are higher between TNBC and other subtypes of BC (Hormone receptor [HR+] and HER2) patients, than between patients diagnosed with HR+ and HER2 BCs (Figure 17a), regardless of patients' ancestry (Figure 17b). The MDS plot does not show a clear differential expression profile between AA and White TNBC patients at the first two dimensions (Figure 17c). However, ancestry-associated discrepancies in disease severity and previous reported differences (Chang et al., 2018; Huo et al., 2017; Keenan et al., 2015; Lindner et al., 2013; Stewart et al., 2013; Sugita et al., 2016), motivated further analyses taking into consideration the data substructure.

a



b



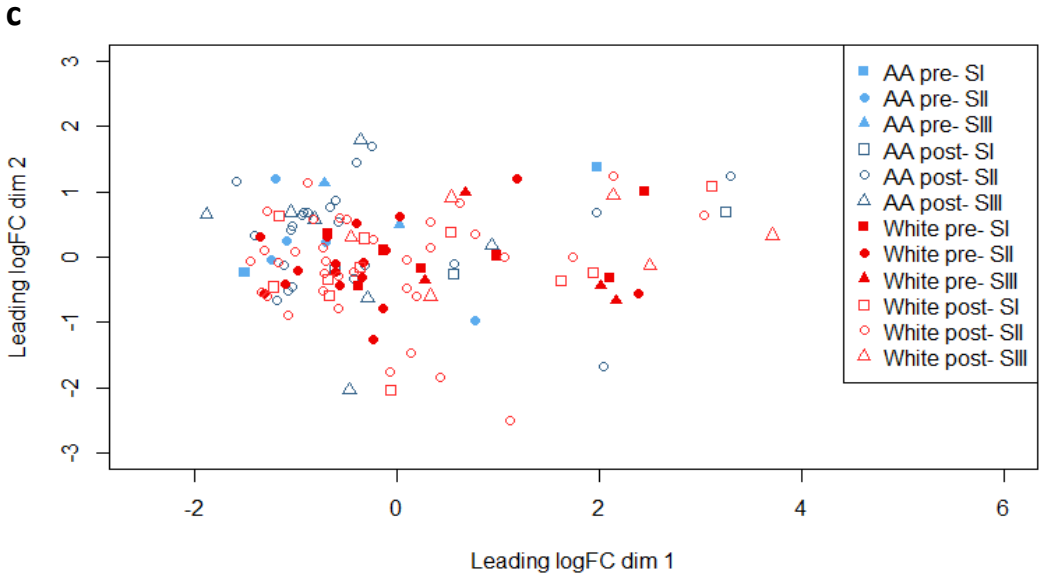


Figure 17 MDS plots of groups of RNA-sequencing data from TCGA. **a** AA patients with different breast cancer subtypes. **b** TNBC vs non-TNBC cases, according to ancestry. **c** TNBC-AA cases vs TNBC-AA cases, according to menopause status and pathological stage of the disease. *AA* African-American. *HR+* Hormone postive breast cancer cases. *HER2* HER2 positive breast cancer cases. *non-TNBC* hormone receptor and HER2 positive breast cancer cases. *pre-* Pre-menopause. *post-* Post-menopause. *SI* Stage I. *SII* Stage II. *SIII* Stage III. *FC* fold-change. *dim* dimension.

Since we observed differences in survival of TNBC-AA patients, comparing to TNBC-White patients, which were statistically significant for patients in premenopausal condition and with tumors diagnosed at stage II (Figure 15), and also considering the ancestry-associated discrepancy regarding the age of onset of BC (Bauer et al., 2007; Carey et al., 2006; Huo et al., 2009; Lund et al., 2009; Shoemaker et al., 2018; Zaky et al., 2009) and a study in which the differential gene expression increased between African-ancestry and White TNBCs in more advanced stages of the disease (Huo et al., 2017; Stewart et al., 2013), the menopause status and the pathological stage of the disease were clinical features taken into account in the DGEA.

Thus, in order to refine the differential gene expression profile between TNBC-AA patients and TNBC-White patients, which may explain the less favorable probability of African-ancestry patients to survive to this disease, the following groups of patients were subjected to DGEA:

1) AA vs White patients diagnosed with TNBC, independently of menopause status and stage of the disease;

2) AA vs White patients diagnosed with TNBC with matched menopause status, not having into account their pathological stage of the disease;

3) AA vs White patients diagnosed with TNBC with matched pathological stage of the disease, independently of patients' menopause status;

4) AA vs White patients diagnosed with TNBC with matched menopause status and pathological stage of the disease.

More details of the groups of patients in each DGEA in Chapter III - Section 1.3.1.1.1.

2.1. DGEA between TNBC-AA and TNBC-White patients

To find differences in the gene expression profile of the total TNBC cases from AA comparing with White patients, 52 TNBC-AA patients and 89 TNBC-White patients were eligible (one of the total 90 White patients with TNBC did not have available RNA-seq data) for this DGEA.

A total of 1122 genes were found to be differentially expressed, false discovery rate (FDR) ≤ 0.05 (Supplementary - DGEA, eTable 1), 638 of which were upregulated and 484 were downregulated in TNBC-AA patients (Table 14, Figure 18).

Table 14 Number of differentially expressed genes with FDR ≤ 0.05 in TNBC-AA patients comparing with TNBC-White patients.

↑ upregulated genes. ↓ downregulated genes

AA African-American patients. *W* White patients.

FDR ≤ 0.05	1122 (638 ↑ + 484 ↓)
# Patients	52 AA vs 89 W

An MD (mean-difference) plot was created to better visualize the dispersion of the 1122 differentially expressed genes between TNBC-AA patients and TNBC-White patients (Figure 18). In an MD plot, the log-fold-change (FC) for each gene is plotted against the average log CPM, which is a measure of how much expressed the genes are. Significantly differentially expressed genes at FDR of 5% are highlighted.

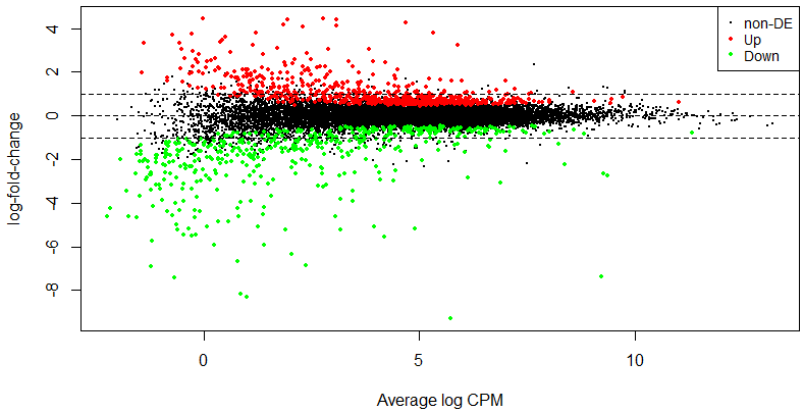


Figure 18 MD plot with log-fold-change against average count size, highlighting the differentially expressed (DE) genes between TNBCs from AA and TNBCs from White patients. *non-DE* non-differentially expressed genes (black dots). *Up* upregulated genes in TNBCs-AA comparing with TNBCs-White (red dots). *Down* downregulated genes in AA TNBCs comparing with White TNBCs (green dots). *CPM* counts per million.

In addition to MD plots, heat maps were also designed to display the differential expression profile of each patient. The advantage of a heat map is that it can show the expression pattern of each

differentially expressed gene (DEG) across all patients involved in that DGEA. Figure 19a displays a heatmap in which all the 1122 DEG were clustered together. Despite some outliers, i.e. White patients presenting a "AA phenotype" for specific DEG, and vice-versa, there is a clear difference in the expression profile between AA and White patients, as can be seen by the color pattern exhibited in the Figure 19a. Also, Figure 19b heatmap shows the 20 most statistically significant DEG identified in TNBC-AA patients.

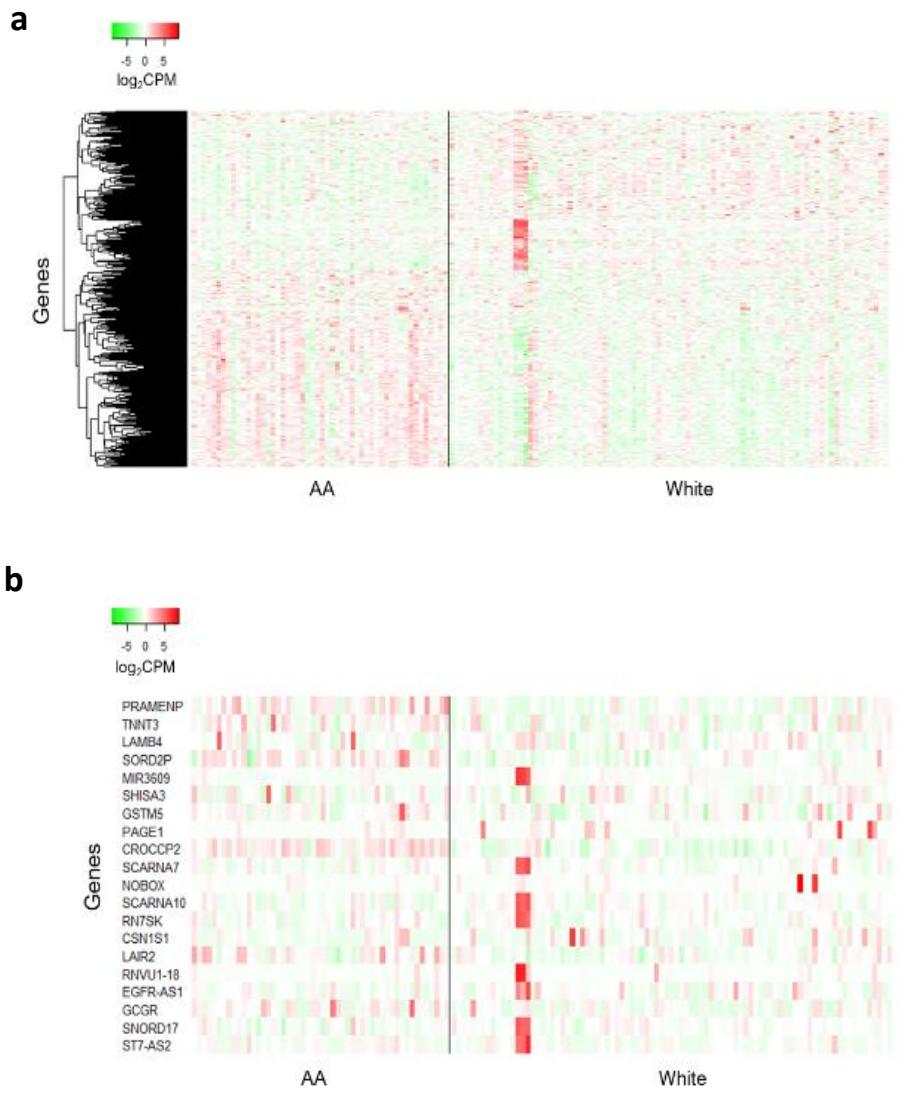


Figure 19 Heat maps across all patients diagnosed with TNBC. **a** Clustering of all differentially expressed genes observed in TNBCs-AA patients vs TNBC-White patients. **b** Top 20 of the most differentially expressed genes identified in TNBC-AA patients vs TNBC-White patients, in ascending order of FDR. *AA* African-American. *CPM* counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

The main aim of this work is to identify ancestry-associated differentially expressed molecules between the two populations, with potential impact in the prognosis and treatment of African-ancestry patients diagnosed with TNBC. For that reason, we first checked if the order of the statistical significance value of the DEG should be the most relevant criterion to the identification of potential candidate prognostic or target genes. Table 15 shows the top 20 DEG observed in TNBC-AA patients,

comparing with their White counterparts, and a brief description of the genes, according to GeneCards summary (https://www.genecards.org/), Entrez Gene summary (https://www.ncbi.nlm.nih.gov/gene) or UniProtKB (https://www.uniprot.org/uniprot/).

Table 15 Top 20 most statistically significant up and downregulated differentially expressed genes in TNBC-AA patients comparing with TNBC-White patients. **a** - Gene description according to GeneCards summary. **b** - Gene description according to Entrez Gene summary. **c** - Gene description according to UniProtKB.

Genes	Description	Log_2FC	FDR
PRAMENP	*Preferentially Expressed Antigen In Melanoma-Like* pseudogene.[a]	4.452	5.12E-11
TNNT3	*Troponin T, Fast Skeletal Muscle*, the binding of Ca(2+) to the trimeric troponin complex initiates the process of muscle contraction.[b]	4.436	8.39E-11
LAMB4	*Laminin Subunit Beta 4*, mediates the attachment, migration and organization of cells into tissues during embryonic development.[c]	3.473	5.36E-10
SORD2P	*Sorbitol Dehydrogenase 2, Pseudogene.*[a]	2.861	7.49E-10
MIR3609	*MicroRNA 3609.*[a]	-8.155	5.67E-09
SHISA3	*Shisa Family Member 3,* transmembrane adaptors that modulate both WNT and FGF signaling by blocking the maturation and transport of their receptors to the cell surface.[b]	4.186	6.03E-09
GSTM5	*Glutathione S-Transferase Mu 5*, encodes one of the mu class of cytosolic and membrane-bound glutathione S-transferase, which is involved in the detoxification of electrophilic compounds, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress, by conjugation with glutathione. Genetic variations in the mu class gene cluster can change an individual's susceptibility to carcinogens and toxins as well as affect the toxicity and efficacy of certain drugs.[b]	4.251	7.83E-09
PAGE1	*PAGE Family Member 1*, belongs to a family of genes that are expressed in many tumors but not in normal tissues, except for the testis. Unlike the other gene family members, this gene does not encode an antigenic peptide. Nothing is presently known about the function of this protein[b] nor if it is expressed in BC.	-8.324	1.28E-08
CROCCP2	*Ciliary Rootlet Coiled-Coil, Rootletin Pseudogene 2.*[a]	1.044	1.87E-08
SCARNA7	*Small Cajal Body-Specific RNA 7*, RNA gene affiliated with the scaRNA class.[a]	-6.866	1.87E-08
NOBOX	*NOBOX Oogenesis Homeobox,* in mice, it is essential for folliculogenesis and regulation of oocyte-specific genes.[b]	-7.427	1.87E-08
SCARNA10	*Small Cajal Body-Specific RNA 10*, RNA gene affiliated with the scaRNA class.[a]	-6.330	1.87E-08
RN7SK	*RNA Component Of 7SK Nuclear Ribonucleoprotein,* RNA gene affiliated with the 7SK RNA class. Diseases associated with *RN7SK* include Acquired Immunodeficiency Syndrome and Gastric Cancer.[a]	-7.358	2.05E-08
CSN1S1	*Casein Alpha S1*, is a protein encoding gene with an important role in the capacity of milk to transport calcium phosphate.[c]	-9.283	3.64E-08
LAIR2	*Leukocyte Associated Immunoglobulin Like Receptor 2*, is a member of the immunoglobulin superfamily. It was identified by its similarity to leukocyte-associated immunoglobulin-like receptor 1, a membrane-bound receptor that modulates innate immune response.b	2.427	5.10E-08
RNVU1-18	*RNA, U1 Small Nuclear 25, Pseudogene*.[a]	-6.900	5.44E-08
EGFR-AS1	*EGFR Antisense RNA 1*, a non-coding RNA gene.[a]	-5.932	5.44E-08
GCGR	*Glucagon Receptor*, important in controlling blood glucose levels. Defects in this gene are a cause of non-insulin-dependent diabetes mellitus.[b]	3.704	6.01E-08
SNORD17	*Small Nucleolar RNA, C/D Box 17*, is an RNA Gene, and is affiliated with the snoRNA class.[a]	-5.202	6.54E-08
ST7-AS2	*ST7 Overlapping Transcript 2 (Antisense Non-Coding RNA)*.[a]	-6.662	9.27E-08

Together with the complete list of DEG identified in this analysis (Supplementary - DGEA, eTable 1), Table 15 confirms that there are indeed biological differences between African-ancestry and White patients diagnosed with TNBC. However, the top 20 most significantly DEG an the complete list of DEG *per se* do not explicit show which are the most relevant genes affecting downstream networks, pathways and cellular mechanisms potentially involved with the observed survival differences between AA and White patients diagnosed with TNBC. Thus, although the highly significantly DEG presented in Table 15 might have important roles in ancestry-associated discrepancy of TNBC, just looking at the most significantly DEG *per se* is not enough to foreshadow their biological relevance.

Therefore, and since the stringent parameters used in the DGEA already restrict our results, we did not exclude the hypothesis that less statistically significant DEG may indeed have an important biological function related with TNBC development in African-ancestry patients. Thus, such DEG may be potential drivers of TNBC aggressiveness in this population and, as such, they should not be neglected. Accordingly, these larger sets of DEG will be further subjected to gene set enrichment analysis (GSEA) and network design in order to identify genes potentially involved in altered pathways and cellular processes involved the ancestry-associated discrepancy of TNBC.

2.2. DGEA between TNBC-AA and TNBC-White patients with matched menopause status, independently of pathological stage

The Kaplan-Meyer survival curves (Figure 15, Section 1.1.) suggested that TNBC-AA patients have a significantly reduced survival probability, comparing with TNBC-White patients, particularly when patients are premenopausal and diagnosed at stage II disease. Thus, we took into consideration that the menopause status and disease's pathological stage are influencing TNBC development and prognosis in the two populations.

To clarify this issue, DGEA was performed with patients presenting the same menopause status, where 41 AA patients and 86 White patients with menopause information were included. The number of DGE resulting from DGEAs of both contrasts with pre- and postmenopausal patients are depicted in Table 16. More information about the groups of patients of each contrast can be consulted in Chapter III, Section 1.3.1.1.1.

Table 16 Number of differentially expressed genes in TNBC-AA patients comparing with TNBC-White patients (with FDR ≤ 0.05) according to patients' menopause status.

↑ upregulated genes. ↓ downregulated genes

AA African-American patients. *W* White patients

	Pre-menopause	**Post-menopause**
FDR ≤ 0.05	83 (76 ↑ + 7 ↓)	718 (352 ↑ + 366 ↓)
# Patients	9 AA vs 29 W	32 AA vs 57 W

Complete DGEA results of each contrast can be consulted in Supplementary - DGEA file, eTables 2-3. The numbers of DEG of each contrast, as well as its FC difference and relative abundance, is highlighted in the following MD plots (Figure 20).

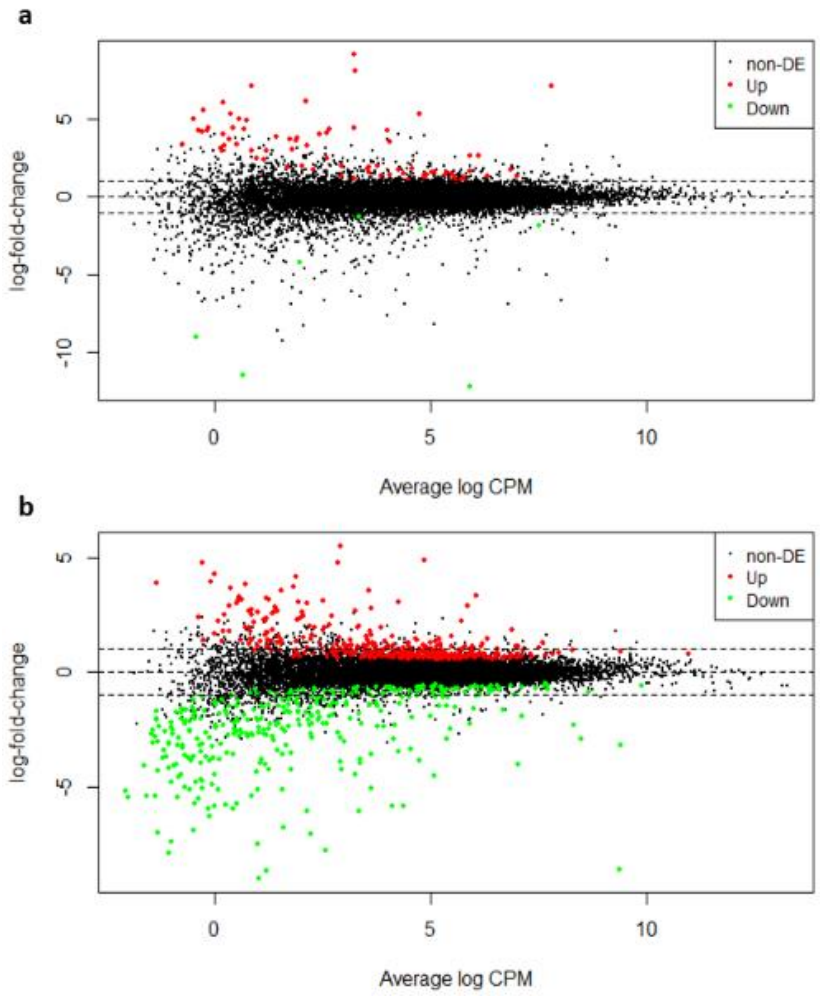


Figure 20 MD plots with log-fold-change against average count size, highlighting the differentially expressed (DE) genes in TNBCs from AA comparing with menopause status-matched TNBCs from White patients. **a** Pre-menopause matched; **b** Post-menopause matched. *non-DE* non-differentially expressed genes (black dots). *Up* upregulated genes in AA TNBCs comparing with White TNBCs (red dots). *Down* downregulated genes in AA TNBCs comparing with White TNBCs (green dots). *CPM* counts per million.

Less DGE were identified in the DGEA of premenopausal patients (Table 16, Figure 20a) than in postmenopausal patients (Table 16, Figure 20b), which, in part, may be due to the lower number of patients included in the premenopausal contrast and not necessary due to the fact that there might be less biological dissimilarities between younger AA and White TNBC patients.

Heatmaps showing the gene expression profile resulting from the DGEA of the premenopausal contrast are represented in Figure 21. Figure 21a represents the gene clustering of all identified DEG and Figure 21b shows the top 20 DEG. Again, the genes appearing in the top 20 (Figure 21b) *per se* may not explain the biological causes behind the increased TNBC incidence and lower survival of

premenopausal AA patients. However, considering all the 83 DEG, we highlight the fact that PIK3CA is the most statistically significant overexpressed gene (Figure 21b). This oncogene is often mutated in BC and PI3K pathway, being involved in tumorigeneses and patients' poor prognosis (Aleskandarany et al., 2010; Berns et al., 2007).

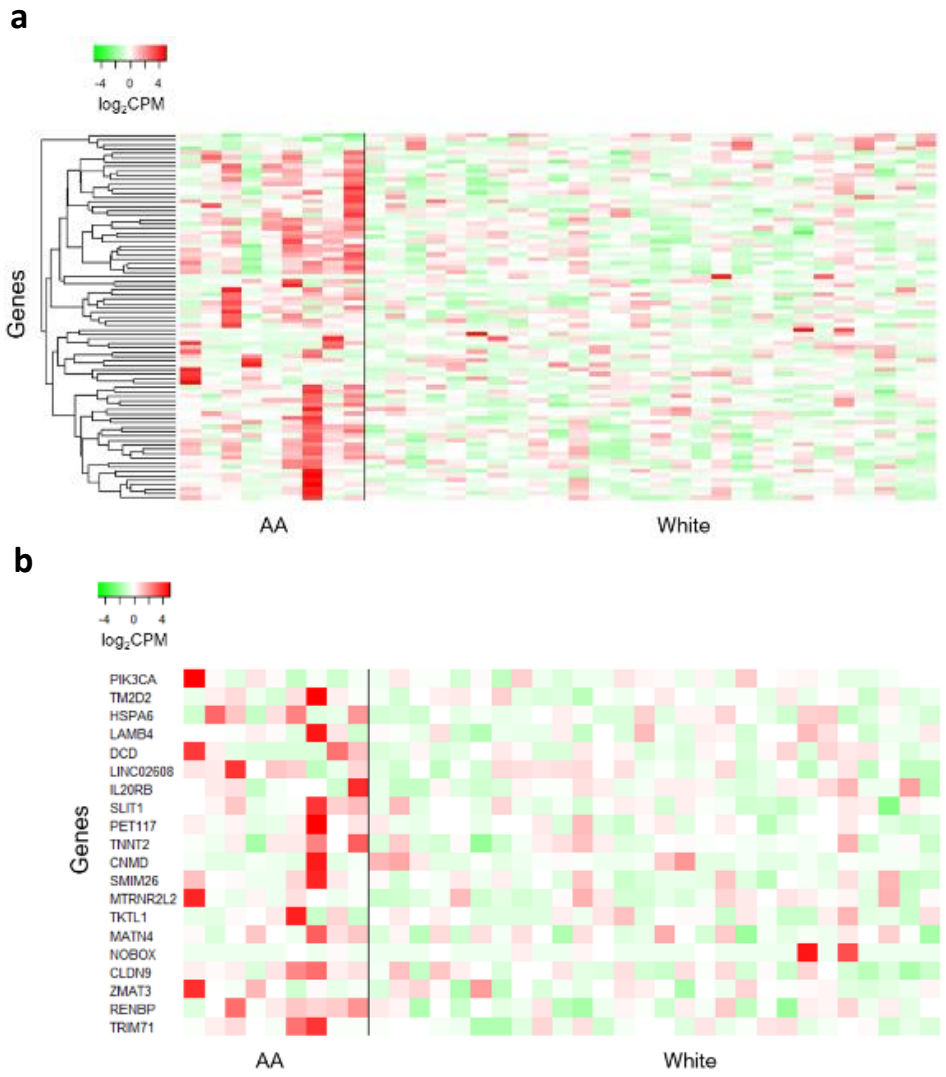


Figure 21 Heat maps of premenopausal TNBC patients. **a** Clustering of all differentially expressed genes observed in TNBCs-AA patients vs TNBC-White patients. **b** Top 20 of the most differentially expressed genes identified in TNBC-AA patients vs TNBC-White patients, in ascending order of FDR. *AA* African-American. *CPM* counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

On the other hand, the DGEA of the contrast with postmenopausal patients (32 AA patients vs 57 White patients) resulted in 718 DEG (Table 16). Figure 22 shows the heatmaps representing the clustering of the DEG identified in postmenopausal TNBC-AA patients (Figure 22a). Given the relatively large number of DEG identified, different clusters (red and green patterns) can be clearly visualized (Figure 22a). Nonetheless, with a large set of identified DEG, the top 20 DEG *per se* (Figure 22b) does not show clear relevant information regarding differences in TNBC biology between AA and White premenopausal patients.

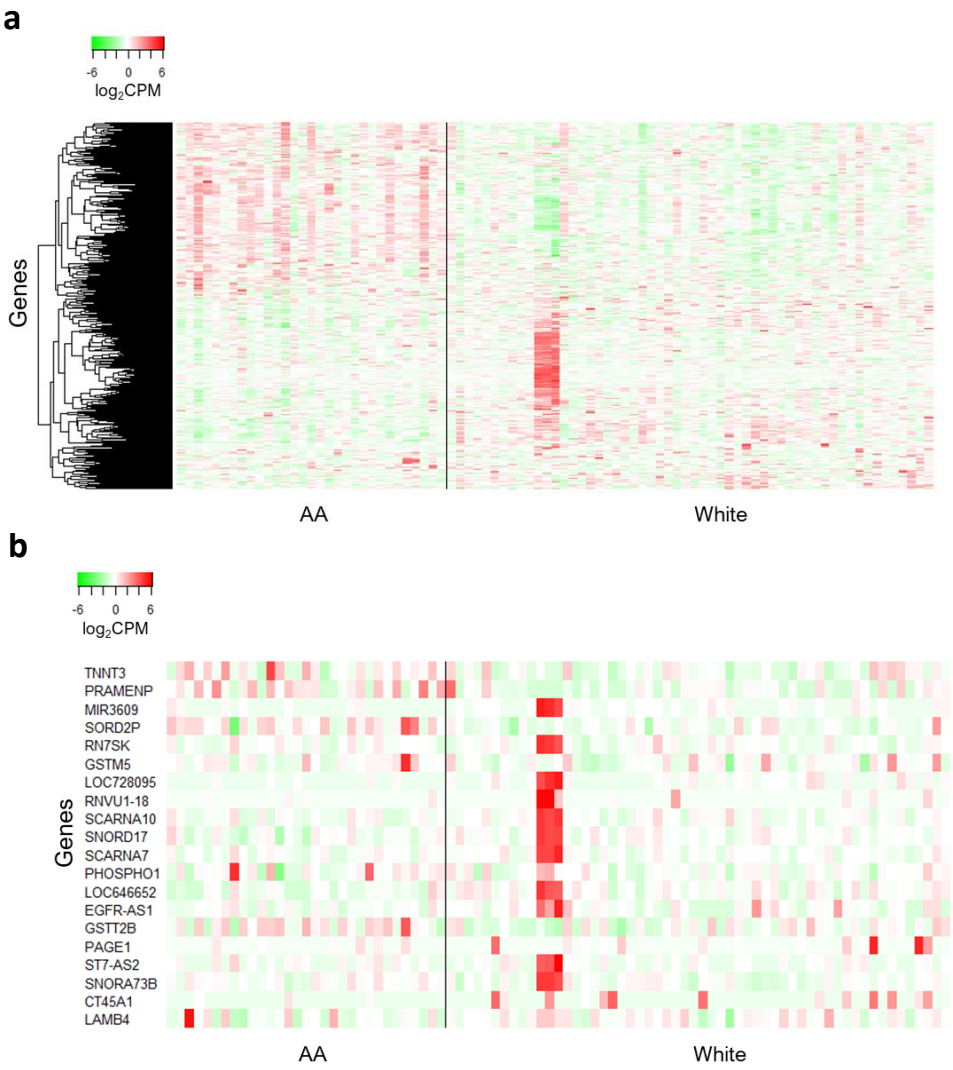


Figure 22 Heat maps of postmenopausal TNBC patients. **a** Clustering of all differentially expressed genes observed in TNBCs-AA patients vs TNBC-White patients. **b** Top 20 of the most differentially expressed genes identified in TNBC-AA patients vs TNBC-White patients, in ascending order of FDR. *AA* African-American. *CPM* counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

In general, the results obtained in this section, showing that are indeed DEG when controlling for patients' pre- or post-menopause condition, suggest that the hormonal features of each menopausal phase also influences the molecular ancestry-associated disparities.

2.3. DGEA between TNBC-AA and TNBC-White patients with matched pathological stage of the disease, independently of menopause status

The following DGEA were performed in patients with the same pathological stage of the disease at TNBC diagnosis, independently of their menopause status. These DGEA included a total of 51 AA patients and 90 White patients with TNBC staging data. Table 17 shows the number of DEG resulting from the DGEA of each of the three analyzed pathological stages. Additional information regarding the

groups of patients of each contrast can be consulted in Chapter III, Section 1.3.1.1.1. Complete DGEA results of each stage-matched contrast can be consulted in Supplementary - DGEA file, eTables 4-6.

Table 17 Number of differentially expressed genes in TNBC-AA patients comparing with TNBC-White patients (with FDR ≤ 0.05) according to patients' stage of the disease.

↑ upregulated genes. ↓ downregulated genes

AA African-American patients. *W* White patients

	Stage I	**Stage II**	**Stage III**
FDR ≤ 0.05	30 (22 ↑ + 8 ↓)	1776 (1073 ↑ + 703 ↓)	94 (36 ↑ + 58 ↓)
# Patients (AA vs W)	9 AA vs 19 W	32 AA vs 59 W	10 AA vs 12 W

Figure 23 shows MD plots for each of the analyzed stage-matched contrasts and significantly DEG are highlighted. Interestingly, from all the DGEA performed in this project, we identified the largest set of DEG in the contrast with stage II disease patients (1776 DEG, 1073 up- and 703 downregulated DEG; Table 17 and Figure 23b). The set of DEG found in patients with disease at this stage is even larger than the set of DEG identified with all the TNBC patients (1122 DEG, 638 up- and 484 downregulated DEG; Table 14 and Figure 18), even though there is a lower number of patients presenting stage II disease (32 AA and 59 White patients, Table 17) comparing with the total number of patients included in the DGEA with all TNBC cases (52 AA and 89 White patients, Table 14).

Furthermore, regarding the other stages, we observed that although stage I contrast has more patients than the stage III, there are three times more DEG in the contrast with stage III disease patients than in the contrast with stage I disease patients (Table 17, Figure 23a and 23c).

Overall, and as previously described in a study with BC patients from TCGA (Stewart et al., 2013), there seems to be an increased expression profile distance between AA and White patients throughout the evolution of the disease.

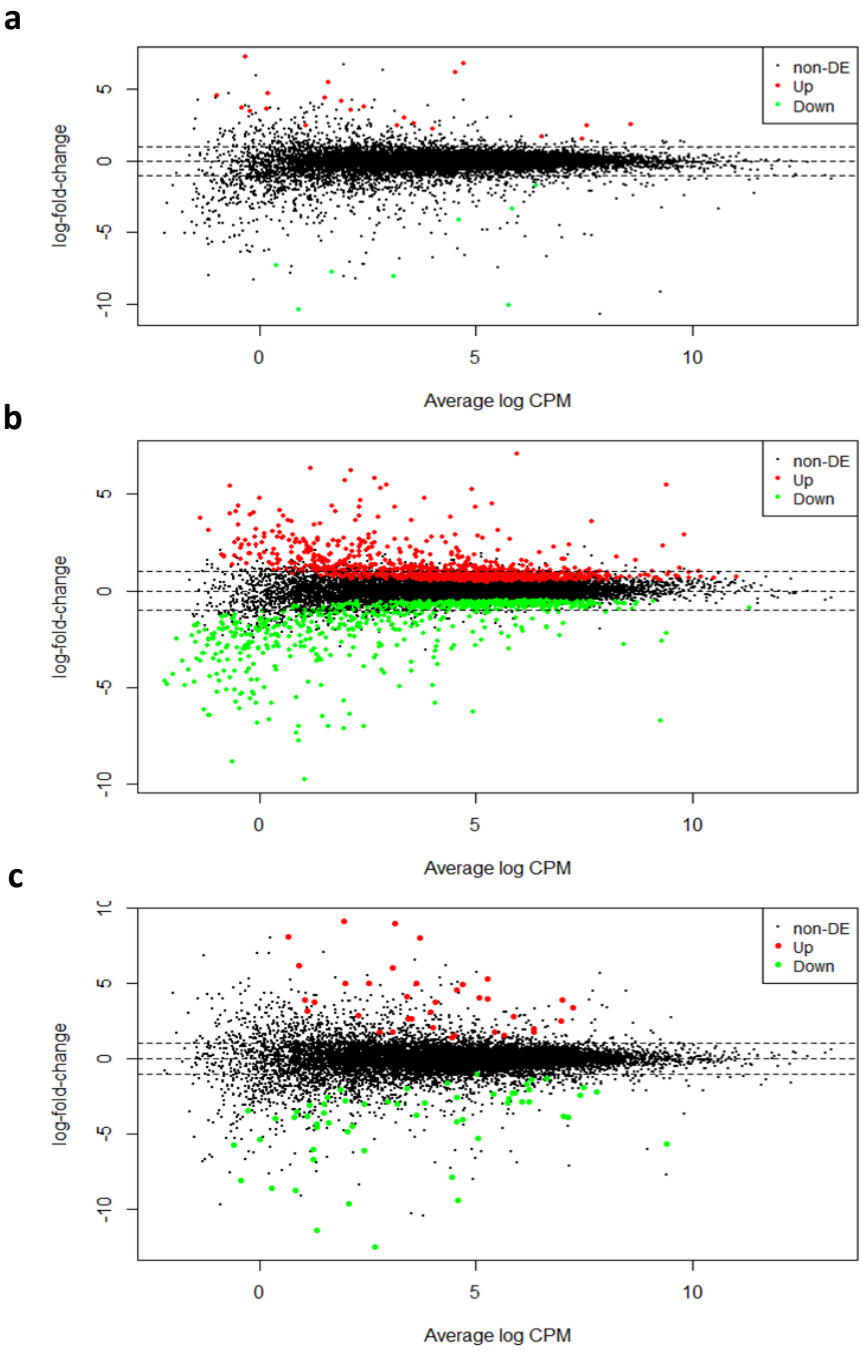


Figure 23 MD plots with log-fold-change against average count size, highlighting the differentially expressed (DE) genes in TNBCs from AA comparing with stage-matched TNBCs from White patients. **a** Stage I matched. **b** Stage II matched. **c** Stage III matched. *non-DE* non-differentially expressed genes (black dots). *Up* upregulated genes in AA TNBCs comparing with White TNBCs (red dots). *Down* downregulated genes in AA TNBCs comparing with White TNBCs (green dots). *CPM* counts per million.

Figure 24 displays the heat maps with clustering of all DEG diagnosed with stage I disease (Figure 24a) and their top 20 DEG (Figure 24b). Given the reduced number of DEG identified in this group of patients, no clear DEG clustering can be visualized (Figure 24a). This DGEA only resulted in 30 DEG and the top 20 DEG *per se* (Figure 24b) does not seem to reflect the biological context of the ancestry-

associated TNBC discrepancy in the early stage of the disease. Importantly, and as aforementioned, the gene expression profile dissimilarities seem to increase throughout TNBC progression. Thus, in an early stage of development, TNBC gene expression may not be so dissimilar between AA and White patients.

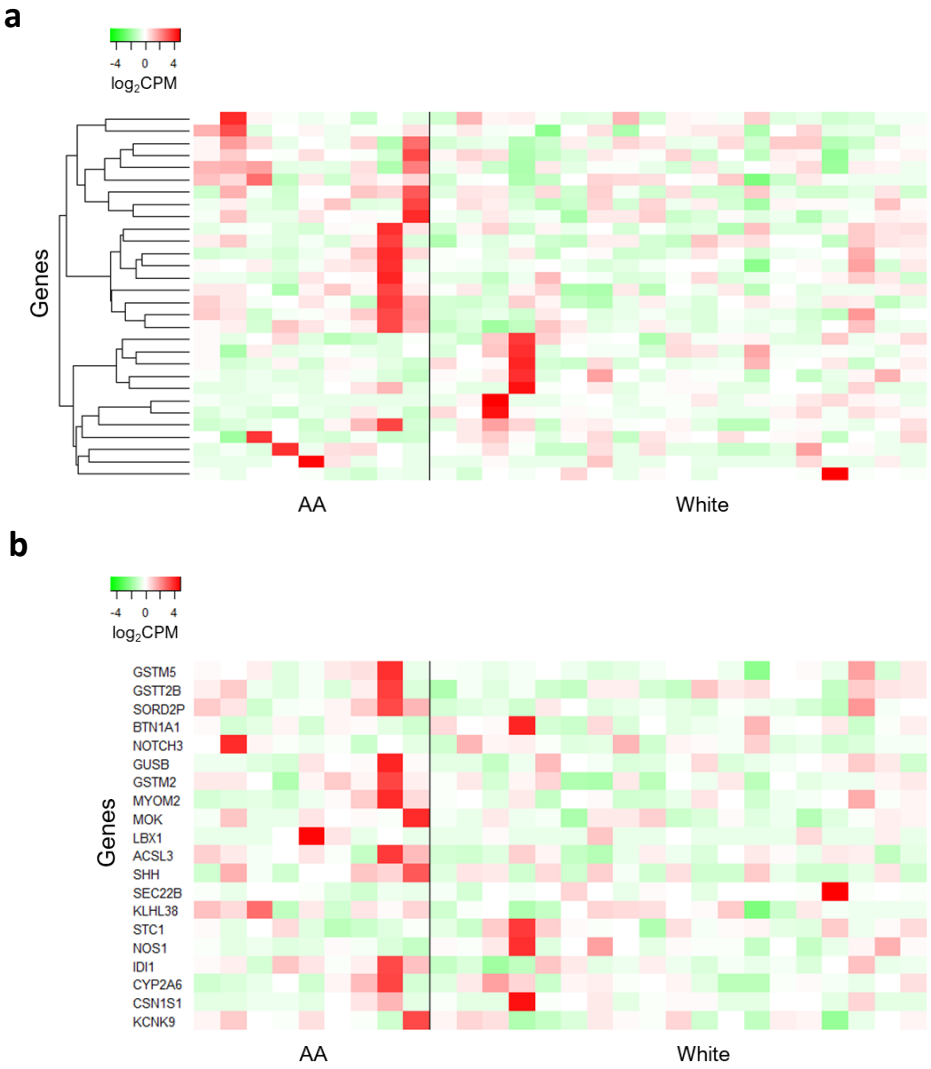


Figure 24 Heat maps of stage I TNBC patients. **a** Clustering of all differentially expressed genes observed in TNBCs-AA patients vs TNBC-White patients. **b** Top 20 of the most differentially expressed genes identified in TNBC-AA patients vs TNBC-White patients, in ascending order of FDR. *AA* African-American. *CPM* counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

Figure 25 shows the heat maps with clustering of all DEG identified in TNBC-AA patients presenting stage II disease comparing with their White counterparts (Figure 25a) and their top 20 DEG (Figure 25b). Given the total 1776 DEG identified within stage II disease and the relatively large number of patients included in this group (32 AA and 59 White patients, Table 17), there is a clear difference in the gene expression profile between the two populations (Figure 25a). However, by observing only the top 20 DEG (Figure 25b) we may be overlooking relevant clues regarding relevant pathways and

cellular processes affecting the ancestry-associated TNBC discrepancy observed in stage II disease patients.

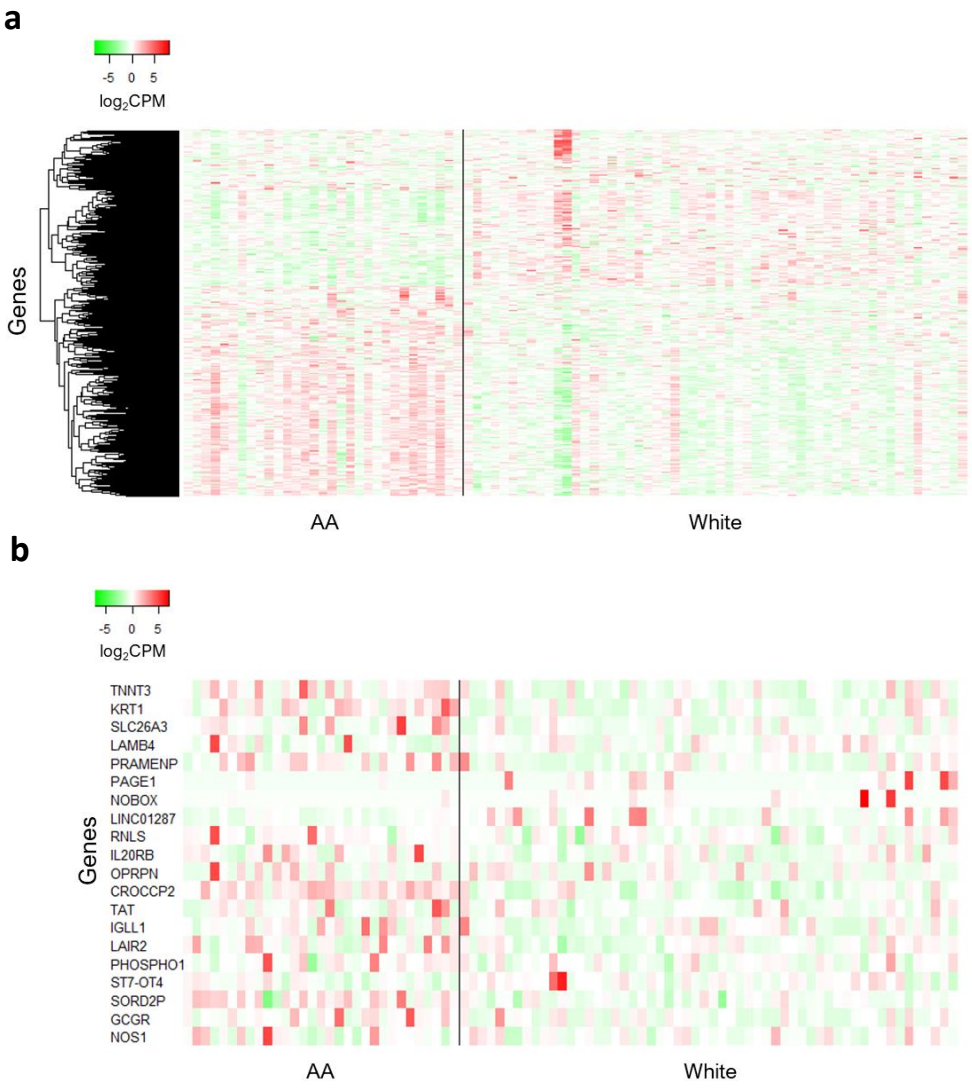


Figure 25 Heat maps of stage II TNBC patients. **a** Clustering of all differentially expressed genes observed in TNBCs-AA patients vs TNBC-White patients. **b** Top 20 of the most differentially expressed genes identified in TNBC-AA patients vs TNBC-White patients, in ascending order of FDR. *AA* African-American. CPM counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

Finally, Figure 26 depicts the heat maps with clustering of all DEG identified in TNBC-AA patients presenting stage III disease (Figure 26a) and their top 20 (Figure 26b). The clustering formation of the 94 DEG identified in TNBC-AA patients, comparing with TNBC-White patients (Figure 26a), is an indicative of the clear distinct molecular profile between patients of the two populations diagnosed with stage III disease. Furthermore, *PIK3CA* is, again, the most statistically significant DEG (Figure 26b).

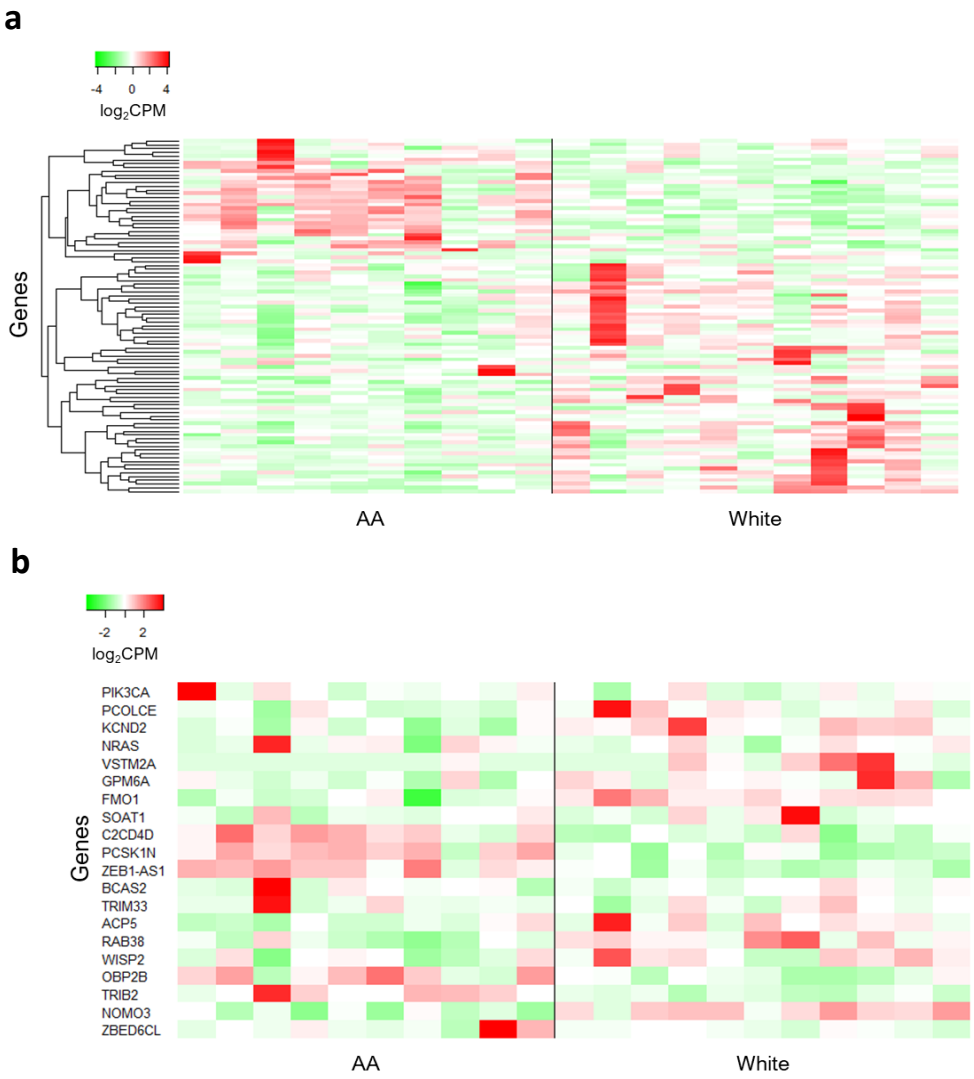


Figure 26 Heat maps of stage III TNBC patients. **a** Clustering of all differentially expressed genes observed in TNBCs-AA patients vs TNBC-White patients. **b** Top 20 of the most differentially expressed genes identified in TNBC-AA patients vs TNBC-White patients, in ascending order of FDR. AA African-American. *CPM* counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

Altogether, these observations suggest that there is an increased differentiation in the ancestry-specific gene expression profile throughout the staging progress of TNBC, which may be due to an increased rate of genomic instability and alterations in the tumor microenvironment, as previously suggested by other works (Negrini et al., 2010; Sonugür and Akbulut, 2019).

2.4. DGEA between TNBC-AA and TNBC-White patients with matched menopause status and disease stage

Finally, we grouped the patients taking in account both their menopause status and disease stage. In this analysis, 41 TNBC-AA patients and 81 TNBC-White patients, with both menopause and disease stage information, were considered. The number of DEG resulting from DGEA for each contrast

is shown in Table 18. Additional information regarding the groups of patients of each contrast can be consulted in Chapter III, Section 1.3.1.1.1.

Table 18 Number of differentially expressed genes in TNBC-AA patients comparing with TNBC-White patients (with FDR ≤ 0.05) according to patients' menopause status and pathological stage of the disease.

↑ upregulated genes. ↓ downregulated genes

AA African-American patients. *W* White patients

	Stage I	**Stage II**	**Stage III**
Pre-menopause			
FDR ≤ 0.05	0	63 (60 ↑ + 3 ↓)	2 (2 ↑)
# Patients	2 AA vs 7 W	5 AA vs 17 W	2 AA vs 4 W
Post-menopause			
FDR ≤ 0.05	10 (10 ↑)	1360 (642 ↑ + 718 ↓)	2 (2 ↓)
# Patients	3 AA vs 11 W	22 AA vs 36 W	7 AA vs 6 W

Complete DGEA results for each contrast are presented in Supplementary - DGEA file, eTables 7-12. DGE of each contrast are highlighted in the following MD plots (Figure 13).

From Table 18 and Figure 27 we can observe that some contrasts do not have DEG (taking in account the FDR ≤ 0.05), such as the contrast involving premenopausal patients with disease at stage I, or have very few DEG, namely the contrasts of pre- and postmenopausal patients with disease at stage III, where only two DEG were identified.

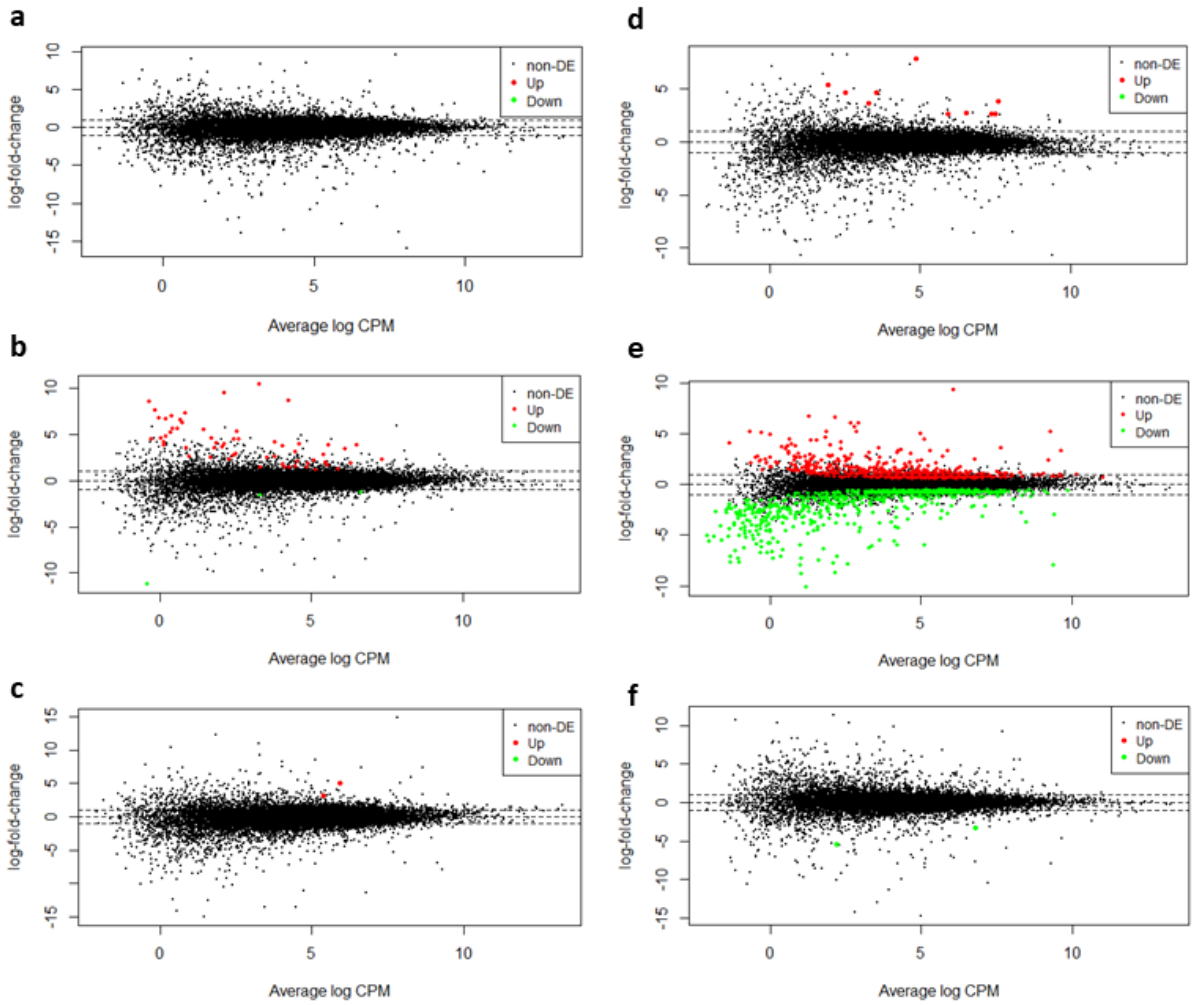


Figure 27 MD plots with log-fold-change against average count size, highlighting the differentially expressed (DE) genes *in* TNBC-AA comparing TNBC-White patients with matched menopause status and stage. **a** All patients in pre-menopause and disease at stage I. **b** All patients in pre-menopause and disease at stage II. **c** All patients in pre-menopause with the disease at stage III. **d** All patients in post-menopause and disease at stage I. **e** All patients in post-menopause and stage II. **f** All patients in post-menopause and disease at stage II. *non-DE* non-differentially expressed genes (black dots), *Up* upregulated genes in TNBC-AA comparing with TNBC-White patients (red dots), *Down* downregulated genes in in TNBC-AA comparing with TNBC-White patients (green dots), *CPM* counts per million.

Looking specifically for patients with disease at pathological stage I, only the ones with postmenopausal status showed a differential expression profile between the two populations. Thus, since such low number of DEG does not allow for biological contextualization through gene set enrichment analysis (GSEA) or network design, we prepared Table 19 with a brief description of the 10 DEG identified in TNBC-AA patients with post-menopause condition and disease at stage I, in order to ascertain if there is a biological pattern among this set of DEG.

Table 19 Differentially expressed genes in stage I disease postmenopausal TNBC-AA patients comparing with matching TNBC-White patients. **a** - Gene description according to GeneCards summary. **b** - Gene description according to Entrez Gene summary. **c** - Gene description according to UniProtKB.

Genes	Description	Log_2FC	FDR
GSTM5	*Glutathione S-Transferase Mu 5*, encodes one of the mu class of cytosolic and membrane-bound glutathione S-transferase, which is involved in the detoxification of electrophilic compounds, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress, by conjugation with glutathione. Genetic variations in the mu class gene cluster can change an individual's susceptibility to carcinogens and toxins as well as affect the toxicity and efficacy of certain drugs.[b]	7.848	9.61E-06
GSTM2	*Glutathione S-Transferase Mu 2,* encodes one of the mu class of cytosolic and membrane-bound glutathione S-transferase, which is involved in the detoxification of electrophilic compounds, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress, by conjugation with glutathione. Genetic variations in the mu class gene cluster can change an individual's susceptibility to carcinogens and toxins as well as affect the toxicity and efficacy of certain drugs.[b]	4.688	1.95E-04
GUSB	*Glucuronidase Beta,* encodes a hydrolase, located in the lysosome, that degrades glycosaminoglycans, including heparan sulfate, dermatan sulfate, and chondroitin-4,6-sulfate.[b]	3.828	9.33E-04
GSTT2B	*Glutathione S-Transferase Theta-2B*, encodes a member of a superfamily of proteins that catalyze the conjugation of reduced glutathione to a variety of electrophilic and hydrophobic compounds.[b]	5.368	2.07E-03
ACSL3	*Acyl-CoA Synthetase Long Chain Family Member 3,* encodes an isozyme that converts free long-chain fatty acids into fatty acyl-CoA esters, and thereby play a key role in fatty acid degradation and lipid biosynthesis.[b]	2.636	2.07E-03
IDI1	*Isopentenyl-Diphosphate Delta Isomerase 1,* encodes a peroxisomally-localized enzyme that catalyzes the interconversion of isopentenyl diphosphate (IPP) to its highly electrophilic isomer, dimethylallyl diphosphate (DMAPP), which are the substrates for the successive reaction that results in the synthesis of farnesyl diphosphate and, ultimately, cholesterol.[b]	2.726	4.59E-03
SORD2P	*Sorbitol Dehydrogenase 2, Pseudogene*.[a]	4.717	4.59E-03
ELOVL5	*ELOVL Fatty Acid Elongase 5,* encodes an endoplasmic reticulum-bound enzyme that catalyzes the first and rate-limiting reaction of the four reactions that constitute the long-chain fatty acids elongation cycle.[c]	2.646	5.31E-03
AACS	*Acyl-CoA Synthetase Family Member 1*, encodes a protein that activates acetoacetate to acetoacetyl-CoA. May be involved in utilizing ketone body for the fatty acid-synthesis during adipose tissue development.[c]	2.632	9.20E-03
FMO5	*Flavin Containing Dimethylaniline Monoxygenase 5*, encodes a NADPH-dependent flavoenzymes that catalyzes the oxidation of soft nucleophilic heteroatom centers in drugs, pesticides, and xenobiotics.[b]	3.645	3.33E-02

Interestingly, most of these 10 DEG, which are all upregulated in TNBC-AA patients, are related to detoxifying mechanisms and fatty acids or other lipid synthesis. Also, *GSTM5*, *GUSB*, *GSTM2*, *ACSL3* and *IDI1* are differentially expressed in the DGEA with stage I patients (Figure 24b), independently of menopause status. This observation suggests that these TNBC-AA patients, comparing to TNBC-White patients, may have a higher drug metabolism, which could be responsible for the impairment of the response to treatment, even in this early-stage disease.

Figure 28 represents the heatmap organized according to the clustering of the DEG identified in TNBC-AA patients with post-menopause condition and disease at stage I. This figure shows that, although the DEG identified in this group of TNBC-AA patients seem to be suggestively associated to detoxifying mechanisms and lipid synthesis (Table 19), this hypothesis needs some caution, since there are only 3 TNBC-AA patients in this contrast, and only one of them is highly expressing all of these 10 DEG, which may be skewing the DGEA.

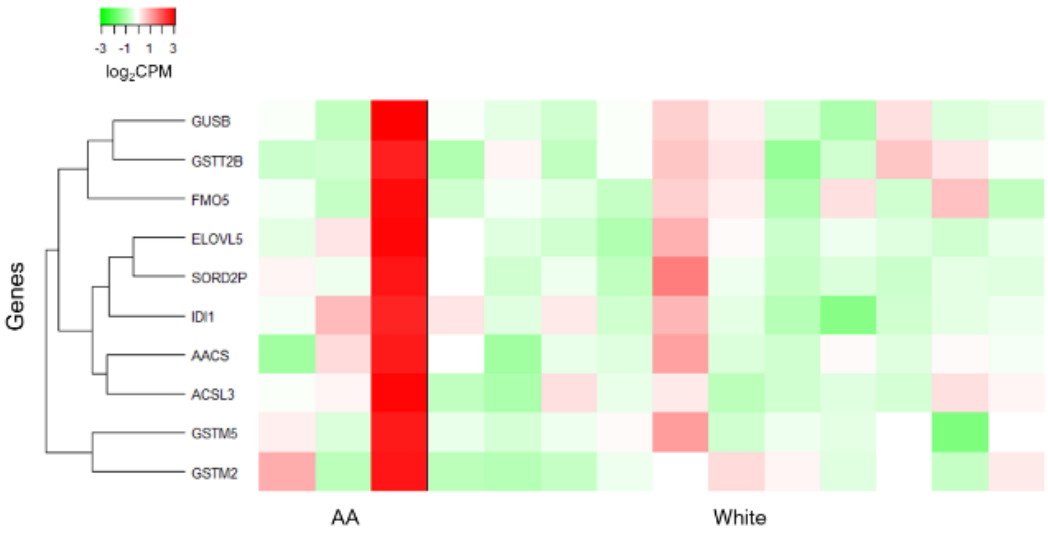


Figure 13 Heat map with gene cluster of the the differentially expressed genes in TNBC-AA patients with post-menopause condition and disease at stage I, comparing with matching TNBC-White patients. *AA* African-American. CPM counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

Most TNBC patients from the study cohort present stage II disease. DGEA of the contrast with premenopausal and stage II disease patients resulted in 63 DEG in TNBC-AA patients, comparing with TNBC-White patients (Table 18 and Figure 27b). Curiously, from the 63 DEG, only 3 are downregulated. Figure 29 shows the heat maps of the DEG in this group. Figure 29a corresponds to the gene clustering of the complete list of DEG, and Figure 29b corresponds to the top 20 most statistically significant DEG. Contrary to the contrast with postmenopausal stage I disease patients, with only 3 TNBC-AA patients (Figure 28), in Figure 29 we observed that the 63 DEG (Figure 29a) are less likely due to the massive up- or downregulation of those genes in a single patient. Nonetheless, and as previously discussed, the top 20 most significantly DEG (Figure 29b) *per se* may not be enough to point out some of the most biologically relevant genes in TNBC development and prognosis in premenopausal and stage II disease AA patients, comparing with their White counterparts.

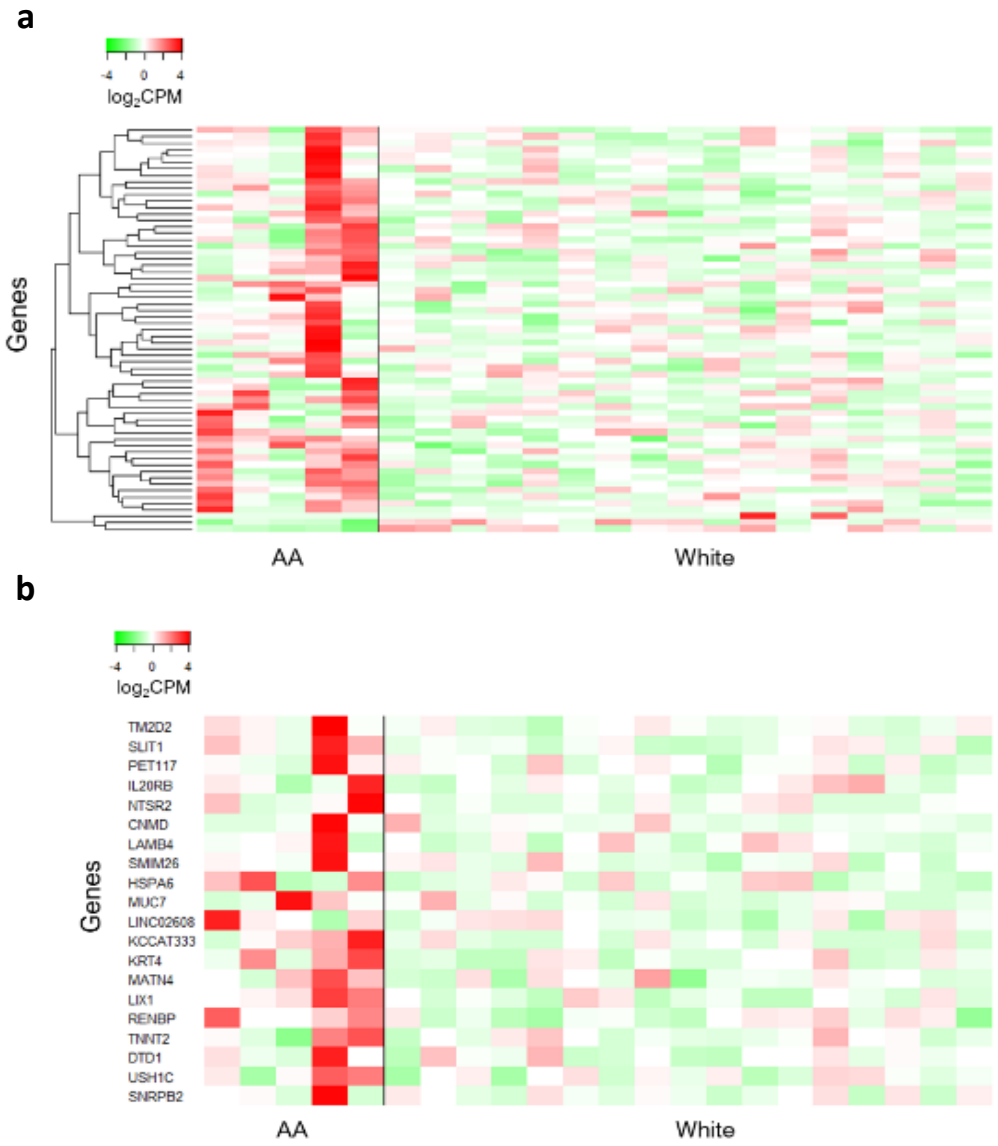


Figure 14 Heat maps of premenopausal stage II TNBC patients. **a** Clustering of all differentially expressed genes observed in TNBCs-AA patients vs TNBC-White patients. **b** Top 20 of the most differentially expressed genes identified in TNBC-AA patients vs TNBC-White patients, in ascending order of FDR. AA African-American. *CPM* counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

On the other hand, DGEA of the postmenopausal stage II disease contrast resulted in a higher number of DEG in TNBC-AA patients (1360, with 642 upregulated and 718 downregulated, Table 18 and Figure 27e) than in the contrast with all TNBC patients (1122, with 638 upregulated and 484 downregulated, Table 14), even though it has less patients included. Interestingly, although including a smallest group of patients, the analysis of this contrast lead to a larger number of DEG than the contrast of postmenopausal patients (718 DEG, with 32 AA patients and 57 White patients, Table 16). This observation suggests that the gene expression profile distance between TNBC-AA patients and TNBC-White patients in postmenopausal condition is more pronounced when patients also exhibit stage II disease.

Figure 30 shows the heat maps corresponding to gene clustering of the complete list of DEG (Figure 30a) and the top 20 most statistically significant DEG (Figure 30b) identified in postmenopausal stage II disease TNBC-AA patients, comparing with matching White patients. In Figure 30a, the considerably larger number of DEG identified in TNBC-AA patients allows for the visualization of a clear distinction in differential gene expression profile between the two groups of patients. However, given the large number of DEG obtained in this DGEA, the most statistically significant DEG (Figure 30b) may not include more relevant driver genes involved in ancestry-associated discrepancy of TNBC.

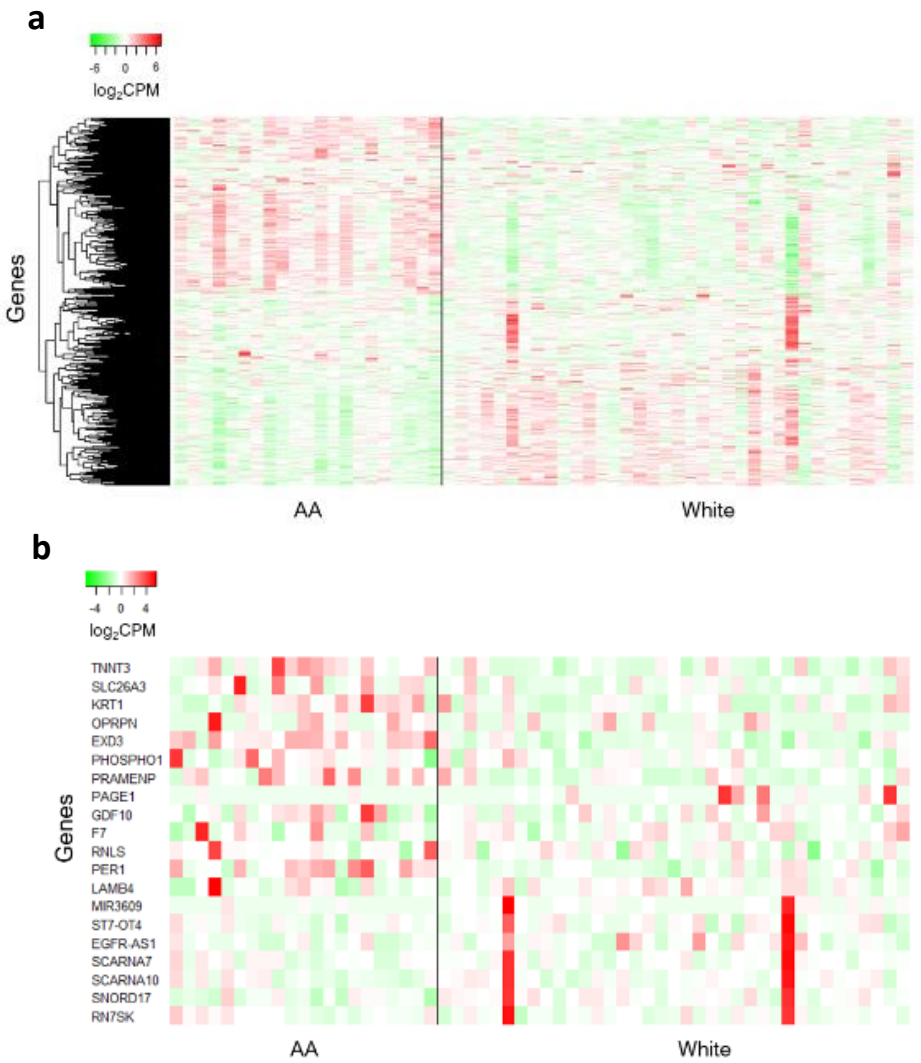


Figure 30 Heat maps of postmenopausal stage II TNBC patients. **a** Clustering of all differentially expressed genes observed in TNBCs-AA patients vs TNBC-White patients. **b** Top 20 of the most differentially expressed genes identified in TNBC-AA patients vs TNBC-White patients, in ascending order of FDR. *AA* African-American. *CPM* counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

Finally, both pre- and post-menopause contrasts with stage III disease patients have in common the fact that few patients are included, similarly to the contrasts involving stage I disease patients (Table 18). Particularly, the premenopausal patients with disease at stage III have the smaller number of patients (2 AA and 4 White) among all contrasts. However, there were two significantly upregulated

genes, *PIK3CA* and *ZMAT3* in premenopausal stage III disease AA-TNBC patients (Table 20). Due to the small number of obtained DEG, as well as the reduced number of patients, it was not possible to design heatmaps. Nonetheless, we identified differences in the expression profile of the two populations within this contrast.

Table 20 Differentially expressed genes in stage III disease premenopausal TNBC-AA patients comparing with matching TNBC-White patients. **a** - Gene description according to UniProtKB. **b** - Gene description according to Entrez Gene summary.

Genes	Description	Log_2FC	FDR
PIK3CA	*Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha*, promotes activation of signaling cascades involved in cell growth, survival, proliferation, motility and morphology, as well as cellular signaling in response to various growth factors.[a]	5.003	4.36E-08
ZMAT3	*Zinc Finger Matrin-Type 3*, the mRNA and the protein of this gene are upregulated by wildtype p53 and overexpression of this gene inhibits tumor cell growth, suggesting that this gene may have a role in the p53-dependent growth regulatory pathway.[b]	3.251	2.83E-02

Curiously*, PIK3CA* and *ZMAT3* are in close proximity in the cytogenetic location 3q26.32. This suggests that the 2 premenopausal AA patients with stage III disease might have a duplication of that chromosomic region, causing the amplification of gene expression. This might be an interesting hypothesis, especially considering that PI3K pathway is involved in drug resistance and poor outcome of BC patients (Aleskandarany et al., 2010; Berns et al., 2007).

Although the contrasts including postmenopausal and stage I disease (Table 19, Figure 28) and premenopausal and stage III disease (Table 20) patients resulted in a small number of DEG, the ones identified are suggestively associated with drug-resistance and bad prognosis.

Finally, the postmenopausal, stage III disease contrast also resulted in two DEG, *KCND2* and *PLAU,* both being downregulated in TNBC-AA patients (Table 21). But, taking in consideration the literature, *KCND2* and *PLAU* do not have any reported involvement in cancer development or with each other. Again, no heatmaps were design for this contrast due to the small number of resulting DEG.

Table 21 Differentially expressed genes in stage III disease postmenopausal TNBC-AA patients comparing with matching TNBC-White patients. a - Gene description according to UniProtKB.

Genes	Description	Log_2FC	FDR
KCND2	*Potassium Voltage-Gated Channel Subfamily D Member 2*, a protein that mediates transmembrane potassium transport in excitable membranes, primarily in the brain.[a]	-5.384	1.75E-02
PLAU	*Plasminogen Activator, Urokinase*, a protein that specifically cleaves the zymogen plasminogen to form the active enzyme plasmin.[a]	-3.288	2.34E-02

Although DGEA with small sampling is possible, such as the cases of premenopausal stage I disease and premenopausal stage III disease contrasts, which have two AA patients each (Table 18), a descriptive analysis with respective fold change of the identified DEG in the contrast is the more appropriate way to interpret these results, as recommended by the *edgeR*'s user's guide (Chen et al.,

2018). For this reason, we did not proceed with further complementary analysis to interpret the DEG identified in these contrasts, in a biological context in African-ancestry TNBC patients.

To conclude the DGEA task of this project, we also identified which DEG (FDR≤0.05) are in common in DGEA with patients with the same or different menopause and disease staging conditions, i.e., genes that are more specific to a menopause status and/or TNBC staging. The complete list of matching DEG in different DGEA can be consulted in Supplementary - Matching DEG file. Figure 31 shows the number of matching DEG between different contrasts, with the exception of the contrasts that resulted in 2 or less DEG, as well as the total number of significant DEG in each contrast. As expected, since patients are not mutually exclusive among the contrasts, stage II contrast generated the larger number of DEG in common with other contrasts, 474, and, of those, 453 are in common with the contrasts with all TNBC patients, postmenopausal patients and patients both postmenopausal and presenting stage II disease. Given that such DEG are shared among a large portion of AA patients, with the most common variables in study (post-menopause status and stage II TNBC at diagnosis), they may be interesting candidate ancestry-associated markers.

Overall, the sets of DEG found in TNBC-AA patients, comparing with TNBC-White patients, show a clear difference in expression profile among these populations. Furthermore, these results suggest that differential expression is indeed influenced by the menopause status and pathological stage of the disease, as different gene expression profiles were found in patients with distinct menopause status or distinct disease stage. These findings have a potential clinical impact in the prognosis and treatment of African-ancestry patients.

Nonetheless, due to the small number of patients of some contrasts, such as those including pre- and postmenopausal patients with stage III disease (Table 18), to perform extrapolations of the results obtained within these contrasts into clinical practice, further investigation with a larger number of patients presenting those characteristics are still required. Also, stage I contrasts resulted in null or reduced sets of DEG. This may be due to the fact that TNBC gene expression profile is progressively more distant between the two populations throughout disease development, as previously reported (Huo et al., 2017; Stewart et al., 2013).

Since DGEA results showed that there are indeed gene expression ancestry-associated differences to justify TNBC's epidemiological and clinical observations, we proceeded to refine DGEA results, in order to find potential leading molecules involved in the ancestry-associated TNBC discrepancy.

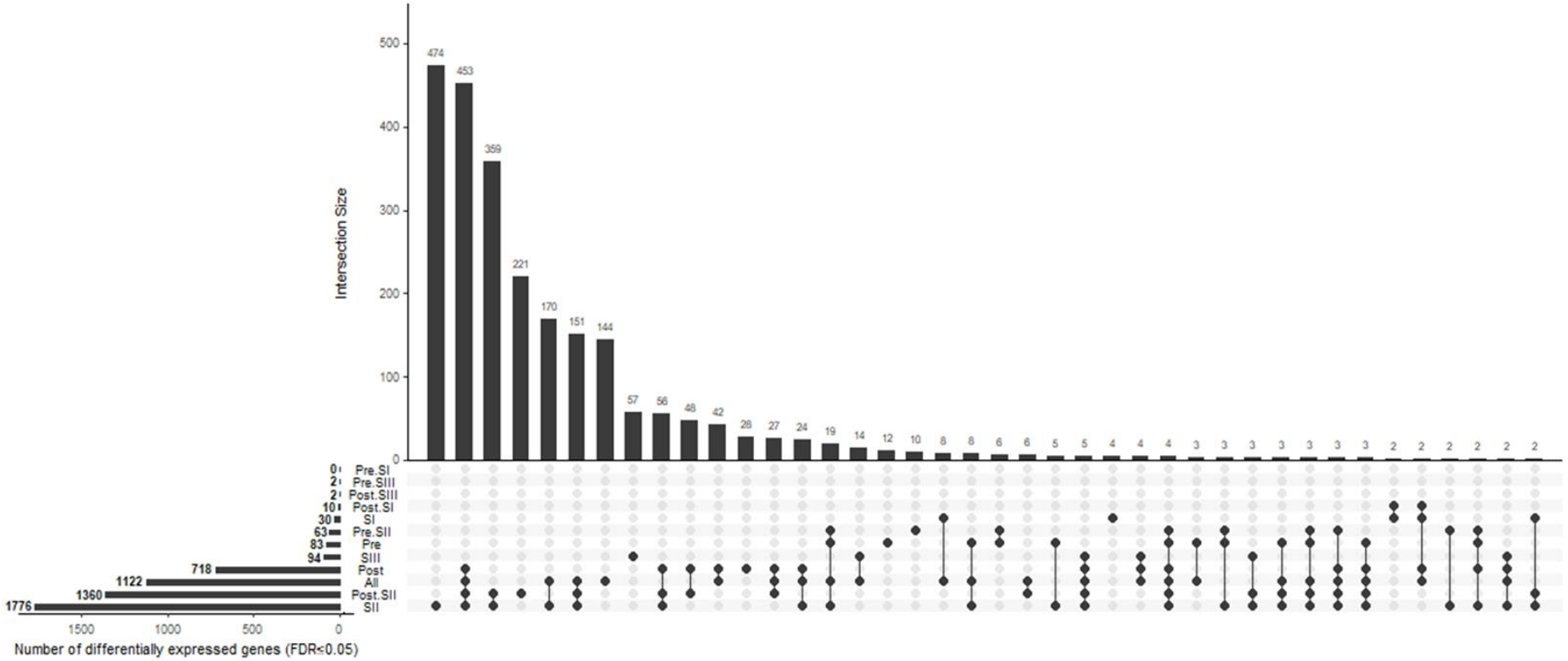


Figure 31 UpSet plot of the differentially expressed genes (FDR≤0.05) idenfied in TNBC-AA patients comparing with TNBC-White patients, obtained in the different differential gene expression analyses. The intersection size represents the number of matching genes. *All* DGEA considering all TNBC patients (52 AA vs 89 White). *Pre* DGEA considering premenopausal patients (9 AA vs 29 White). *Post* DGEA considering postmenopausal patients (32 AA vs 57 White). *SI* DGEA considering stage I disease patients (9 AA vs 19 White). *SII* DGEA considering stage II disease patients (32 AA vs 59 White). *SIII* DGEA considering stage III disease patients (10 AA vs 12 White). *PreSI* DGEA considering premenopausal stage I disease patients (2 AA vs 7 White). *PreSII* DGEA considering premenopausal stage II disease patients (5 AA vs 17 White). *PreSIII* DGEA considering premenopausal stage III disease patients (2 AA vs 4 White). *PostSI* DGEA considering postmenopausal stage I disease patients (3 AA vs 11 White). *PostSII* DGEA considering postmenopausal stage II disease patients (22 AA vs 36 White). *PostSIII* DGEA considering postmenopausal stage III disease patients (7 AA vs 6 White). *AA* - African-American.

3. GENE SET ENRICHMENT ANALYSIS AND INGENUITY PATHWAY ANALYSIS HIGHLIGHT PATHWAYS AND GENES DIFFERENTIALLY EXPRESSED IN TNBC-AA PATIENTS

To shed some light on interesting candidate ancestry-associated markers to be further investigated in the validation cohort, the following analyses were performed: gene set enrichment analysis (GSEA), using collections of diseases and pathways from Kyoto Encyclopedia of Genes and Genomes (KEGG), and Ingenuity Pathway Analysis (IPA)'s core analysis, with focus on the network design, to aid visualize in what way the DEG of a given contrast interact and which genes have a central role in the altered networks. Both GSEA and IPA's network design were performed using as input the complete gene lists obtained from DGEA results, including non-DEG and the respective FC and FDR of each gene.

3.1. Gene set enrichment analysis

With the exception of DGEA results obtained in pre- and postmenopausal stage I and stage III disease contrasts, all the other DGEA results were subjected to GSEA. Gene sets from KEGG Pathway, a curated database of molecular pathways and disease signatures (Kanehisa et al., 2017, 2016; Kanehisa and Goto, 2000), were considered enriched when FDR≤0.05, taking into consideration both up- and downregulated genes in the set (FDR.mixed). The following contrasts resulted in enriched gene sets (Supplementary - GSEA file, eTables 1, 3, 5 and 11, respectively): all TNBC patients, postmenopausal patients regardless the disease stage, postmenopausal patients independently of disease stage, patients diagnosed with stage II disease regardless of menopause status and postmenopausal stage II disease patients. Complete lists of GSEA's results can be consulted in Supplementary - GSEA file.

In order to focus on genes that may have greater impact on TNBC development and in the prognosis and treatment of African-ancestry patients, the following gene set KEGG collections were considered: signal transduction, cellular community - eukaryotes and cell motility, cell growth and death, cancer: overview, where the BC gene set was included, as well as the drug resistance - antineoplastic collection, women-specific endocrine system collection gene sets, and gene sets related to immuno oncology, from the immune system collection.

The enriched gene sets in TNBC-AA patients, comparing with matching TNBC-White patients, are displayed in Figure 32. Green color represents gene sets with a global negative net direction (Down) and red color represent gene sets with a global positive net direction (Up), with color intensity proportional to $-\log_2(\text{FDR})$. As mentioned in Chapter III, Section 1.4., the net direction of an enriched gene set in a given contrast is influenced by gene interactions in that pathway or disease and by the FC and FDR obtained in the DGEA, and does not mean that the majority of the genes in that enriched gene set are up- or downregulated.

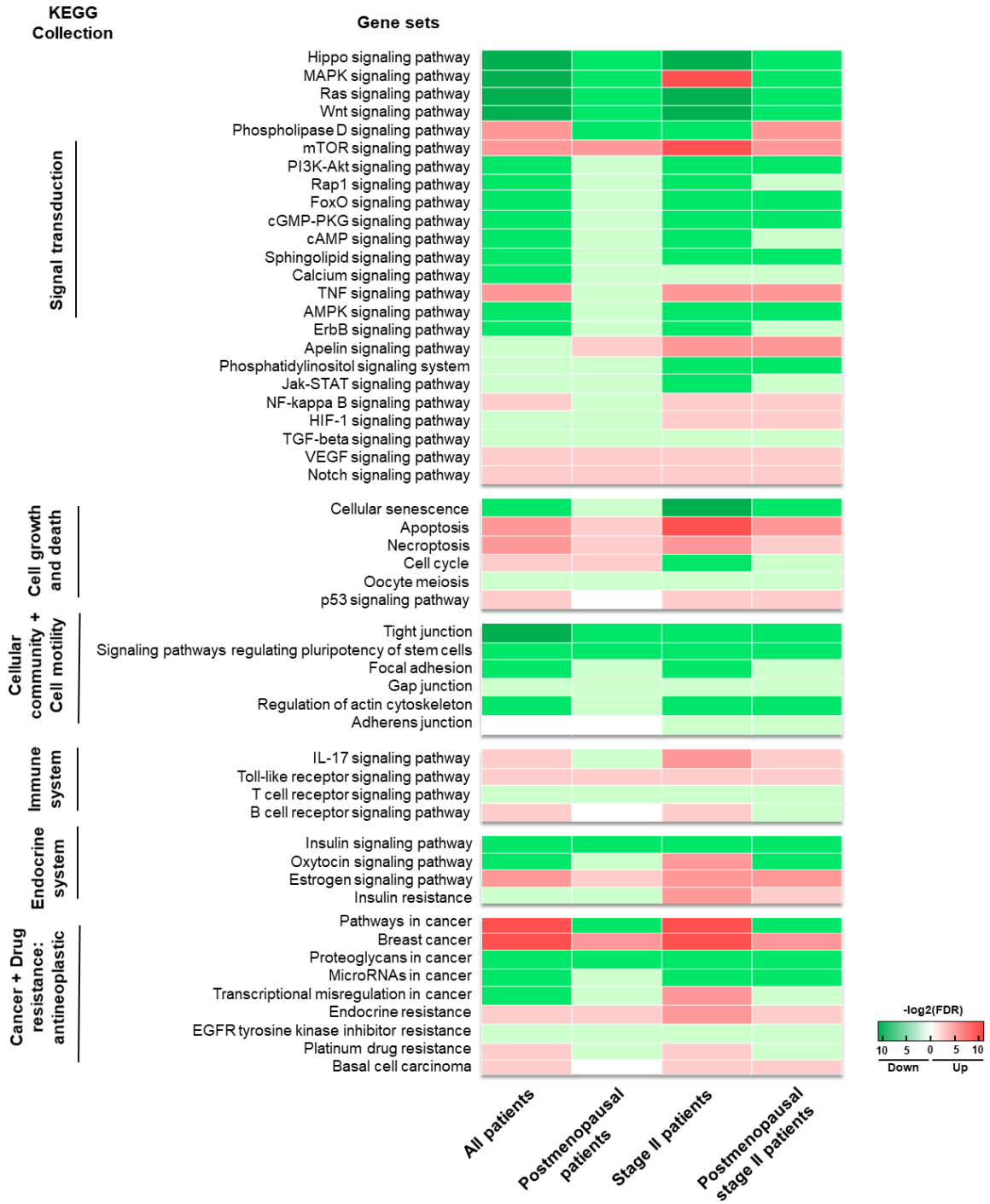


Figure 32 Enriched gene sets in African-American patients. Each column corresponds to a contrast, in the following order: all TNBC patients, postmenopausal patients, stage II disease patients and postmenopausal stage II disease patients. Green color - negative net direction (Down). Red color - positive net direction (Up). White color - not statistically significant (FDR≥0.05). Color intensity is proportional to -$\log_2$(FDR) value.

Some enriched gene sets' share the same net direction between the contrasts, with differences being at the level of FDR value. This was somewhat expected, considering that the different analyzed groups include patients mainly at post-menopause status and/or with stage II disease. On the other hand, opposite net direction of the enriched gene sets in a specific contrast suggests that menopause status and pathological stage of the disease indeed influence specific pathways and cellular

mechanisms. Additionally, enriched gene sets with consistent net directions between these groups suggest that those pathways may contribute to TNBC development in AA patients independently of their menopause status or disease stage.

Overall, most enriched gene sets are negatively regulated in TNBC-AA patients, comparing with matching White patients. Nevertheless, positively regulated gene sets were seen among all the analyzed groups, and include mTOR, VEGF and Notch signaling pathways, Apoptosis and Necroptosis cell processes, Toll-like signaling pathway, Estrogen signaling pathway, Breast cancer and Endocrine resistance.

In the Signal transduction collection, postmenopausal patients and stage II disease patients revealed the most distinct differences regarding the net direction of the enriched gene sets. For instance, MAPK pathway is exclusively positively regulated in AA patients from the contrast composed by patients with stage II disease, comparing with their White counterparts. On contrary, TNF, NF-κB and HIF-1 pathways appeared to be negatively regulated in AA patients presenting postmenopausal condition, regardless the disease stage. Regarding Cell growth and death collection, Cell cycle gene set is positively regulated in AA patients from the contrasts with all TNBC patients, independently of menopause status and stage of the disease, and in the postmenopausal contrast. Gene sets from the Cellular community and motility group have consistently a negative net direction in all four contrasts. In the Immune system collection, IL-17 signaling pathway is negatively regulated in postmenopausal AA patients, and B cell receptor signaling pathway is negatively regulated in AAs from the contrast composed by postmenopausal stage II disease patients. In the Endocrine system collection, Oxytocin signaling pathway is exclusively positively regulated in AA patients from the stage II disease contrast, independently of their menopause status; also, Insulin resistance gene set is positively regulated in AAs from the postmenopausal status and stage II disease contrast and in the contrast comprised by stage II disease patients, independently of menopause status. Finally, in the Cancer and antineoplastic drug resistance collection, Pathways in cancer, Transcriptional misregulation in cancer and Platinum drug resistance gene sets are positively regulated in stage II AA patients, being negatively regulated in AAs in the postmenopausal contrast.

Of all the enriched gene sets and respective net direction in common in the four groups of TNBC-AA patients, here we highlight “Hippo signaling pathway” (Figure 32). This gene set stands out by having the lowest FDR value and by being negatively regulated in AA patients, comparing with matching White patients, in all four contrasts. Figure 33 shows the DEG of this gene set identified in each contrast, as well as its respective FC. Different DEG contribute to the enrichment of the Hippo signaling pathway gene set in each of these four contrasts. *FZD8*, *MYC*, *SCRIB*, *DVL1*, *CCND3* and *LATS1* are the most consistent DEG, being identified in the four contrasts, and contributing to the enrichment of this gene set. According to the literature, downregulation of Hippo pathway components, including *LATS1*, may promote the transcription of genes involved in cell proliferation and competition, cell death inhibition, epithelial-to-mesenchymal transition (EMT), tumor metastasis, and tumorigenesis (Harvey et al., 2013; Lei et al., 2008; Marti et al., 2015; Piccolo et al., 2014; Zanconato et al., 2016; Zhao et al., 2010).

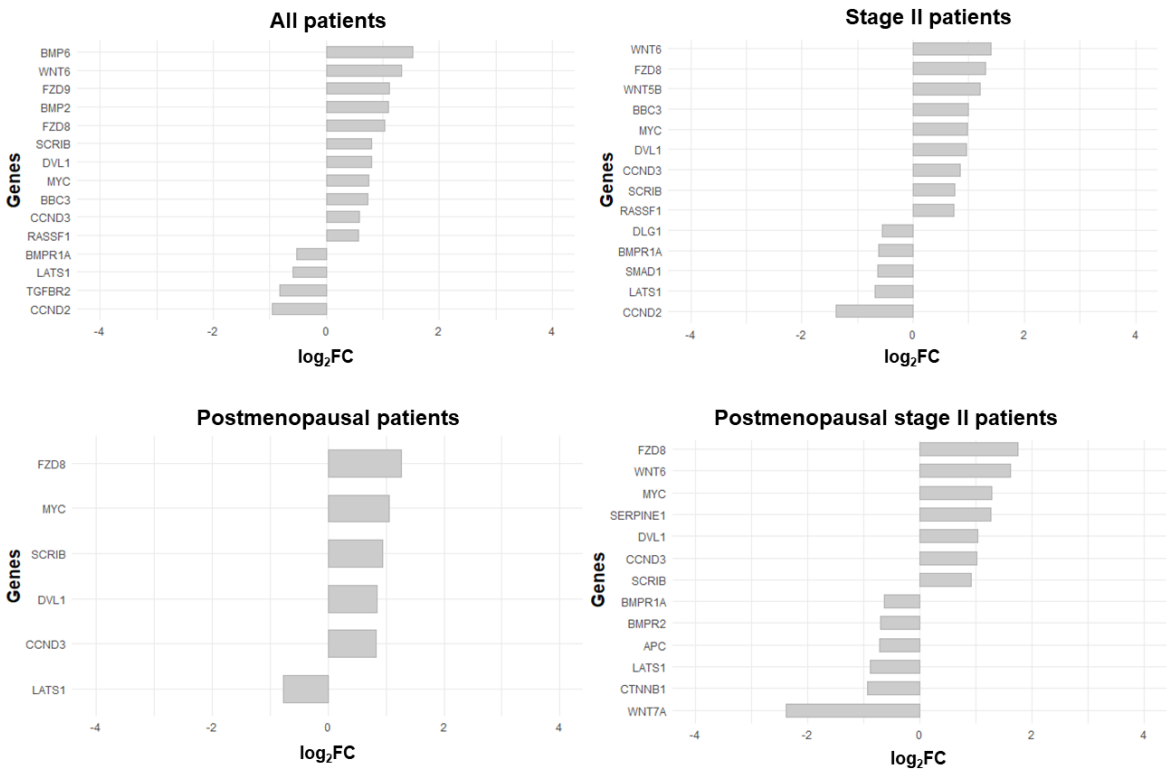
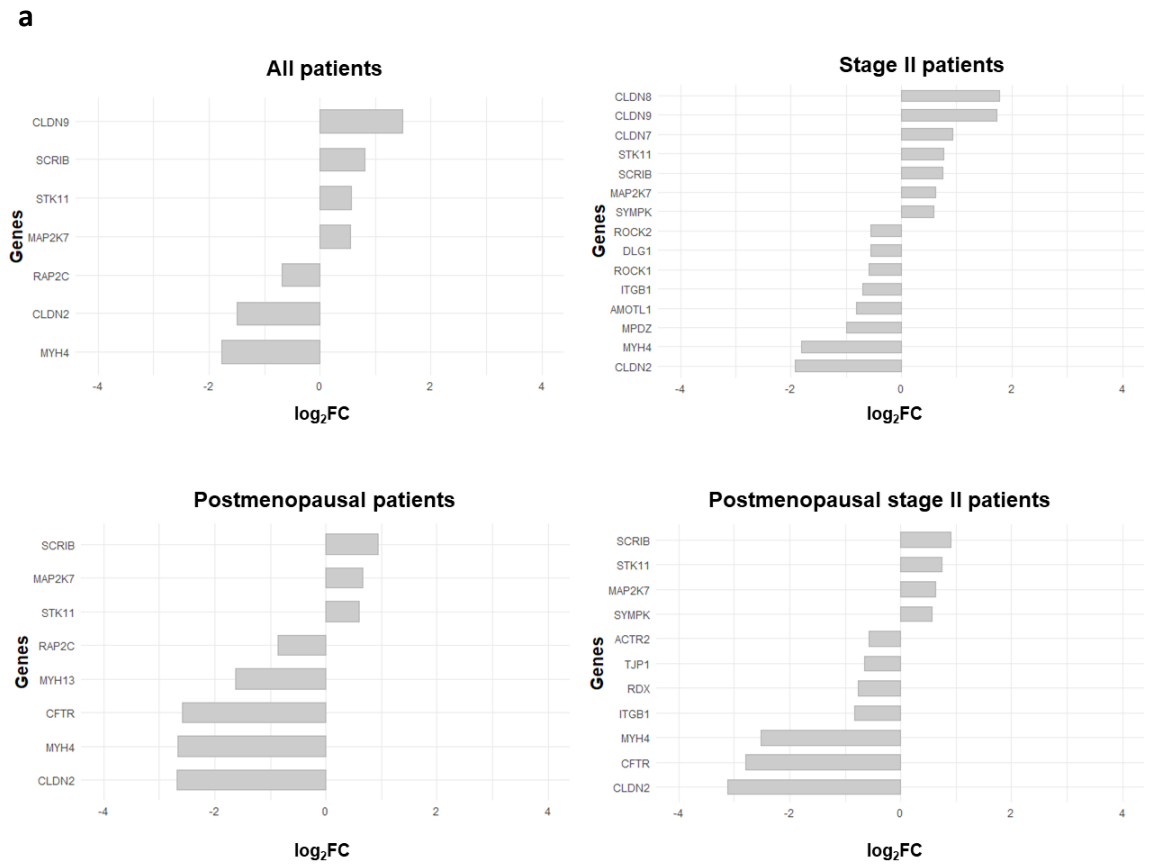
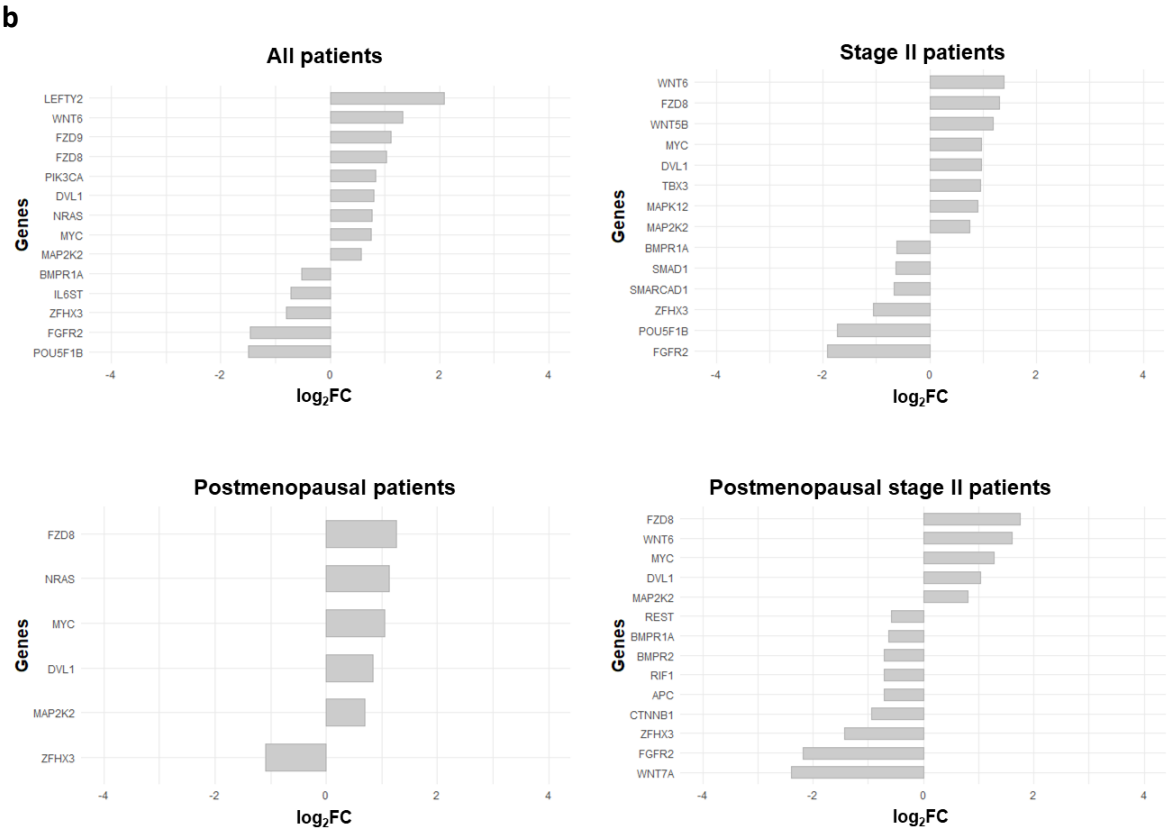


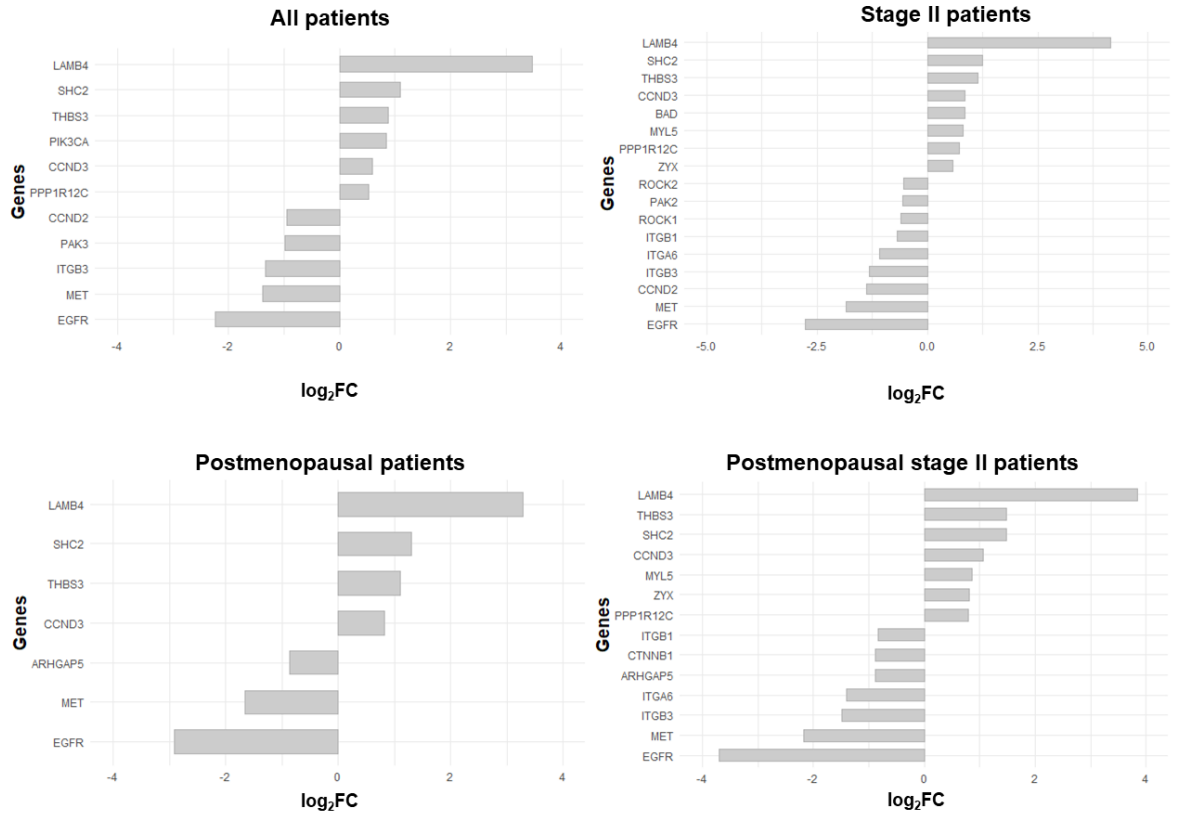
Figure 33 Differentially expressed genes identified in the Hippo signaling pathway gene set from the Signal transduction collection in the following groups of patients: all TNBC patients, postmenopausal patients, stage II disease patients and postmenopausal and stage II disease patients. *FC* Fold-change.

We also highlight the cellular community and cell motility gene set collection as having the most consistent negative net direction in all contrasts (Figure 32). This consistent downregulation of the cellular community and cell motility gene sets suggests that the integrity of cell-cell contacts and actin cytoskeleton organization, fundamental in epithelial tissue homeostasis, are impaired (Bhat et al., 2019; Cavallaro and Christofori, 2004), promoting tumor cells dissociation and subsequent metastasis (Bogenrieder and Herlyn, 2003). Figure 34 displays the DEG of the enriched cellular community and cell motility gene sets in each group of patients.

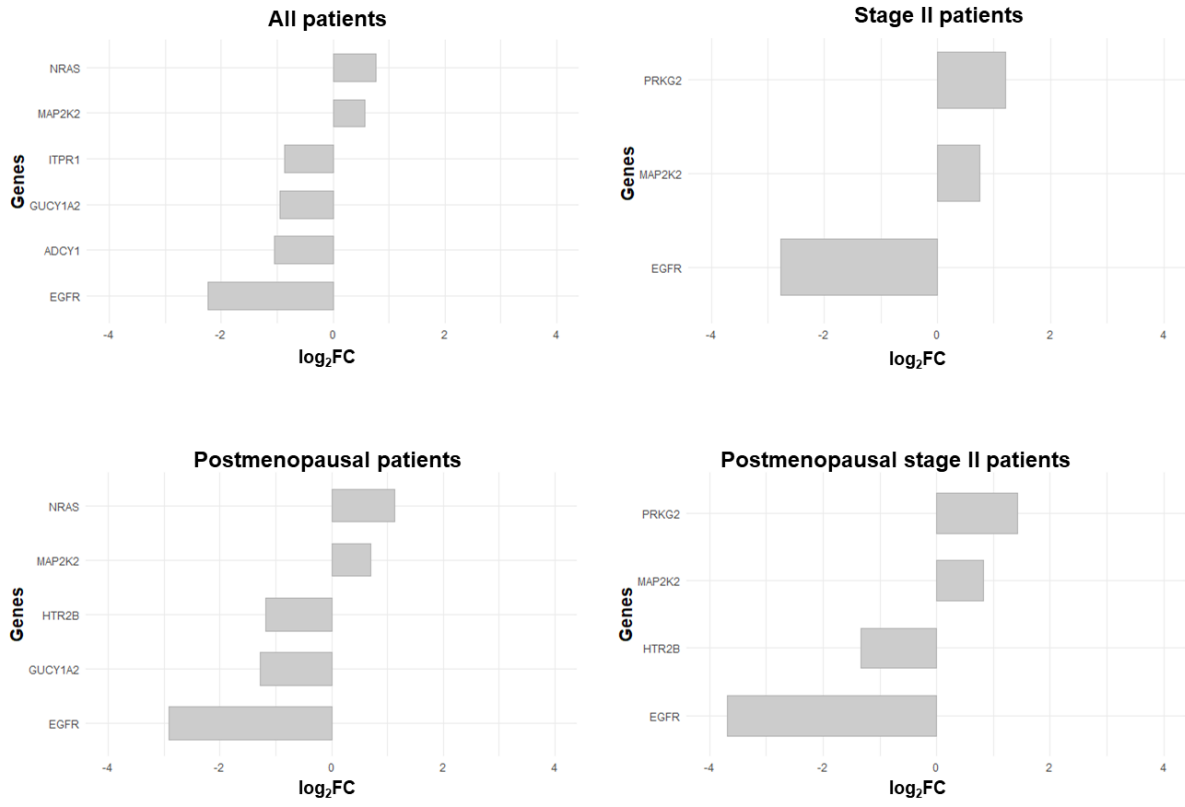

a
All patients
CLDN9
SCRIB
STK11
MAP2K7
RAP2C
CLDN2
MYH4
Genes
-4
-2
0
2
4
log2FC
Stage II patients
CLDN8
CLDN9
CLDN7
STK11
SCRIB
MAP2K7
SYMPK
ROCK2
DLG1
ROCK1
ITGB1
AMOTL1
MPDZ
MYH4
CLDN2
Genes
-4
-2
0
2
4
log2FC
Postmenopausal patients
SCRIB
MAP2K7
STK11
RAP2C
MYH13
CFTR
MYH4
CLDN2
Genes
-4
-2
0
2
4
log2FC
Postmenopausal stage II patients
SCRIB
STK11
MAP2K7
SYMPK
ACTR2
TJP1
RDX
ITGB1
MYH4
CFTR
CLDN2
Genes
-4
-2
0
2
4
log2FC


b
All patients
LEFTY2
WNT6
FZD9
FZD8
PIK3CA
DVL1
NRAS
MYC
MAP2K2
BMPR1A
IL6ST
ZFHX3
FGFR2
POU5F1B
Genes
-4
-2
0
2
4
log2FC
Stage II patients
WNT6
FZD8
WNT5B
MYC
DVL1
TBX3
MAPK12
MAP2K2
BMPR1A
SMAD1
SMARCAD1
ZFHX3
POU5F1B
FGFR2
Genes
-4
-2
0
2
4
log2FC
Postmenopausal patients
FZD8
NRAS
MYC
DVL1
MAP2K2
ZFHX3
Genes
-4
-2
0
2
4
log2FC
Postmenopausal stage II patients
FZD8
WNT6
MYC
DVL1
MAP2K2
REST
BMPR1A
BMPR2
RIF1
APC
CTNNB1
ZFHX3
FGFR2
WNT7A
Genes
-4
-2
0
2
4
log2FC

c


All patients
Genes
LAMB4
SHC2
THBS3
PIK3CA
CCND3
PPP1R12C
CCND2
PAK3
ITGB3
MET
EGFR
log2FC
Stage II patients
LAMB4
SHC2
THBS3
CCND3
BAD
MYL5
PPP1R12C
ZYX
ROCK2
PAK2
ROCK1
ITGB1
ITGA6
ITGB3
CCND2
MET
EGFR
log2FC
Postmenopausal patients
LAMB4
SHC2
THBS3
CCND3
ARHGAP5
MET
EGFR
log2FC
Postmenopausal stage II patients
LAMB4
THBS3
SHC2
CCND3
MYL5
ZYX
PPP1R12C
ITGB1
CTNNB1
ARHGAP5
ITGA6
ITGB3
MET
EGFR
log2FC

d


All patients
Genes
NRAS
MAP2K2
ITPR1
GUCY1A2
ADCY1
EGFR
log2FC
Stage II patients
PRKG2
MAP2K2
EGFR
log2FC
Postmenopausal patients
NRAS
MAP2K2
HTR2B
GUCY1A2
EGFR
log2FC
Postmenopausal stage II patients
PRKG2
MAP2K2
HTR2B
EGFR
log2FC

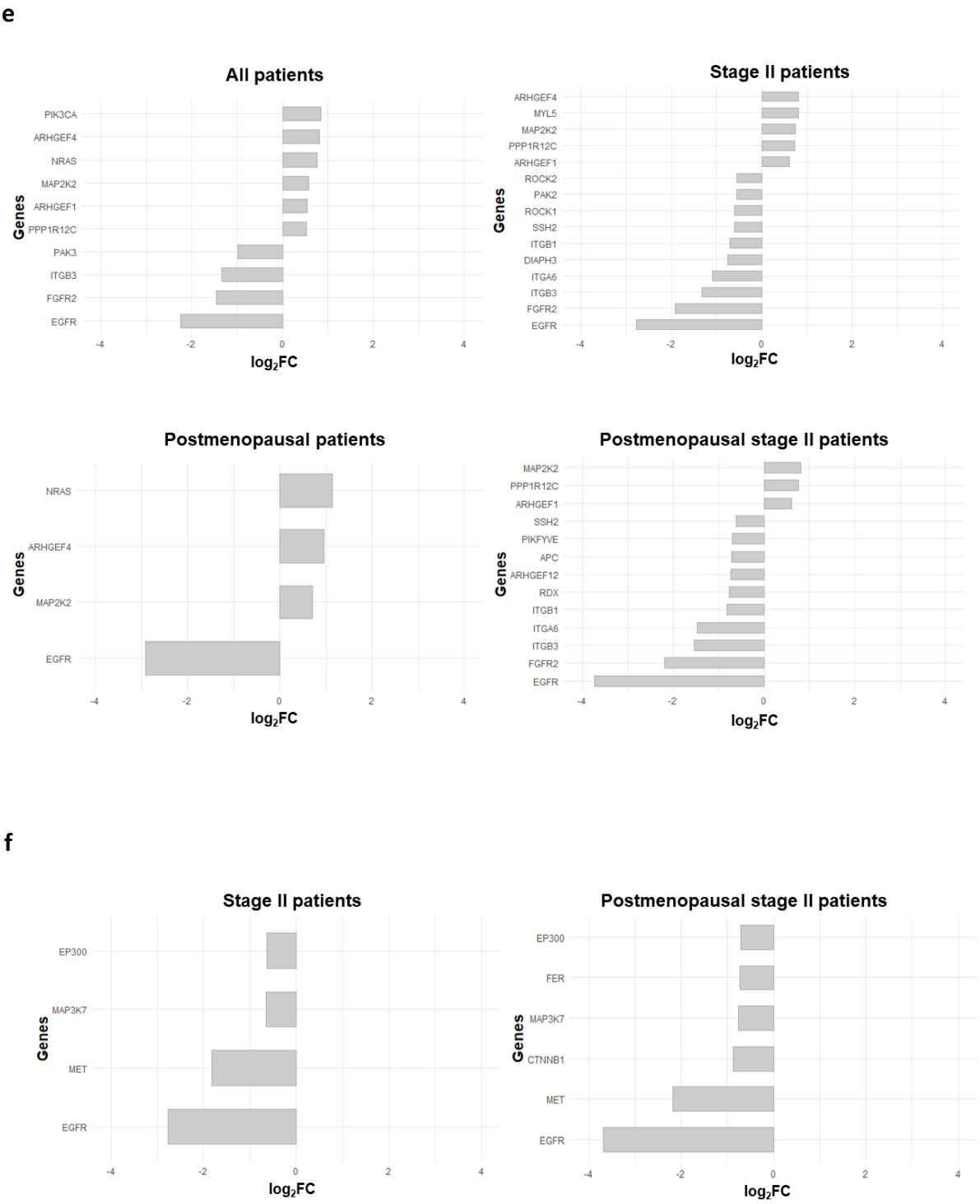


Figure 34 Differentially expressed genes identified in highlighted enriched KEGG gene sets from the group of Cellular community and Cell motility collection (FDR ≤0.05) in the following contrasts: all TNBC patients, postmenopausal patients, stage II disease patients and postmenopausal and stage II disease patients. **a** Tight junction gene set; **b** Signaling pathways regulating pluripotency of stem cells gene set; **c** Focal adhesion gene set; **d** Gap junction gene set; **e** Regulation of actin cytoskeleton gene set; **f** Adherens junction gene set. *FC* Fold-change.

We also highlight BC gene set as being consistently upregulated in AA patients in all contrasts (Figure 32). BC gene set includes, among others, *EGFR*, *MYC* or *CTNNB1* genes (Figure 35), and the fact that is upregulated in AA patients, comparing with White patients, suggests that AA women have a higher activation of components and processes involved in BC development, which may be translated into TNBC's faster progression and aggressive behavior.

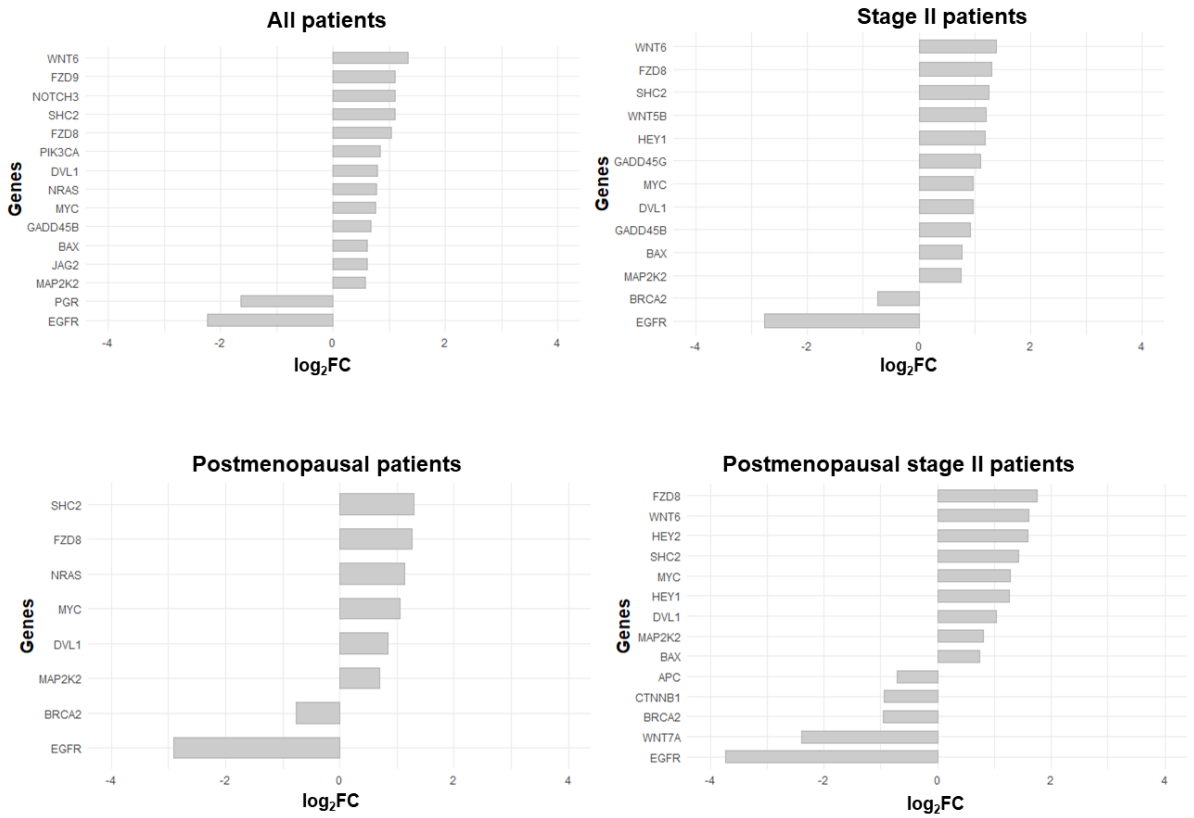


Figure 35 Differentially expressed genes identified in Breast cancer gene set from the Cance collectionr: specific types collection, in the following groups of patients: all TNBC patients, postmenopausal patients, stage II disease patients and postmenopausal and stage II disease patients. *FC* Fold-change.

Lastly, we highlight MAPK signaling pathway gene set which, interestingly, is exclusively upregulated in stage II disease AA patients (Figure 32). MAPK cascades regulate a variety of cellular processes, including proliferation, differentiation, apoptosis and stress responses, playing a crucial role in the survival and development of cancer cells (Guo et al., 2020). Thus, DEG involved in MAPK pathway exclusively identified in stage II AA patients (Figure 36), including *NTRK1*, *GADD45C*, *MAPK12*, *TRADD* or *FGFR2*, may contribute to the observed worst prognosis of TNBC-AA patients when comparing with matching TNBC-White patients.

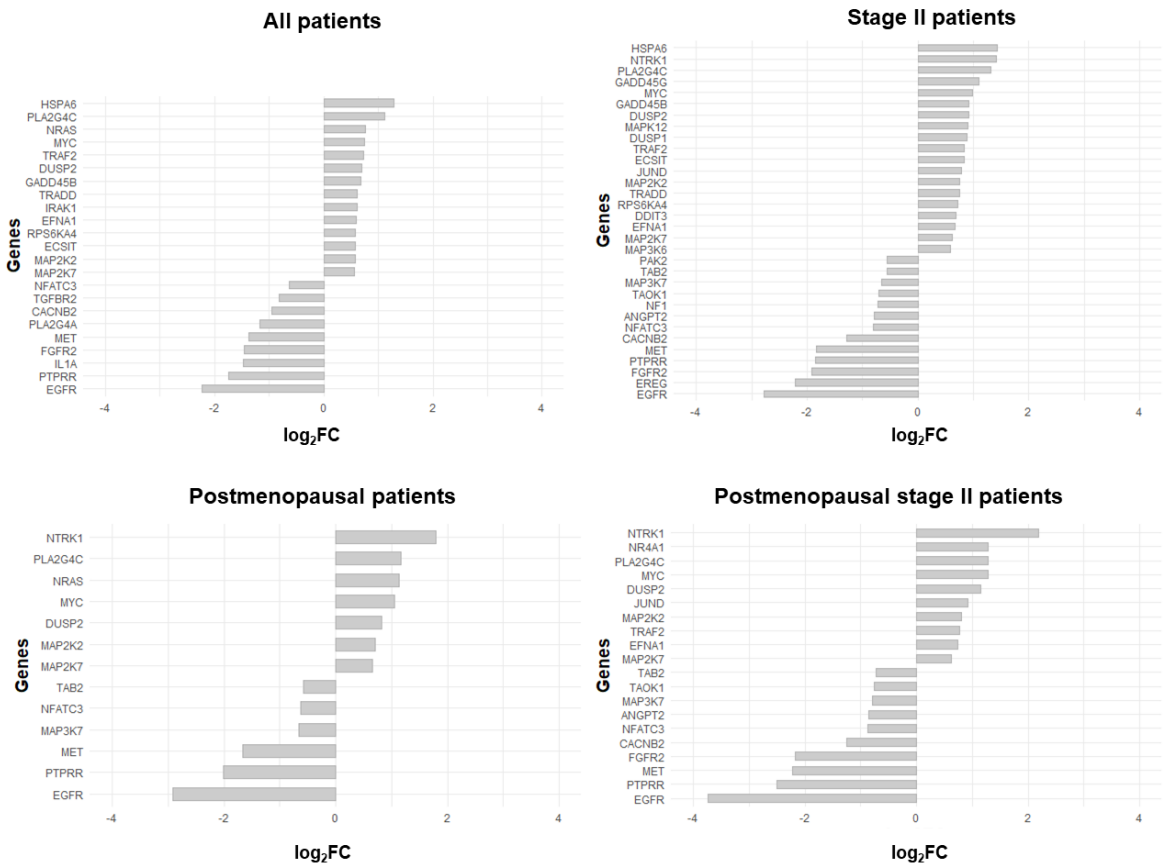


Figure 36 Differentially expressed genes identified in MAPK signaling pathway gene set from the Signal transduction collection, in the following groups of patients: all TNBC patients, postmenopausal patients, stage II disease patients and postmenopausal and stage II disease patients. *FC* Fold-change.

GSEA provides a method for the identification of the pathways and/or cellular mechanisms that are over- or under- represented considering the sets of genes resulting from the DGEA of the different contrasts. To complement GSEA results, and to help to identify potential driving molecules involved in the ancestry-associated TNBC discrepancy we also performed IPA's core analysis with focus on network design and identification of central network regulators.

3.2. Ingenuity Pathway Analysis

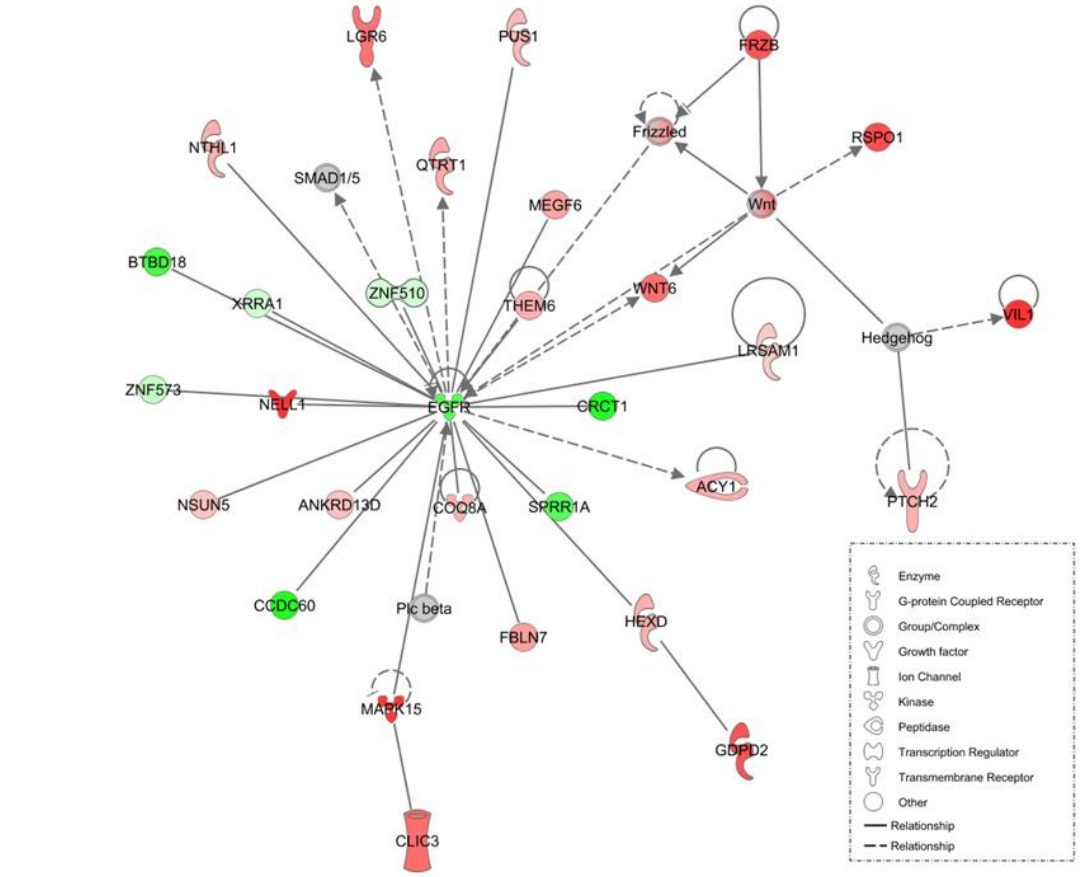
IPA's core analysis with focus on network design allows the visualization of the interactions between genes resulting from DGEA, displaying how they might work together at molecular level and which is the central network regulator. Central network regulators are the DEG with more direct and indirect connections in their network and highly interconnected networks are likely to represent significant biological functions. Only networks with a score above 30 were considered. Network scoring is based on the right-tailed Fisher's Exact test p-value calculation and they are scored based on the number of molecules from the inputted data set they contain. The higher the score, the lower the probability of finding the observed molecules from the data set in the designed network by random chance. In the context of this project, central network regulators of highly scored networks are potential drivers for ancestry-associated TNBC differences.

IPA's network design was performed using the complete lists of genes obtained through DGEA for each group, except pre- and postmenopausal stage I and III disease, due to their low number of DEG (Table 18). The complete lists of genes and respective FC and FDR were used as input. Complete lists of networks and respective central regulators, scores, molecules involved, and top diseases and functions, according to patient's menopause status and/or staging, are listed in Supplementary - Networks file.

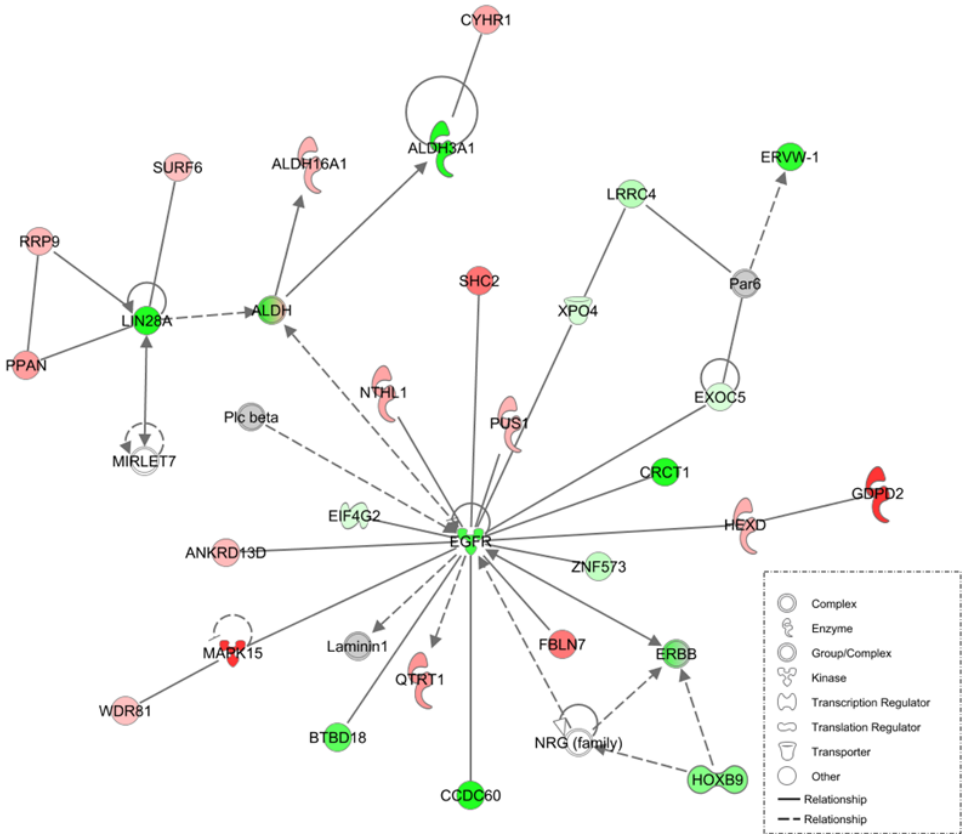
IPA's network design, through central regulator identification and network scoring, successfully recognized EGFR, Myc and Bcl2 as potential candidate ancestry-associated markers, as follows:

EGFR as a central regulator had among the highest scores and was observed in the following contrasts: all TNBC patients, postmenopausal patients regardless the stage of the disease, stage II disease patients regardless their menopause status and postmenopausal stage II disease patients (Supplementary - Networks file, eTables 1, 3, 5 and 7, respectively). Figure 37 represents the networks in which EGFR is a central regulator. EGFR is a transmembrane glycoprotein, member of the protein kinase superfamily, and it is a receptor for members of the epidermal growth factor family. EGFR activation promotes cell proliferation, motility, and survival via activation of various downstream signaling pathways, such as Ras-Raf-MEK-ERK, PI3K-AKT-mTOR, and Src-STAT3 (Yarden and Sliwkowski, 2001). Figure 37a represents the EGFR network in AA patients from the contrast comprised by all TNBC cases, independently of patients' menopause status or pathological stage of the disease, with a score of 40 and 30 focus molecules included. Here we observed that low EGFR expression seems to be indirectly affecting Wnt signaling pathway members, such as RSPO1, WNT6 or Frizzled class receptors, which are upregulated in this group of patients, as indicated by the red color. Interestingly, Wnt signaling pathway gene set has a negative net direction (Figure 32), which may be due to strong effect that the under-expression of EGFR may be causing. Additionally, according to IPA, this network is potentially involved in the following diseases and functions: Cancer, Cellular Movement, and Gastrointestinal Disease (Supplementary - Networks file, eTable 1). Figure 37b shows the EGFR network in postmenopausal TNBC-AA patients and has a score of 41 and 29 focus molecules. In this network we highlight that transcriptional regulators involved in embryonic development and cancer stemness maintenance, including LIN28A and ALDH proteins, are indirectly negatively regulated by the under-expressed EGFR. Furthermore, according to IPA, this network is involved in the following diseases and functions: Amino Acid Metabolism, Cellular Movement, and Connective Tissue Disorders (Supplementary - Networks file, eTable 3). Figure 37c displays the EGFR network in TNBC-AA patients diagnosed with stage II disease and has a score of 42 and 34 focus molecules. To our best knowledge, most molecules included in this network, such as TRIML2, GDPD2 or GRIK4, have unknown or poorly impacting roles in cancer development. Also, according to IPA, this network is involved in the following diseases and functions: Cellular Movement, Developmental Disorder, and DNA Replication, Recombination and Repair (Supplementary - Networks file, eTable 5). Finally, Figure 37d represents the EGFR network in the postmenopausal stage II disease AA patients, with a score of 49 and 35 focus molecules. Here, most of the molecules included in the design of this network, mainly those with larger log_2FC value (such as *MTRNR2L1*, *GDPD2* or *UNC80*), to our best knowledge, do not have known roles in cancer development. Also, according to IPA, this network is mainly involved with the following diseases and functions: Amino Acid Metabolism, Cellular Movement, and Connective Tissue Disorders (Supplementary - Networks file, eTable 7).

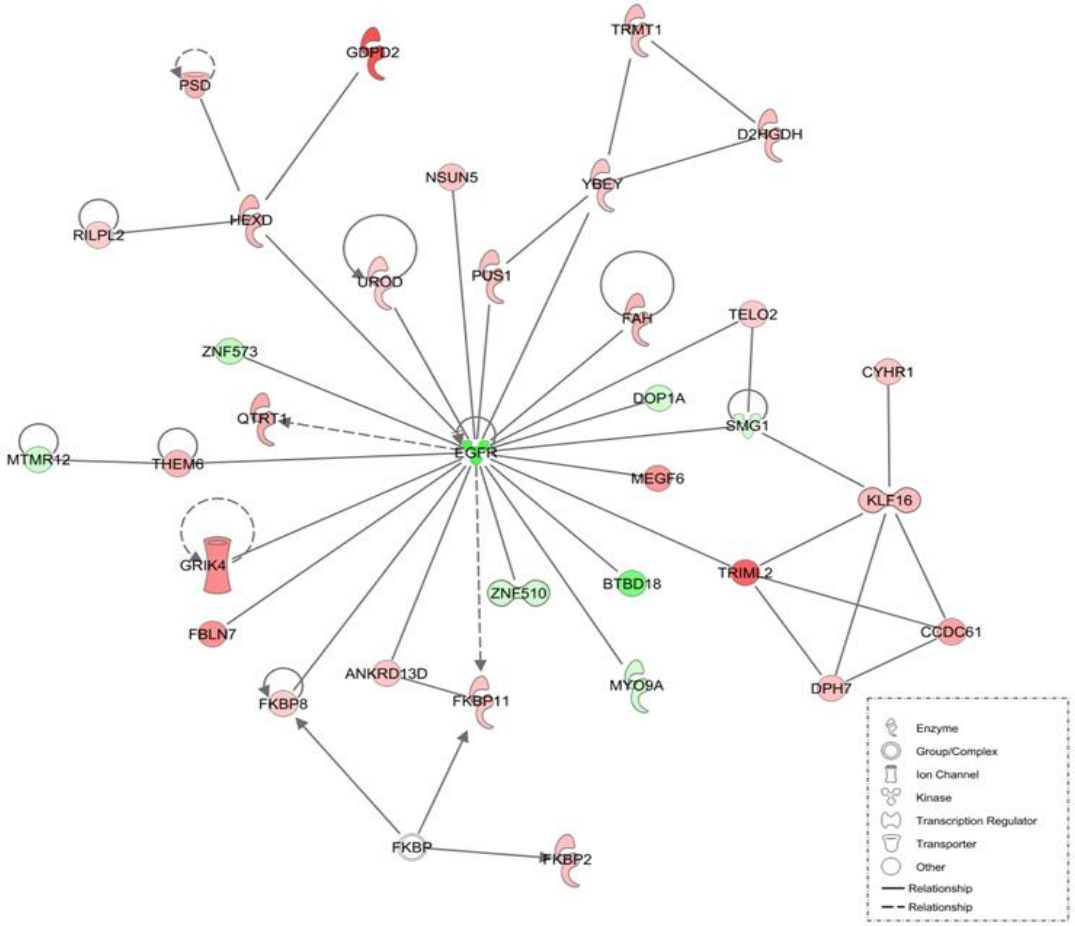
a


LGR6
PUS1
FRZB
Frizzled
RSPO1
NTHL1
SMAD1/5
QTRT1
MEGF6
Wnt
BTBD18
ZNF510
THEM6
WNT6
XRRA1
VIL1
LRSAM1
Hedgehog
ZNF573
NELL1
EGFR
CRCT1
ACY1
PTCH2
NSUN5
ANKRD13D
COQ8A
SPRR1A
CCDC60
Plc beta
FBLN7
HEXD
MAPK15
GDPD2
CLIC3
Enzyme
G-protein Coupled Receptor
Group/Complex
Growth factor
Ion Channel
Kinase
Peptidase
Transcription Regulator
Transmembrane Receptor
Other
Relationship
Relationship

b


CYHR1
ALDH3A1
SURF6
ALDH16A1
ERVW-1
LRRC4
RRP9
SHC2
Par6
LIN28A
ALDH
XPO4
PPAN
NTHL1
EXOC5
Plc beta
PUS1
MIRLET7
CRCT1
GDPD2
EIF4G2
HEXD
EGFR
ANKRD13D
ZNF573
MAPK15
Laminin1
FBLN7
ERBB
QTRT1
WDR81
BTBD18
NRG (family)
HOXB9
CCDC60
Complex
Enzyme
Group/Complex
Kinase
Transcription Regulator
Translation Regulator
Transporter
Other
Relationship
Relationship

c


TRMT1
GDPD2
PSD
D2HGDH
NSUN5
YBEY
HEXD
RILPL2
UROD
PUS1
FAH
TELO2
ZNF573
CYHR1
DOP1A
QTRT1
SMG1
MTMR12
THEM6
EGFR
MEGF6
KLF16
GRIK4
TRIML2
BTBD18
ZNF510
CCDC61
FBLN7
ANKRD13D
MYO9A
DPH7
FKBP8
FKBP11
FKBP
FKBP2
Enzyme
Group/Complex
Ion Channel
Kinase
Transcription Regulator
Transporter
Other
Relationship
Relationship

d

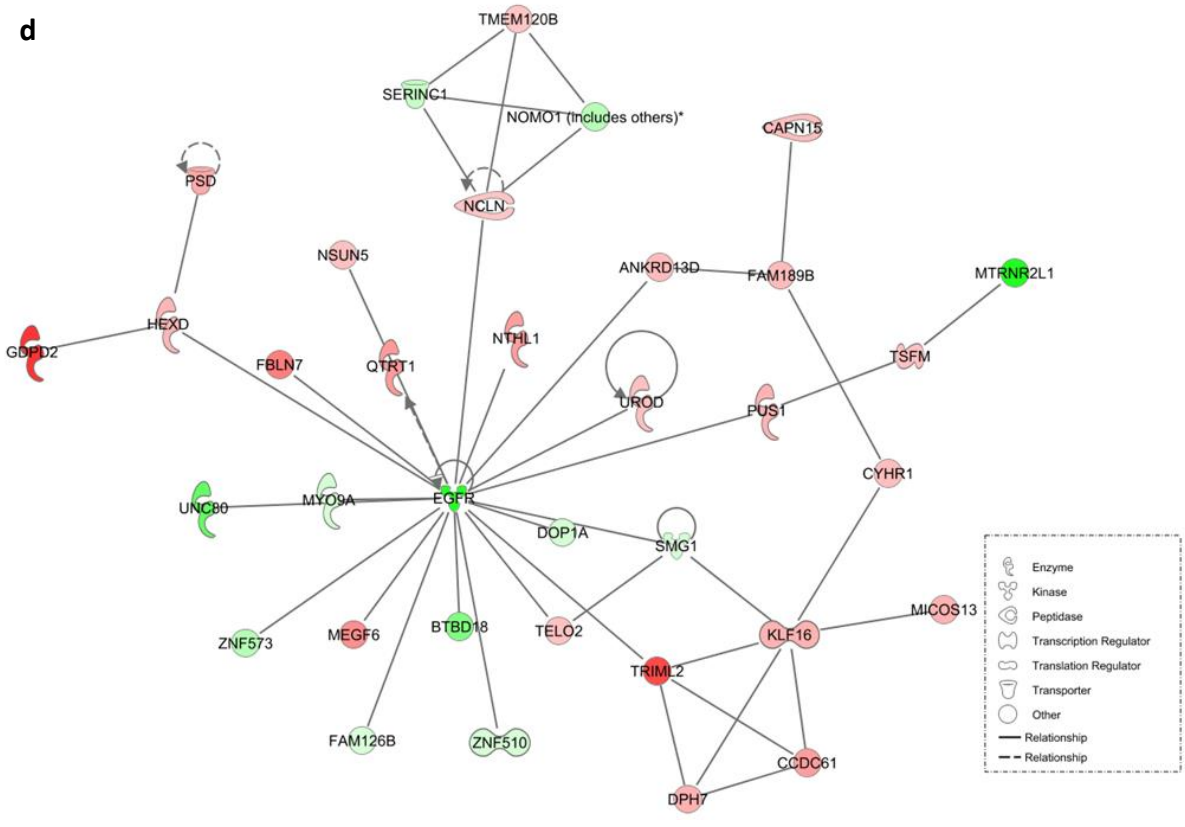
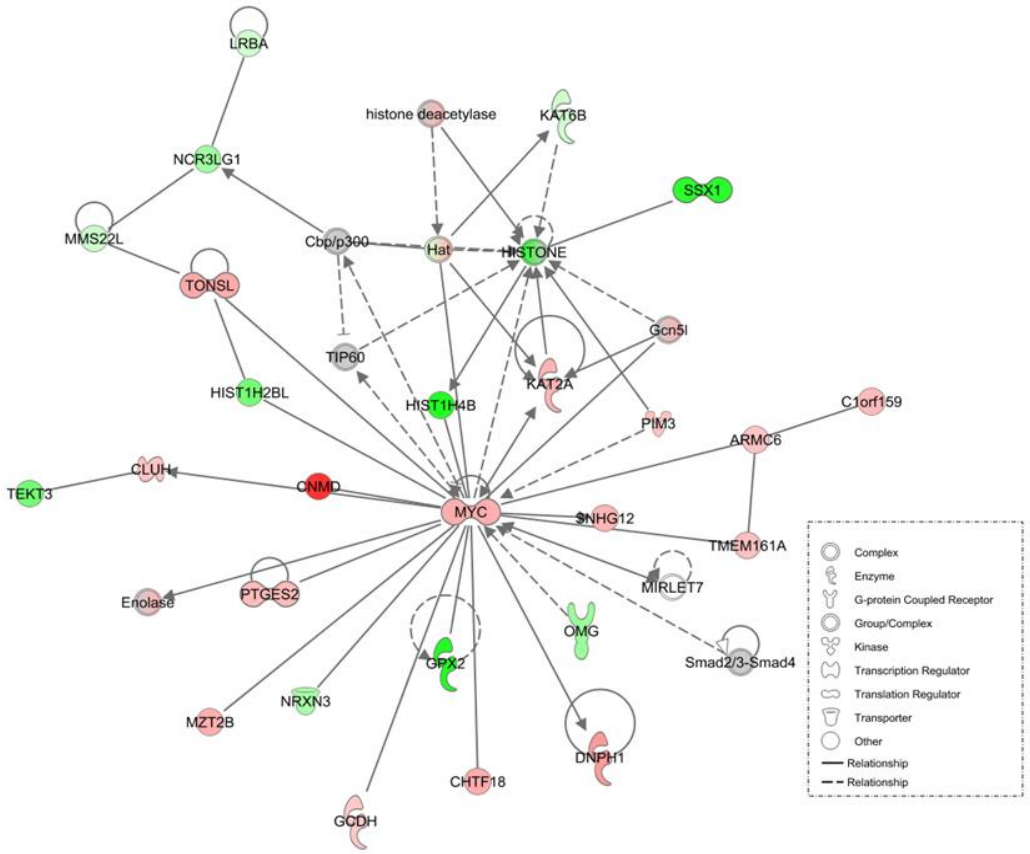

TMEM120B
SERINC1
NOMO1 (Includes others)*
CAPN15
PSD
NCLN
NSUN5
ANKRD13D
FAM189B
MTRNR2L1
HEXD
GDPD2
NTHL1
FBLN7
QTRT1
TSFM
UROD
PUS1
CYHR1
UNC80
MYO9A
EGFR
DOP1A
SMG1
MICOS13
ZNF573
MEGF6
BTBD18
TELO2
KLF16
TRIML2
FAM126B
ZNF510
CCDC61
DPH7
Enzyme
Kinase
Peptidase
Transcription Regulator
Translation Regulator
Transporter
Other
Relationship
Relationship

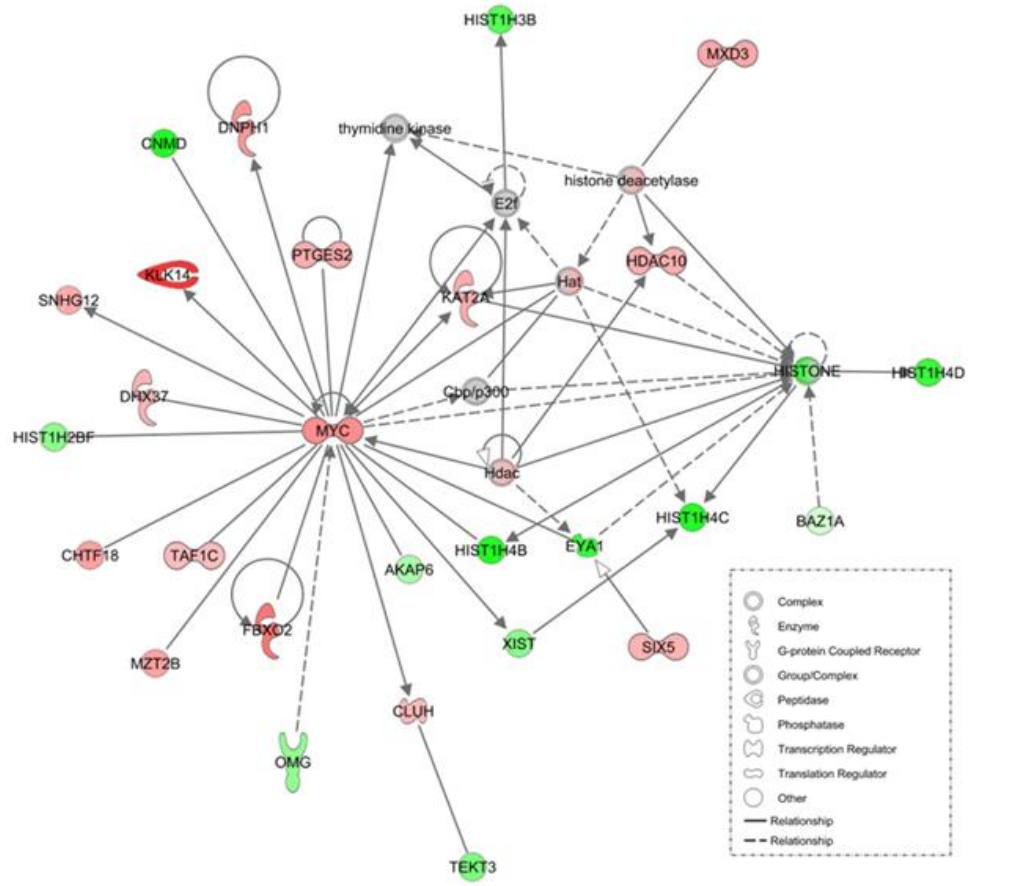
Figure 37 IPA's networks with EGFR as central regulator in TNBC-AA patients. **a** EGFR network in the group with all patients, with a score of 40 and 30 focus molecules, and mainly involved with the following diseases and functions: Cancer, Cellular Movement and Gastrointestinal Disease. **b** EGFR network in the group with postmenopausal patients, with a score of 41 and 29 focus molecules, and mainly involved with the following diseases and functions: Cancer, Cell Cycle and Organismal Injury and Abnormalities. **c** EGFR network in the group with stage II disease patients, with a score of 42 and 34 focus molecules, and mainly involved with the following diseases and functions: Cellular Movement, Developmental Disorder and DNA Replication, Recombination, and Repair. **d** EGFR network in the group with postmenopausal stage II disease patients, with a score of 49 and 35 focus molecules, and mainly involved with the following diseases and functions: Amino Acid Metabolism, Cellular Movement and Connective Tissue Disorders. Genes are represented by nodes whose shape indicates the type of molecule/functional class. The relationship between the nodes is showed by edges. Nodes in red represent statistically significant upregulated genes (FDR ≤ 0.05), nodes in green represent statistically significant downregulated genes (FDR ≤ 0.05), grey nodes stand for non-differentially expressed genes and white nodes correspond to genes not included in the DGEA. Solid lines indicate direct relationship between molecules and dashed lines indicate indirect relationship. Color intensity is proportional to log_2FC value.

Although having lower network scores, Myc also stands out (Figure 38) as a central network regulator in the contrasts with all TNBC patients, postmenopausal patients and postmenopausal stage II disease patients (Supplementary - Networks file, eTables 1, 3 and 7, respectively). Myc regulates up to 15% of human genes, mostly by regulating RNA expression, both protein and non-coding, and is involved in cell growth, proliferation, metabolism, differentiation, stress pathways, mechanisms of drug resistance and apoptosis, thus, sustaining growth of many types of cancers (Camarda et al., 2016; Dang et al., 2006; Fallah et al., 2017). Figure 38a shows the Myc network in TNBC-AA patients from the contrast comprised by all TNBC cases, with a score of 31 and 26 focus molecules included. Here we highlight the fact that, besides Myc itself, this network is composed by genes and proteins evolved in misregulation in cancer, such as histone elements and regulators (HIST1H2BL, HIST1H4B), which are downregulated in this group of patients, while the genes involved in histone deacetylase complex are upregulated. Also, IPA's suggests that this network is involved in the following diseases and functions: Cell Cycle, Cell Death and Survival, and Cellular Assembly and Organization (Supplementary - Networks file, eTable 1). Figure 38b displays Myc as a central regulator in postmenopausal TNBC-AA patients, having a score of 41 and 28 focus molecules. Again, this network is also composed by genes and proteins involved in misregulation in cancer, including the downregulated genes and proteins associated to histone regulation and structure. This statement is also supported by the diseases and functions affected by this network, which include Cellular Assembly and Organization, DNA Replication, Recombination and Repair, and Post-Translational Modification (Supplementary - Networks file, eTable 3). Finally, Figure 38c represents Myc network in postmenopausal stage II disease TNBC-AA patients, with a score of 37 and 30 focus molecules. As in the contrast with all TNBC-AA patients, the elements represented in this network are also suggestively associated with cancer misregulation. In fact, IPA's suggests that this network is involved in the following diseases and functions: Acid Metabolism, RNA Post-Transcriptional Modification, and Small Molecule Biochemistry (Supplementary - Networks file, eTable 7). Overall, these results show that Myc upregulation in TNBC-AA, relatively to TNBC-White patients, and its consequent effect on Myc-dependent genes, is consistent among the groups of patients mostly presenting postmenopausal condition.

a


LRBA
histone deacetylase
KAT6B
NCR3LG1
SSX1
MMS22L
Cbp/p300
Hat
HISTONE
TONSL
Gcn5l
TIP60
KAT2A
HIST1H2BL
HIST1H4B
PIM3
C1orf159
ARMC6
CLUH
TEKT3
CNMD
MYC
SNHG12
TMEM161A
MIRLET7
Enolase
PTGES2
OMG
GPX2
Smad2/3-Smad4
NRXN3
MZT2B
DNPH1
CHTF18
GCDH
Complex
Enzyme
G-protein Coupled Receptor
Group/Complex
Kinase
Transcription Regulator
Translation Regulator
Transporter
Other
Relationship
Relationship

b


HIST1H3B
MXD3
DNPH1
CNMD
thymidine kinase
histone deacetylase
E2f
PTGES2
HDAC10
KLK14
Hat
SNHG12
KAT2A
HISTONE
HIST1H4D
DHX37
Cbp/p300
HIST1H2BF
MYC
Hdac
HIST1H4C
BAZ1A
CHTF18
TAF1C
HIST1H4B
EYA1
AKAP6
FBXO2
XIST
SIX5
MZT2B
CLUH
OMG
TEKT3
Complex
Enzyme
G-protein Coupled Receptor
Group/Complex
Peptidase
Phosphatase
Transcription Regulator
Translation Regulator
Other
Relationship
Relationship

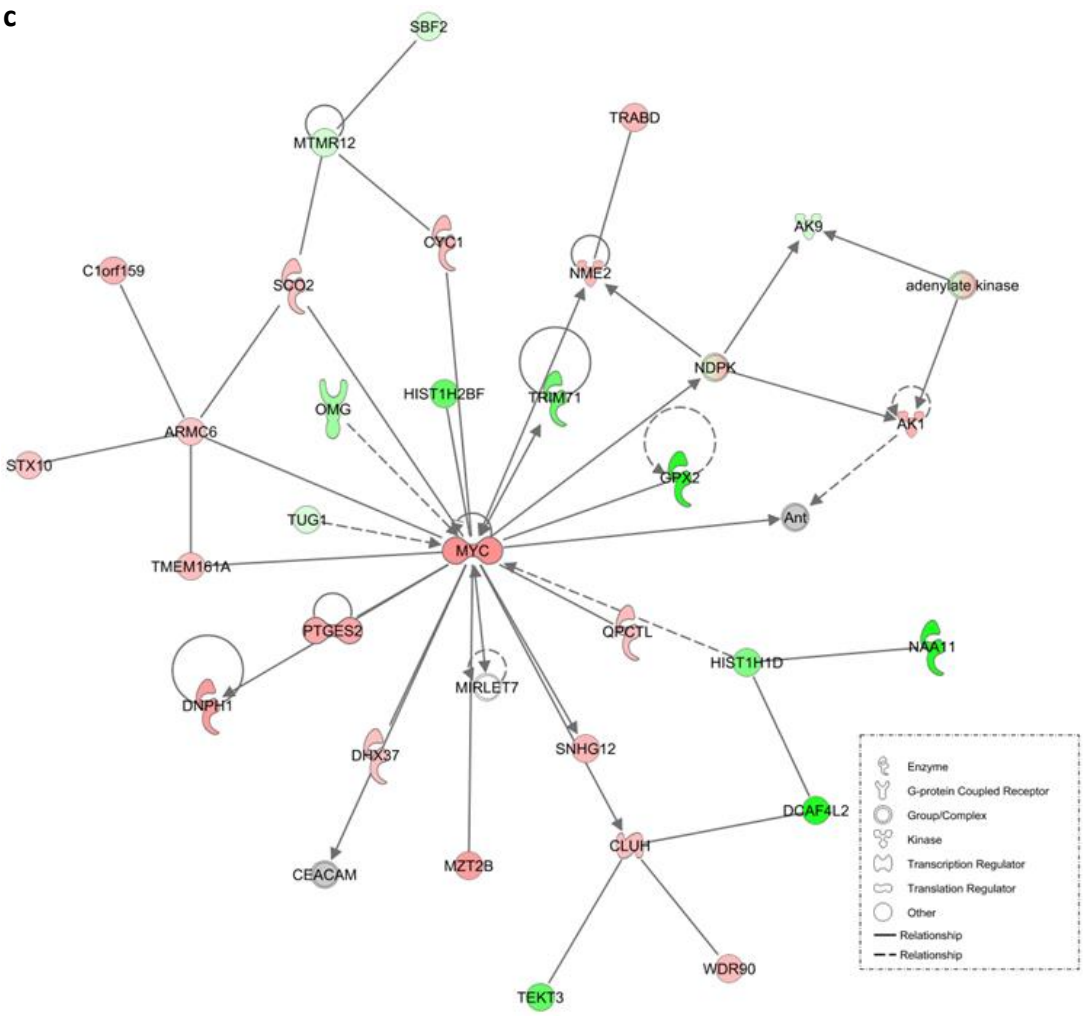


Figure 38 IPA's networks with MYC as central regulator in TNBC-AA patients. **a** MYC network in the group with all TNBC patients, with a score of 31 and 26 focus molecules, and mainly involved with the following diseases and functions: Cell Cycle, Cell Death and Survival and Cellular Assembly and Organization. **b** MYC network in the group with postmenopausal patients, with a score of 41 and 28 focus molecules, and mainly involved with the following diseases and functions: Cellular Assembly and Organization, DNA Replication, Recombination, and Repair and Post-Translational Modification. **c** MYC network in the group with postmenopausal stage II patients, with a score of 37 and 30 focus molecules, and mainly involved with the following diseases and functions: Nucleic Acid Metabolism, RNA Post-Transcriptional Modification and Small Molecule Biochemistry. Genes are represented by nodes whose shape indicates the type of molecule/functional class. The relationship between the nodes is showed by edges. Nodes in red represent statistically significant up-regulated genes (FDR ≤ 0.05), nodes in green represent statistically significant downregulated genes (FDR ≤ 0.05), grey nodes stand for non-differentially expressed genes and white nodes correspond to genes not included in the DGEA. Solid lines indicate direct relationship between molecules and dashed lines indicate indirect relationship. Color intensity is proportional to log_2FC value.

Finally, Bcl2 stands out as being downregulated in premenopausal TNBC-AA comparing with TNBC-White patients (Figure 39). Bcl2 is an apoptosis regulator that usually acts by promoting cell survival rather than by driving cell proliferation (Adams and Cory, 2007). Bcl2 network is the only network having a score above 30 in the premenopausal group of patients, with a score of 47. This is

one of the highest scores obtained through IPA's network design, only second to EGFR network in postmenopausal stage II disease patients, which has a score of 49 (Supplementary - Networks file, eTable 2 and 7, respectively). Moreover, although it was not possible to observe statistically significant enriched gene sets (FDR ≤ 0.05) for this group of patients in the GSEA (Section 3.1.), through IPA's Network design it was possible to identify Bcl2 network as being significantly altered in premenopausal TNBC-AA patients relatively to TNBC-White patients. Nonetheless, although scored with a high value, this network only has 21 key molecules. Interestingly, according to this network, upregulated PIK3CA is indirectly regulating Bcl2 downregulation, and vice-versa. Furthermore, according to IPA, the disease and functions affected by this network include Cancer, Organismal Injury and Abnormalities, and Skeletal and Muscular System Development and Function (Supplementary - Networks file, eTable 2).

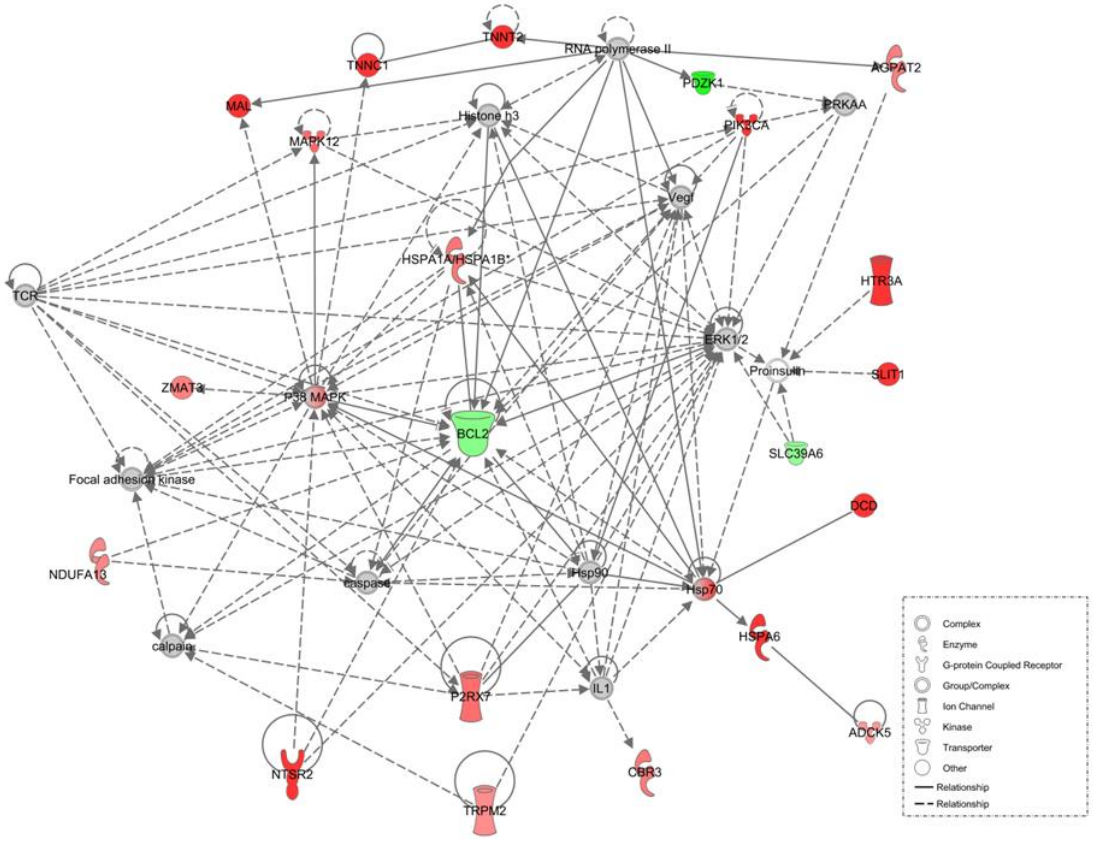


Figure 39 IPA's network with Bcl2 as central regulator in premenopausal TNBC-AA patients. Bcl2 network has a score of 47 and 21 focus molecules, and mainly involved with the following diseases and functions: Cancer, Organismal Injury and Abnormalities and Skeletal and Muscular System Development and Function. Genes are represented by nodes whose shape indicates the type of molecule/functional class. The relationship between the nodes is showed by edges. Nodes in red represent statistically significant up-regulated genes (FDR ≤ 0.05), nodes in green represent statistically significant downregulated genes (FDR ≤ 0.05), grey nodes stand for non-differentially expressed genes and white nodes correspond to genes not included in the DGEA. Solid lines indicate direct relationship between molecules and dashed lines indicate indirect relationship. Color intensity is proportional to $\log_2$FC value.

GSEA and IPA's Network design, together with DGEA and literature review, contributed to the selection of candidate genes to be further explored at the protein level using clinical samples.

However, we must take into consideration that the observed differences between TNBC-AA and TNBC-White patients could be due to differences associated to intrinsic ancestry-associated factors,

and not related to TNBC development. Thus, it is important to scrutinize which differences are due to the disease progression and those which are normally found in breast tissue between these two groups of patients with different ancestry.

4. DIFFERENCIAL EXPRESSION PROFILE BETWEEN TNBC-AA AND TNBC-WHITE PATIENTS ARE DISEASE-SPECIFIC AND NOT ANCESTRY-ASSOCIATED

To ascertain if the differential gene expression profile observed between TNBC-AA and TNBC-White patients are due to differences with features of the disease or just due to normal ancestry-associated gene expression in the mammary tissue, DGEA was performed using RNA-seq data from normal-adjacent BC tissue, removed during a mastectomy or lumpectomy, available in TCGA.

4.1. Differential expression profile between normal-adjacent BC tissue of AA patients vs White patients

DGEA to compare normal-adjacent tissue from patients with different subtypes of BC was performed using the same methodology as with the RNA-seq data from TNBC samples (Chapter III, Section 1.3.). This analysis included RNA-seq data from normal-adjacent BC tissue obtained from 6 AA cases and 105 White cases (Supplementary - Patients file, eTable 2). Figure 40 shows the overall gene expression differences between patients through an MDS plot. In general, there is no clear difference between the expression profile of normal tissue between the two populations, although some White patients have their expression profile further apart from the main cluster. A hypothesis that can justify these apparent outliers may be the fact that, although considered histologically normal, this samples are from tissue surrounding cancer and, therefore, they may have stromal infiltration due to the adjacent cancer, which can affect gene expression.

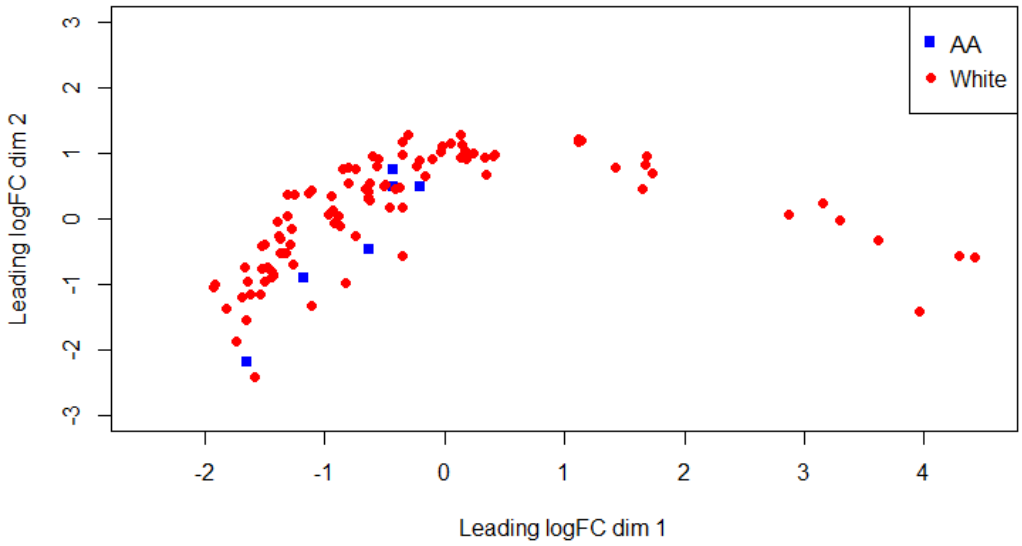


Figure 40 MDS plot of RNA-sequencing data from normal-adjancent breast cancer tissue from TCGA patients, according to patients ancestry. *AA* African-American.

This DGEA resulted in 76 DEG (Supplementary - DGEA file, eTable 13), all of which are upregulated in Normal-AA cases comparing Normal-White cases (Table 22, Figure 27).

Table 22 Number of differentially expressed genes with FDR ≤ 0.05 in normal-adjacent BC tissue from AA patients, comparing with matching tissue from White patients.

↑ upregulated genes. ↓ downregulated genes

AA African-American patients. *W* White patients

FDR ≤ 0.05	76 (76 ↑ + 0 ↓)
# Patients	6 AA vs 105 W

The DEG identified in Normal-AA cases, comparing with Normal-White cases, are highlighted in Figure 40. In this MD plot we can observe that all DEG in Normal-AA tissue are upregulated.

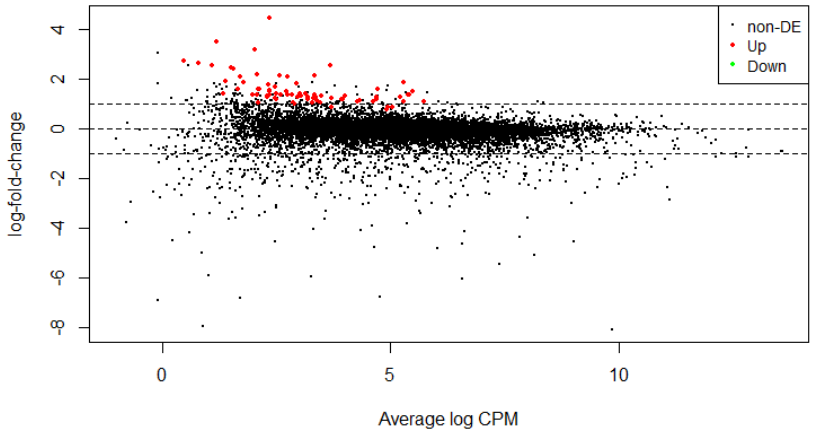
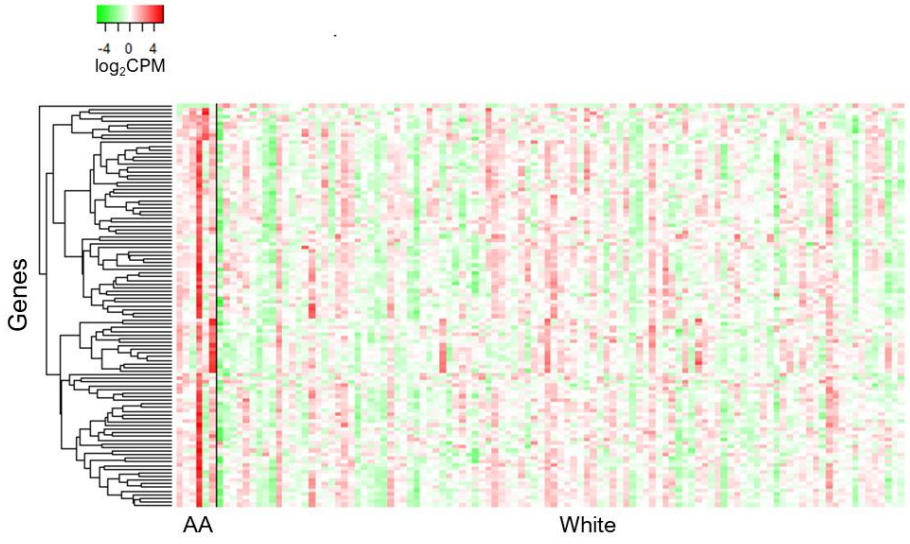


Figure 41 MD plot with log-fold-change against average count size, highlighting the differentially expressed (DE) genes between normal-adjacent breast cancer tissue from AA and matched tissue from White patients. *non-DE* non-differentially expressed genes (black dots), *Up* upregulated genes in TNBC tissue comparing with normal tissue (red dots), *Down* downregulated genes in TNBC tissue comparing with normal tissue (green dots). *CPM* counts per million.

Figure 42 displays the heatmaps showing the gene expression profile resulting from the DGEA with RNA-seq from normal-adjacent breast tissue. Figure 42a heat map shows the gene clustering of all the 76 DEG identified in normal tissue. On the other hand, Figure 42b shows the top 20 most statistically significant DEG.

a



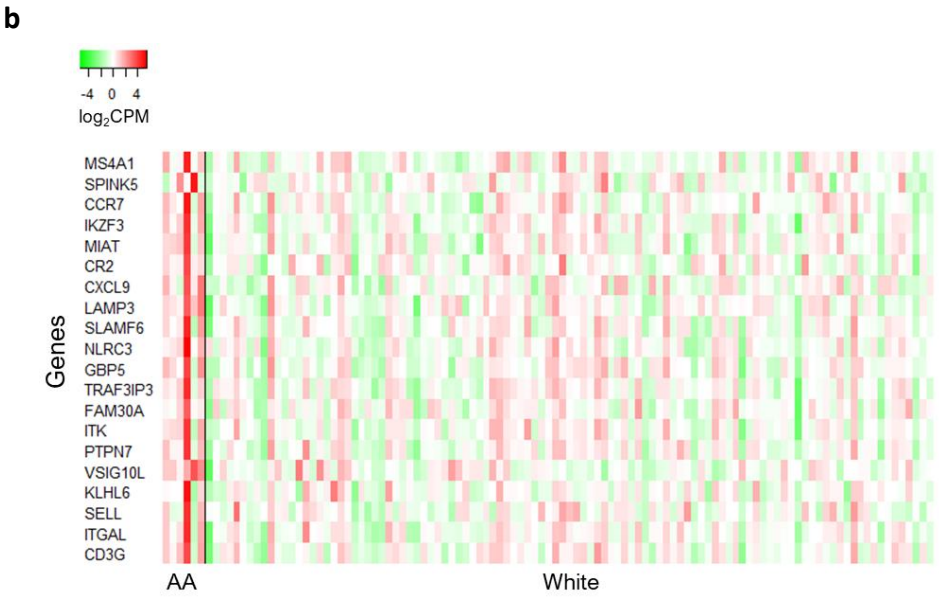


Figure 42 Heat maps of normal-adjacent tissue from breast cancer patients. **a** Clustering of all differentially expressed genes observed in normal-adjacent breast cancer tissue from AA patients. **b** Top 20 of the most differentially expressed genes identified in normal-adjacent breast cancer tissue from AA patients, in ascending order of FDR. AA African-American. *CPM* counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

This DGEA showed that the large majority of the DEG previously identified are associated with TNBC progression in the two patients' populations, as those differences are not visible in normal mammary tissue. In fact, only two genes - *CXCL10* and *LTB* - were found to be simultaneously differentially expressed between AA and White patients in both TNBC and normal tissue (Supplementary - Matching DEG file). Hence, these results reinforce the idea that the expression profile differences observed in TNBC patients seems to be mostly due to differences in the disease biology, and not to inherent features related to patients' ancestry.

4.2. Differential expression profile between TNBC-AA patients and Normal-AA cases

Normal and TNBC tissue from AA patients were also compared in order to identify genes that lost or acquired an abnormal expression during TNBC progression in AA patients.

DGEA between 52 TNBC-AA patients and 6 Normal-AA cases resulted in 4137 DEG (Table 23). This high number of DEG is a strong indicator that there is indeed a molecular alteration of patients' mammary tissue throughout the TNBC development.

Table 23 Number of differentially expressed genes in TNBC-AA patients comparing with Normal-AA cases (with FDR ≤ 0.05).

↑ upregulated genes. ↓ downregulated genes

AA African-American patients. *NB* Normal-adjacent breast cancer tissue.

FDR ≤ 0.05	4137 (2592 ↑ + 1545↓)
# Patients	52 TNBC vs 6 NB

Accordingly, the following MDS plot (Figure 43) shows that there is, indeed, a clear distance regarding the expression profile of TNBC-AA patients and Normal-AA cases, as suggested by the cluster formation.

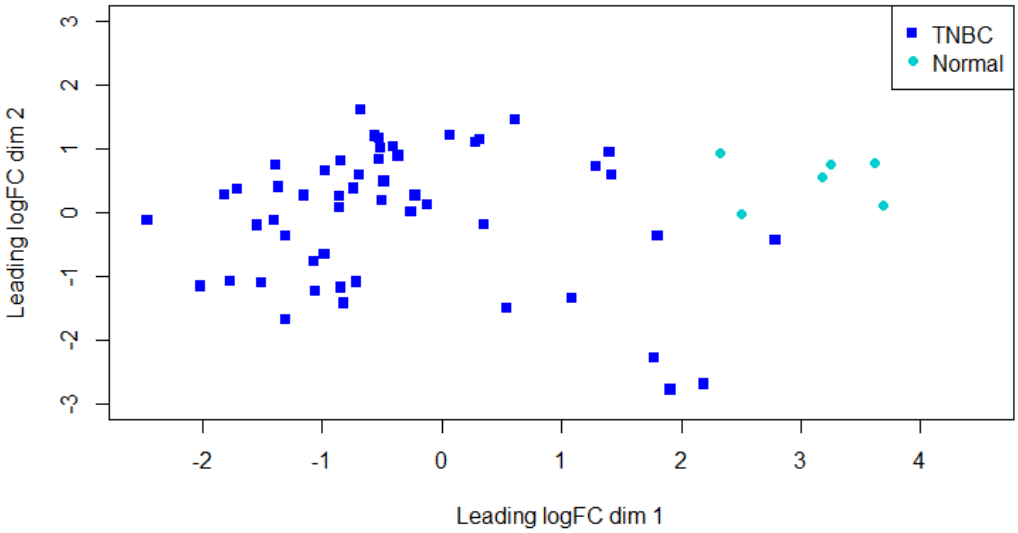


Figure 43 MDS plot of RNA-sequencing data from TNBC tissue vs normal-adjacent breast cancer tissue. *AA* African-American. *Normal* Normal-adjacent breast cancer tissue from AA patients.

Figure 30 is a representation of the 4137 DEG in the form of an MD plot, where most DEG, each represented by a red or green dot, are indistinguishable from each other, given their large number. Complete list of DGE identified in the contrast with TNBC-AA patients and Normal-AA cases can be consulted in Supplementary – DGEA file, eTable 14.

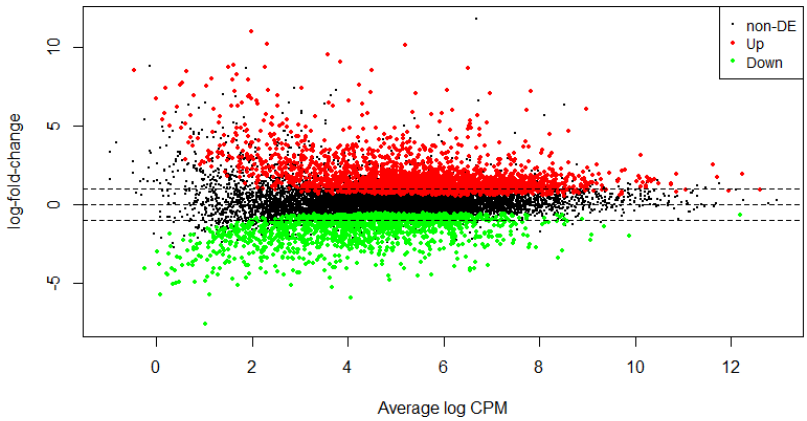
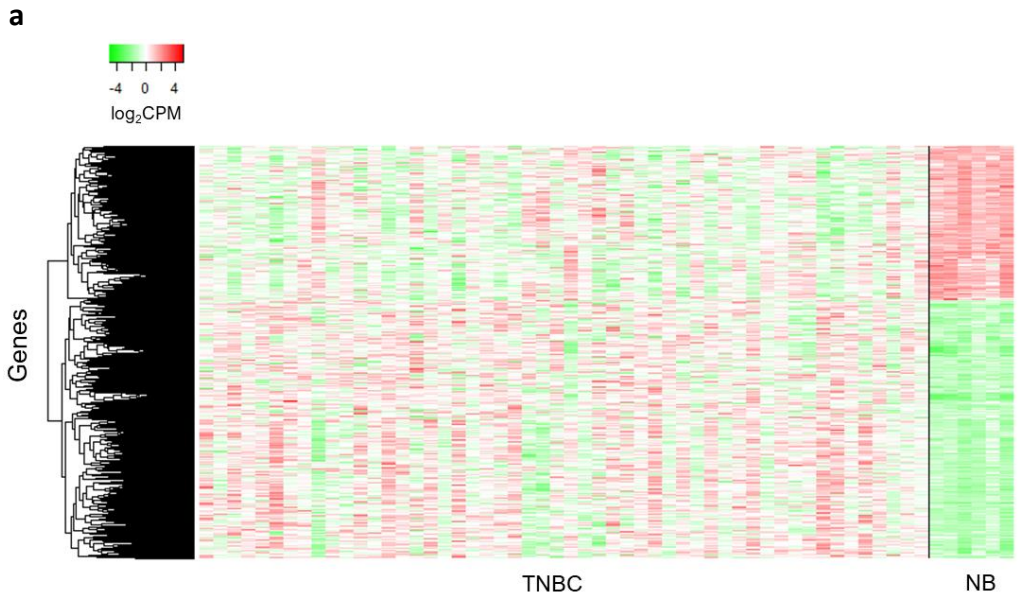


Figure 44 MD plot with log-fold-change against average count size, highlighting the differentially expressed (DE) genes between TNBC tissue and normal-adjacent breast cancer tissue from AA patients. *non-DE* non-differentially expressed genes (black dots), *Up* upregulated genes in TNBC tissue comparing with normal tissue (red dots), *Down* downregulated genes in TNBC tissue comparing with normal tissue (green dots). *CPM* counts per million.

Figure 45 shows a heat maps with the DEG identified in TNBC-AA patients when comparing with Normal-AA cases. Figure 45a represents the gene clustering of all 4137 DEG identified in TNBC-AA patients, with a clear cluster formation between the normal and TNBC tissue. Figure 45b shows the top 20 most statistically significant DEG, where none of these Top 20 most statistically significant DEG match with any of the DGEA with TNBC patients.



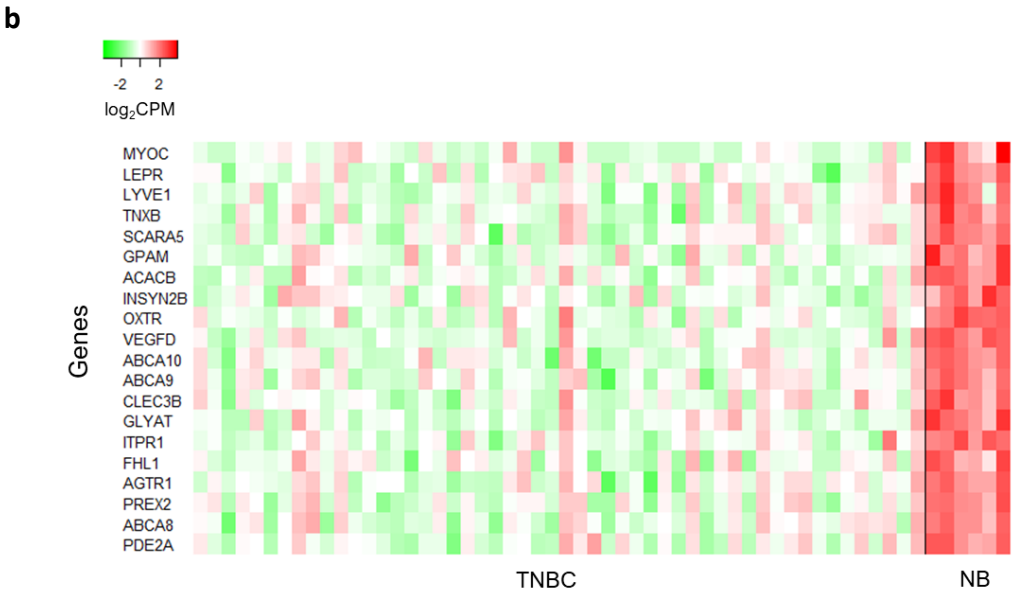


Figure 45 Heat maps of TNBC and normal-adjacent breast cancer tissue from AA patients. a. Clustering of all differentially expressed genes observed in TNBC tissue from AA patients. b Top 20 of the most differentially expressed genes identified in TNBC tissue from AA patients, in ascending order of FDR. AA African-American. CPM counts per million. Green - downregulated gene. Red - Upregulated gene. Color intensity is proportional to log_2CPM value.

Some of the DEG identified in this DGEA match with the DEG identified in the contrasts comparing TNBC-AA patients with menopause and/or stage matched TNBC-White patients (Supplementary - Matching DEG file). Thus, such set of DEG is not only involved in TNBC development as they may be exclusively to TNBC development in women of African-ancestry.

5. EGFR, MYC, BCL2 AND B-CATENIN AS POTENTIAL PROGNOSIS MARKERS AND TERAPEUTIC TARGETS IN AFRICAN-ANCESTRY PATIENTS DIAGNOSED WITH TNBC

Overall, based on the reported analysis, *EGFR*, *MYC* and *BCL2* were the DEG selected for further validation, at the protein expression level, using formalin-fixed paraffin-embedded (FFPE) TNBC samples.

Furthermore, we additionally investigated β-catenin protein expression in the clinical samples. β-catenin is a key downstream component of the canonical Wnt signaling pathway as well as a critical link in a chain of proteins that strengthen cell-cell adhesion by coupling cadherins to the actin cytoskeleton (Shang et al., 2017). In the breast, in addition to creating and stabilizing the mammary gland architecture, the catenin-cadherin complexes are involved in cell fate, polarity, motility and survival (Hatsell et al., 2003). Also, the role of Wnt/β-catenin signaling in BC and TNBC tumorigenesis (Dey et al., 2013b, 2013a; Xu et al., 2015) and worse prognosis (Dey et al., 2013b; Geyer et al., 2011; Khramtsov et al., 2010; López-Knowles et al., 2010) is well described. For these reasons, we decided to ascertain β-catenin expression, which is encoded by *CTNNB1* gene. Also, this protein is directly or indirectly involved in some of our findings. For instance, according to GSEA, we observed that Wnt signaling pathway is one of the most negatively regulated pathways identified in TNBC-AA patients (Figure 32). Also, in the group of patients presenting postmenopausal status and stage II disease, β-

catenin, which is downregulated (log_2FC = -0.935, FDR = 1.63E-02), is involved in the Hippo signaling pathway (Figure 33), signaling pathways regulating pluripotency of stem cells (Figure 34b) and in the breast cancer gene set (Figure 35). Moreover, in addition to Wnt signaling pathway, β-catenin is an important intermediate in many other signal transduction pathways, including the PI3K-AKT pathway (Ozawa et al., 1989). Also, Myc is among the many β-catenin downstream effectors (He et al., 1998; Xu et al., 2015). Since Wnt/β-catenin signaling is usually upregulated in BC, as well as in the triple-negative subtype, most treatments consist in Wnt/β-catenin inhibitors, such as PRI-724, a small molecule Wnt co-activator antagonist, or Sorafenib and Refametinib, a tyrosine kinase inhibitor and an MEK inhibitor, respectively, which inactivate β-catenin signaling (Krishnamurthy and Kurzrock, 2018; O'Toole et al., 2013). However, if African-ancestry patients with TNBC have indeed a lower expression of Wnt/β-catenin, these treatments would be particularly harmful, since the Wnt/β-catenin signaling is essential in the lymphopoiesis and hematopoiesis and regeneration of tissues and organs (Staal and Sen, 2008).

All in all, it is important to highlight the fact that other ancestry-associated DEG, either observed in the contrast with all TNBC patients or in the contrasts considering patients' menopause status, pathological stage of the disease, or both, can also be relevant as ancestry-associated markers in African-ancestry TNBC patients, with impact in the prognosis and/or as therapeutics targets.

Thus, in this project, EGFR, Myc, Bcl2 and β-catenin were further validated as potential ancestry-associated markers in African-ancestry patients diagnosed with TNBC.

6. DIFFERENCES IN EGFR, MYC, BCL2 AND β-CATENIN PROTEIN EXPRESSION LEVEL BETWEEN AFRICAN-ANCESTRY TNBC AND WHITE-TNBC PATIENTS, FOLLOWED IN PORTUGAL, ARE ALIGNED WITH DIFFERENCIAL GENE EXPRESSION ANALYSES

The validation cohort included 12 African-ancestry patients and 11 White patients diagnosed with TNBC at Hospital Professor Doutor Fernando Fonseca (HFF) between January 2012 and August 2018. All White patients and one African-ancestry patient are Portuguese, while all the other patients are from the former Portuguese colonies, namely Cabo Verde, São Tomé, Mozambique, Guinea-Bissau and Brazil. The only criterion when choosing the patients was regarding the type of samples, with preference for surgical pieces over biopsies since the last do not have normal-adjacent tissue that would be further used to validate IHC results from malignant tissue in normal-adjacent mammary tissue. Description of clinical characteristics of the validation cohort patients is in Table 24. More information about the validation cohort can be consulted in Supplementary - Patients file, eTable 3.

Differently TNBC patients from TCGA, where no significant differences were observed between AA and White TNBC patients regarding their age of initial diagnosis (55.87 vs 54.23 years old, p = 0.780, Table 13), African-ancestry patients HFF were diagnosed with TNBC at a significantly younger age, comparing with their White counterparts, which are also diagnosed later when comparing with White patients from TCGA (47.75 vs 64.82 years old, p = 0.007, Table 24). Related to this, there is also a significantly higher proportion of premenopausal TNBC cases in the African-ancestry population, regarding the White population (58.33% vs 0, p = 0.010). As with TCGA patients, none of the other clinical variables had statistically significant differences between the two populations of patients from HFF. Kaplan-Meier survival curves were not considered for the validation cohort, due to the small sampling.

Table 24 Description of the validation cohort.

	African-ancestry (n = 12)	White (n = 11)	p-value
Age at diagnosis of **TNBC**, mean (IQR)	47.75 (37 - 54.25)	64.82 (58 - 70)	0.007[a]
Stage at **TNBC** diagnosis, n (%)			
I	1 (8.33)	4 (36.36)	0.262
II	4 (33.33)	2 (18.18)	0.725
III	2 (16.67)	2 (18.18)	> 0.999
IV	4 (33.33)	2 (18.18)	0.725
Indeterminate or not evaluated	2 (16.67)	1 (9.1)	
Menopause status at **TNBC** diagnosis, n (%)			NA
Pre (<6 months since LMP, no prior bilateral ovariectomy and not on estrogen replacement)	7 (58.33)	0	0.010
Post (prior bilateral ovariectomy or >12 months since LMP with no prior hysterectomy)	3 (25)	10 (90.9)	0.006
Indeterminate or not evaluated	2 (16.67)	1 (9.1)	NA
Follow-up **TNBC** patients, mean (IQR)[b]	43 (21 - 58)	57 (45 - 73)	0.131
Vital Status of **TNBC** patients, n (%)			> 0.999
Alive	10 (83.33)	10 (90.9)	NA
Dead	2 (16.67)	1 (9.1)	NA
Time to death of **TNBC** patients, months	22 - 33	53	0.667

[a] Wilcoxon rank-sum test. Otherwise, the comparisons are by Fisher exact test.
[b] Months of follow-up after initial diagnosis.
NA – Not available
IQR – Interquartile range
LMP – Last menstrual period

FFPE TNBC tissue from 12 African-ancestry and 11 White patients collected from HFF (Supplementary - Patients file, eTable 3) was analyzed through immunohistochemistry (IHC) assays to ascertain if EGFR, Myc, Bcl2 and β-catenin protein expression differences among the two groups of patients are in agreement with the differential gene expression observed in TNBC-AA *vs* TNBC- White patients from TCGA. Due to the small sampling, patients were not stratified according to their menopause status or pathological stage of the disease.

Wilcoxon Rank-Sum (Mann-Whitney U) test was used to compare H-score of each protein between African-ancestry and White patients. Briefly, H-score (Ishibashi et al., 2003; John et al., 2009) is a semiquantitative approach in which the intensity of a protein expression is measured using the scale 0-3, 0 being negative and 3 being a very high expression. The H-Score is then calculated by multiplying the intensity value with its area. Remarkably, with no exception, all proteins are significantly differentially expressed between the groups of patients ($p \leq 0.05$) and the net direction of protein expression intensity matches with those of differential gene expression observed in the DGEA (Figure 46). Therefore, these results corroborate that while Myc is upregulated in African-ancestry

patients, EGFR, Bcl2 and β-Catenin are downregulated in these patients, in comparison with White patients.

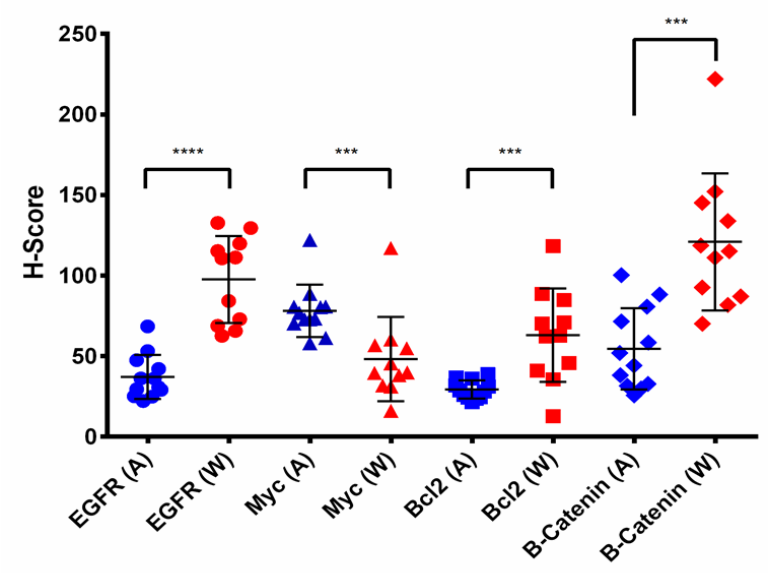


Figure 46 Differences in H-score between African-ancestry (n = 12) and White (n = 11) patients are significantly different for EGFR (p = 5.92E-06), Myc (p = 4.01E-04), Bcl2 (p = 9.81E-04) and β-Catenin (p = 1.44E-04) expression and have the same net direction as the results obtained in DGE analysis. A/Blue - African-ancestry patients. W/Red - White patients. Circles - EGFR expression. Triangles - Myc expression. Squares - Bcl2 expression. Diamonds - β-catenin expression.

Figure 47 shows examples of a positive and a low/negative EGFR staining in African-ancestry and White patients, where we can see that the protein is clearly more expressed in the TNBC samples of White patients. Of note is the fact that different areas of each tissue may show different intensities of EGFR expression and that the examples shown in Figure 47 for positive EGFR staining in African-ancestry patients and negative EGFR staining in White patients were rarely observed. Interestingly, although EGFR downregulation was observed in the contrasts that included postmenopausal patients, stage II disease patients, or both, in clinical samples EGFR protein is under-expressed in a group of African-ancestry patients that is mostly comprised by premenopausal cases.

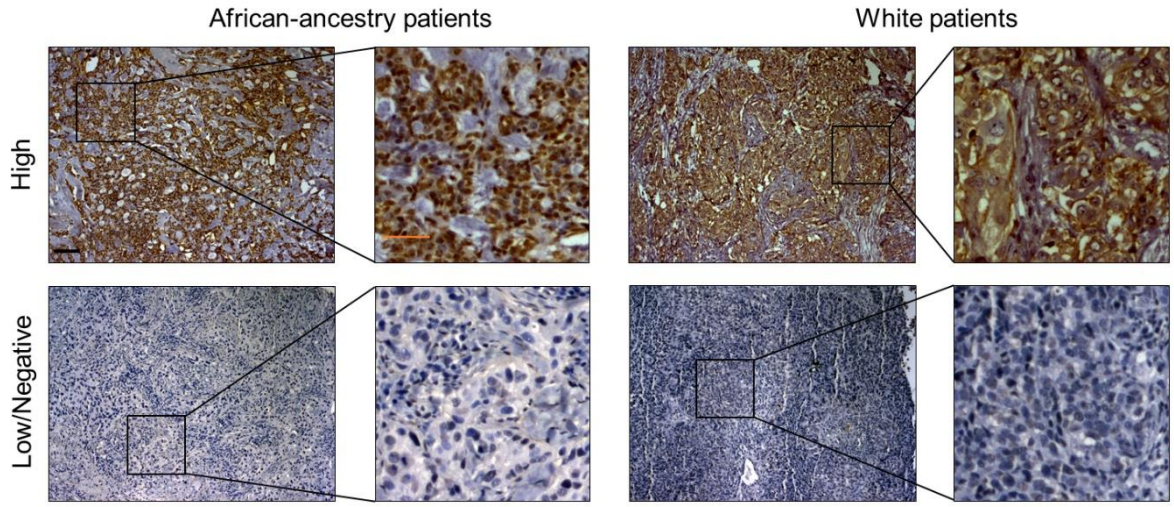


Figure 47 Immunohistochemistry staining of FFPE TNBC tissue from African-ancestry patients (left) and White patients (right) with anti-EGFR IgG. Brown color indicates positive reactivity (top) and shows expression of EGFR. Black scale bar - 100 µm. Orange scale bar - 50 µm.

Examples of Myc positive and low/negative staining are showed in Figure 48, where it can be seen a higher expression of Myc in African-ancestry patients comparing with its expression in White patients. Furthermore, as EGFR, here Myc is overexpressed in a group of African-ancestry patients that are mostly premenopausal.

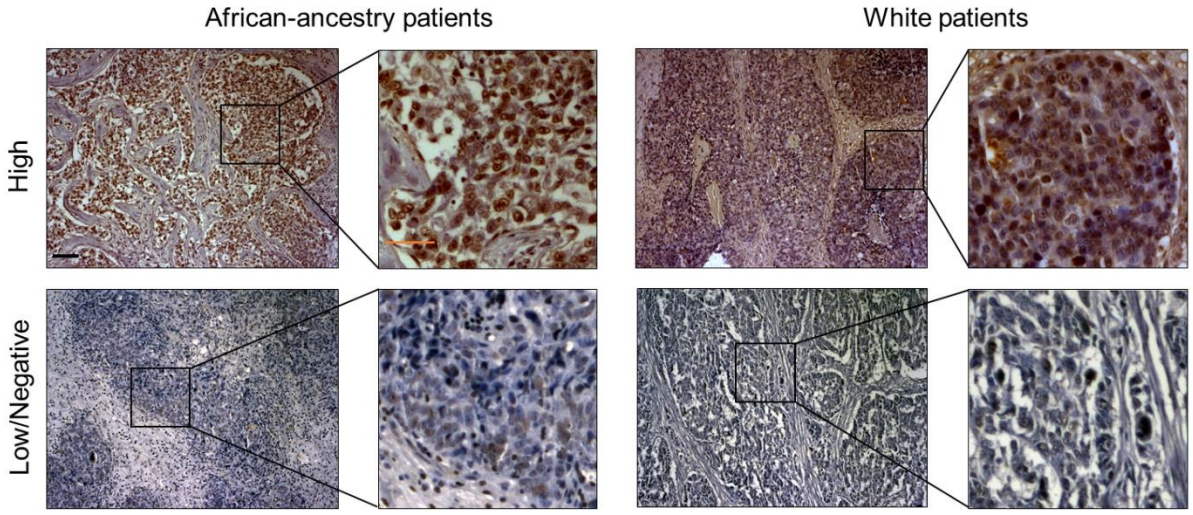


Figure 48 Immunohistochemistry staining of FFPE TNBC tissue from African-ancestry patients (left) and White patients (right) with anti-Myc IgG. Brown color indicates positive reactivity (top) and shows expression of Myc. Black scale bar - 100 µm. Orange scale bar - 50 µm.

Bcl2 IHC results are showed in Figure 49. Interestingly, none of the African-ancestry patients, mostly premenopausal, expressed Bcl2. On contrary, Bcl2 expression can be observed in most White patients.

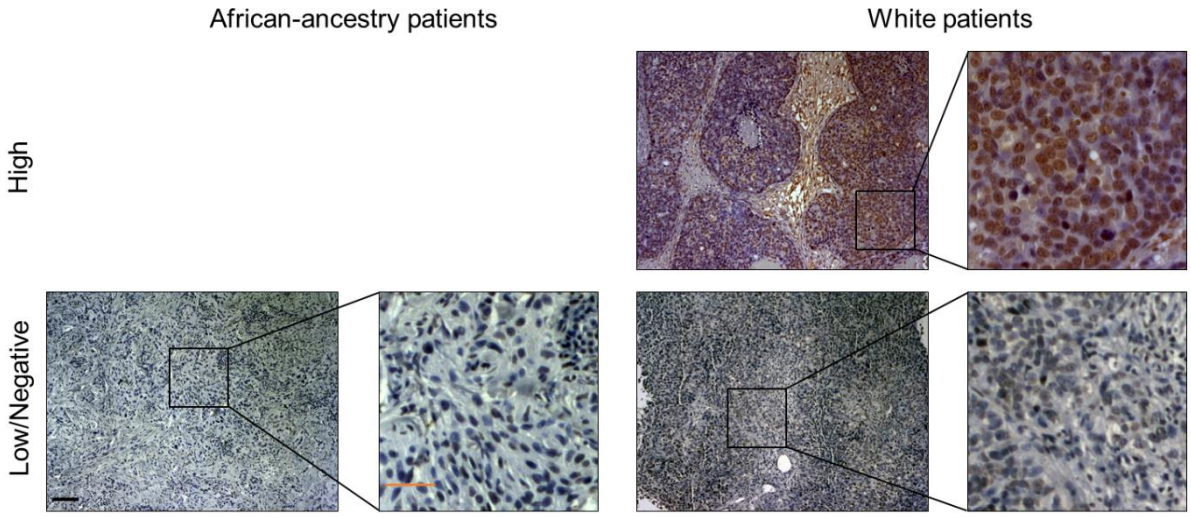


Figure 49 Immunohistochemistry staining of FFPE TNBC tissue from African-ancestry patients (left) and White patients (right) with anti-Bcl2 IgG. Brown color indicates positive reactivity (top) and shows expression of Bcl2. Black scale bar - 100 µm. Orange scale bar - 50 µm.

Finally, both African-ancestry and White patients showed high levels of β-Catenin expression (Figure 50). However, while some African-ancestry patients have low β-Catenin expression, as shown by the H-score of each patient, displayed in Figure 46, β-Catenin is highly expressed in all White patients

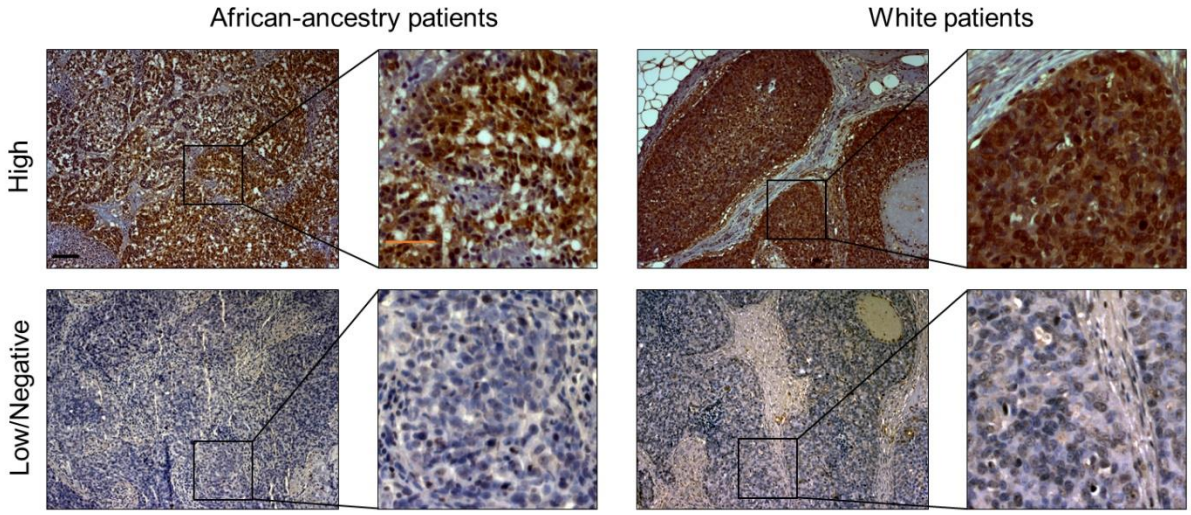


Figure 50 Immunohistochemistry staining of FFPE TNBC tissue from African-ancestry patients (left) and White patients (right) with anti-β-Catenin IgG. Brown color indicates positive reactivity (top) and shows expression of β-Catenin. Black scale bar - 100 µm. Orange scale bar - 50 µm.

To validate these results, we also performed IHC in the normal-adjacent BC tissue from the same patients, excluding those in which only biopsy samples were available. However, the samples of normal-adjacent BC tissue had very few epithelial cells, being mostly composed by fat tissue, more prone to be damaged during FFPE slicing. Although this proportion of fat vs epithelial tissue is expected in a normal breast, it was not enough for a representative staining and quantification. For that reason,

those samples were excluded, and no validation of the candidate molecules was performed in FFPE normal-adjacent breast cancer tissue.

To sum up, we have shown that there is indeed a differential expression profile between AA and White patients diagnosed with TNBC, which vary according to patients' menopause status and pathological stage of the disease. Such ancestry-associated differences at the expression level translate into differences in the activation or repression of signaling pathways and cellular mechanisms, as observed through GSEA. Furthermore, IPA's core analysis, in which we focused on the network design, suggests that some proteins act as central regulators of specific networks found altered in groups of TNBC-AA patients presenting certain menopause status and/or stages of the disease. Finally, among the chosen candidate genes for validation through IHC, using samples collected in Portugal, protein expression of EGFR, Myc, Bcl2 and β-catenin are remarkable in line with the FC direction of the genes from which are encoded, even though study and validation cohorts are composed by geographically and culturally distinct groups of patients. In the next chapter, we will further discuss these findings in the current state of the art regarding ancestry-associated discrepancies and TNBC development and prognosis.

Chapter V

Discussion

1. OVERVIEW

Triple-negative (TNBC) is the breast cancer (BC) subtype that neither expresses estrogen receptor (ER), nor progesterone receptor (PR), nor has Human Epidermal growth factor Receptor 2 (HER2) amplification. Consequently, TNBCs are defined by what they "lack" rather than what they "have" and thus this negative nomenclature provides no actionable information on druggable targets. Pertinently, only in 2020 the first targeted therapy was approved for TNBC, consisting in the combination of the checkpoint inhibitor atezolizumab with nab-paclitaxel (Kang and Syed, 2020; Schmid et al., 2020). However, this immunochemotherapy option is specifically for patients with unresectable locally advanced or metastatic TNBC, in which the tumors have a PD-L1 expression ≥ 1% (Kang and Syed, 2020). Thus, most patients with TNBC are still heavily dependent on conventional chemotherapy (CT), and the survival rate is dismal (Garrido-Castro et al., 2019; Saraiva et al., 2017).

Globally, TNBC accounts for 10-20% of all BC diagnoses. However, it has been recognized that women of African-ancestry are twice as likely to develop TNBC than women of European/Caucasian descent (P. Boyle, 2012; DeSantis et al., 2019), hereafter denominated as White patients, as stipulated in TCGA race list.

TNBC in African-ancestry women present higher mortality rates compared to women of European/Caucasian ancestry. Although factors for such disparity may include advanced disease stage at diagnosis, socioeconomic status, and lack of access to the healthcare treatment, biological disparities are one of the main factors involved in the ancestry-associated TNBC discrepancy (Grunda et al., 2012; Keenan et al., 2015; Siddharth and Sharma, 2018). While White and Asian women tend to have TNBC with a later age of onset and less aggressive clinical course (Kwong et al., 2009; Lim et al., 2009; Saraiva et al., 2017), African-ancestry women are often premenopausal (Carey et al., 2006; Huo et al., 2009; Lund et al., 2009; Zaky et al., 2009), with a faster and more aggressive development of TNBC, independent of other risk factors for BC (Stead et al., 2009). Additionally, whenever possible, TNBC patients are treated with longstanding and costly conventional systemic treatment of classic chemotherapy, which in many cases do not have any effect due to chemo-resistance or, in cases where response to chemotherapy is positive, African-ancestry patients relapse faster when compared with their White counterparts (Copson et al., 2014; Frasci et al., 2009). Overall, ancestry-specific molecular mechanisms involved in the increased aggressiveness and earlier age of onset of TNBC in African-ancestry patients are mostly unknown.

Therefore, identification of ancestry-associated markers should contribute to understand more clearly the features of this disease in both populations. Additionally, this study may also contribute to patients' prognosis and establishment of more effective targeted therapies, potentially closing the widening mortality gap between these populations

2. CHARATERIZATION OF TNBC PATIENTS

In this project, two groups of patients were used. Namely, for the computational analysis, the group was comprised by African-American (AA) and White TNBC patients with RNA-sequencing (RNA-seq) data from The Cancer Genome Atlas (TCGA); and for the validation of the selected molecules given by the computational analysis, clinical samples from African-ancestry and White TNBC patients diagnosed in the Portuguese Hospital Prof. Dr. Fernando Fonseca (HFF), were used.

Other authors have previously explored TCGA data in order to ascertain ancestry-associated discrepancies in BC, and TNBC in particular, at molecular level (Ademuyiwa et al., 2017; Huo et al.,

2017; Keenan et al., 2015; Koboldt et al., 2012; O'Meara et al., 2019; Polak et al., 2017; Stewart et al., 2013). However, this study is one of the few to systematically examine ancestry-associated differences at transcriptomic level with RNA-seq data in TNBC patients, with the particularity of considering patients' menopause status and TNBC pathological stage. Since African-ancestry women are frequently diagnosed with TNBC at pre-menopause and/or with more advanced pathological stages of the disease, and due to the fact that the molecular mechanisms that regulate this TNBC incidence are far from being understood, we proposed to identify potential ancestry-specific markers involved in altered pathways and cellular mechanisms in African-ancestry patients, associated to a menopause status and a pathological stage of the disease.

Comparing our patients' epidemiological characteristics with those from other studies, the incidence of TNBC cases among TCGA's AA BC patients is significantly higher than among White BC patients (28.42 % vs 11.89 %, respectively, $p < 0.0001$), according to data collected on March 6, 2019, following our method of selecting TNBC patients (Chapter III, Section 1.1.2.). Previous TCGA studies reported statistically significant results with the same tendency, namely Stewart et al. (Stewart et al., 2013) in 2013 (18.87 % TNBC-AA vs 11.67 % TNBC-White patients), Keenan et al. (Keenan et al., 2015) in 2015 (36.3 % TNBC-AA vs 13.7 % TNBC-White patients), Ademuyiwa et al. (Ademuyiwa et al., 2017) in 2017 (33.3 % TNBC-AA vs 14.9 % TNBC-White patients), Huo et al. (Huo et al., 2017) in 2017 (36.4 % AA basal-like BC patients vs 14.7 White basal-like BC patients) and O'Meara et al. (O'Meara et al., 2019) in January 2019 (35.8 % TNBC-AA patients vs 16.4% % TNBC-White patients). TNBC incidence of African-ancestry and White patients among BC cases from HFF was not estimated due to the small sampling of TNBC and BC patients with information regarding patients' self-described ancestry, particularly in the case of White patients.

We also compared other epidemiological and clinical variables between the African-Ancestry and White TNBC patients from TCGA and from HFF. Regarding the age of TNBC diagnosis between AA and White patients, there were no significant differences in the groups of patients from TCGA (Chapter IV, Section 1). On contrary, and being in line with earlier reports from other studies (Carey et al., 2006; Huo et al., 2009; Lund et al., 2009; Zaky et al., 2009), African-ancestry patients from HFF were diagnosed with TNBC at a significantly younger age than White patients (mean age 47.75 years vs 64.82 years, respectively, $p = 0.007$; Chapter IV, Section 6).

In line with these results, there were no significant differences between TNBC-AA and TNBC-White patients from TCGA in respect to the menopause status. On the other hand, premenopausal African-ancestry patients from HFF comprise 58.33% of all cases in that population, while, interestingly, none of the White patients is premenopausal (58.33% vs 0, $p = 0.010$).

No significant differences were observed regarding the incidence rate of the different stages of the disease at TNBC diagnosis between both populations from TCGA and HFF. Regarding the patients from TCGA, these results were not surprising since previous studies based on TCGA data never reported statistical significant differences regarding the number of patients of both ancestries diagnosed in a specific stage of the disease (Ademuyiwa et al., 2017; Keenan et al., 2015; O'Meara et al., 2019). Nonetheless, other studies reported a higher incidence of advanced stages (III/IV) in TNBC cases from African-ancestry patients, namely in non-Hispanic AA patients (Amirikia et al., 2011) and in patients from Ghana (Stark et al., 2010).

Concerning TCGA patients' survival up to 12 years of follow-up after initial diagnosis, although not being statistically significant (log-rank $p = 0.085$), TNBC-AA patients have a lower survival probability tendency than TNBC-White patients (Chapter IV - Figure 15a), as previous observed in other TCGA-based studies (Ademuyiwa et al., 2017; Huo et al., 2017), and other epidemiological studies

exploring the ancestry-associated discrepancy of TNBC (Copson et al., 2014; DeSantis et al., 2019; Dietze et al., 2015a; Lund et al., 2009; Sullivan et al., 2014). After stratifying TCGA patients according to menopause status and pathological stage of the disease, premenopausal and stage II disease TNBC-AA patients have significantly lower survival probability comparing with matching TNBC-White patients (Chapter IV - Figure 15b and 15e, log-rank p = 0.019 and 0.0038, respectively). A similar tendency, although not statistically significant, is observed when comparing postmenopausal patients (Chapter IV - Figure 15c, log-rank p = 0.13). These results are in line with the mortality rate observed in past studies, in which African-ancestry TNBC patients have a higher mortality rate when comparing with White patients (Dietze et al., 2015a; Lund et al., 2009; Sullivan et al., 2014), particularly in younger patients (Albain et al., 2009; DeSantis et al., 2019). Also, to our best knowledge, no studies were previously performed regarding mortality rate or overall survival (OS) of these two populations according to the stage of the disease at initial diagnosis. Hence, this is the first study reporting a significantly lower survival probability in African-ancestry patients specifically associated to pathological stage II at TNBC diagnosis, when comparing with matching White patients. Two reasons for this ancestry-associated TNBC discrepancy in the survival of stage II disease patients might be related to the fact that TNBCs in African-ancestry patients are reportedly enriched in populations of poorly differentiated cells [28–30], and due to increased genomic instability throughout the stage of the disease [17]. Thus, we speculate that although this subgroup of TNBC-AA patients was initially diagnosed with stage II TNBC, a number of patients might have resisted systemic CT and the disease relapsed and/or evolved faster than in the subset of stage II TNBC-White patients, leading to a lower survival probability in TNBC-AA patients. Finally, the low mortality among patients diagnosed at stage I, and the low survival probability in the groups of patients from both ancestries diagnosed at stage III were not surprising, since these are early and advanced stages of the disease, respectively. Survival probability was not possible to ascertain in both groups of TNBC patients from HFF due to their small numbers. Nonetheless, Honório et al. (Honório et al., 2016) observed that African-ancestry patients were more frequently presented with late stage disease and worse survival outcome than the non-African patients from HFF, regardless of BC subtype (median survival: 62 vs 15 months, $p < 0.005$).

3. DIFFERENTIAL GENE EXPRESSION PROFILE IN TNBC-AA PATIENTS

Diseases such as cancer are very complex, being controlled by many genes. Furthermore, most oncogenes and tumor suppressor genes are signaling molecules themselves, each of which functions to control the expression of a subset of downstream genes (Sager, 1997; Vogelstein, Bert; Lane, David; J. Levine, 2000). Thus, differential gene expression analysis (DGEA) is one of the most widely used strategies for discovering and understanding the molecular circuitry underlying cancer (Liang and Pardee, 2003), as well as finding new potential targets (Xue et al., 2020).

RNA-seq uses the capabilities of high-throughput sequencing methods to provide insight into the transcriptome of a sample. Compared to previous Sanger sequencing- and microarray-based methods, RNA-seq provides far higher coverage and greater resolution of the dynamic nature of the transcriptome (Wang et al., 2009). Thus, RNA-seq studies possess great potential to provide novel insights into the pathogenesis of complex diseases such as cancer. This study systematically applied bioinformatic methods to identify new ancestry-associated markers with roles in signaling pathways and cellular mechanisms involved in the development and progression of TNBC in African-ancestry women. DGEA was performed between groups of TNBC-AA and TNBC-White patients controlling for

the same conditions. Such contrasts controlled for either the same menopause status, the same pathological stage of the disease, both conditions, or neither of them, accounting for all TNBC cases in that case (for more information, consult Chapter III, Section 1.3.1.1.1.). By performing DGEA of each contrast, our aim was to identify differentially expressed genes (DEG) in TNBC-AA patients associated to one, or more, of the menopause and staging variables, since others (Chang et al., 2018; DeSantis et al., 2019; Huo et al., 2017; Keenan et al., 2015; Lindner et al., 2013; Polak et al., 2017; Stewart et al., 2013; Sugita et al., 2016) and our results concerning patients' survival probability (Chapter IV, Figure 15) suggest that biological differences are involved in a more aggressive TNBC disease in African-ancestry patients. DEG identified in TNBC-AA patients from TCGA, according to patient's menopause and/or staging of the disease, are potentially involved in the characteristics that make TNBC more aggressive in African-ancestry patients than in patients with other ancestries.

Before performing DGEA, MDS plots were designed to show the gene expression profile distance between patients, in two dimensions. MDS plots confirmed that there is a distinct gene expression profile between TNBC-AA patients and AA patients diagnosed with Hormone receptor (HR) positive and HER2+ BCs (Chapter IV, Figure 17a). Additionally, transcriptomes of AA patients diagnosed with HR+ and HER2+ are more similar to each other than to transcriptomes of patients with TNBC, corroborating the idea that TNBC is a BC entity with more distinct characteristics than other less serious, more homogeneous and well characterized BC subtypes, as previously suggested (Huo et al., 2017; Stewart et al., 2013) (Chapter IV, Figure 17b).

Finally, ancestry-associated discrepancies in disease severity and previous reported differences (Chang et al., 2018; Huo et al., 2017; Keenan et al., 2015; Lindner et al., 2013; Stewart et al., 2013; Sugita et al., 2016), motivated further analyses taking into consideration the data substructure. Thus, the first part of this project consisted in the identification of DEG between TNBC-AA patients and matching White patients, evaluating the contribution of patients' menopause status and pathological stage of the disease.

3.1. Power and sample size

Grouping patients in order to identify DEG associated specifically with a menopause status, a pathological stage of the disease, or both, caused a decreased number of patients to be included in the contrasts. However, although a larger sample size would be more suitable in the contrasts with less patients included, we are confident in our results, as explained below.

In the specific case of this study, the power of the differential expression test is the probability of finding statistically significant DEG between AA and White patients as a function of the dimension of the true differences between these two populations. Thus, the more distant the expression profile between the two groups of patients, the higher the power. In other words, if the true difference is high, then the test can have high power even with low sample sizes (Jones et al., 2003; Li et al., 2019). We also implemented a strong stringency in our analysis, including filtering out genes that do not reach 10 CPM in at least two libraries in each group in a contrast (Chapter III – Section 1.3.1.1.4.), implementing the TREAT function when testing for differentially expressed genes (Chapter III – Section 1.3.1.1.9.), and considering a gene as differentially expressed when FDR ≤ 0.05. For these reasons, even when the groups in a contrast have a small number of patients, we are confident that our results are robust.

In fact, *edgeR* package allows for differential expression tests even with only one biological replicate [45]. *edgeR* authors acknowledge that RNA-seq is not only an expensive technology as sometimes happens that only one or two libraries are available for a specific condition [45].

3.2. Ancestry-associated disparity is more pronounced according to disease's progression than to menopause status

Differential gene expression analysis of the contrasts including patients diagnosed with stage II disease (32 TNBC-AA vs 59 TNBC-White patients) and postmenopausal stage II disease patients (22 TNBC-AA vs 36 TNBC-White) resulted in a higher number of DEG (1776 and 1360 DEG, respectively) than the contrast with all TNBC patients. Interesting, the contrast controlling for post-menopause status (32 TNBC-AA vs 57 TNBC-White), although having 22 AA and 37 White patients in common with the contrast controlling for stage II disease, resulted in a considerably lower number of DE genes (718 DEG). Thus, although TNBC has a discrepant age of onset when comparing African-ancestry and White patients, as reported in other studies (Carey et al., 2006; Huo et al., 2009; Zaky et al., 2009) and in the group of TNBC patients from HFF (Chapter IV, Section 6), our results suggest that differences in ancestry-associated transcriptomic programs in TNBC patients are more pronounced when associated with the stage progression of the disease than with patients' hormonal differences associated with their menopause status. In the case of stage II disease TNBC, and as previously discussed in this Chapter (Section 2), these results also support the hypothesis that stage II may be the turning point in the uneven aggressive behavior of TNBC in African-ancestry patients.

3.3. Expression profile of early stage TNBC-AA patients suggests preponderance for chemo- resistance and higher fatty acid metabolism

Discrepancies in TNBC development and prognosis in African-ancestry patients, when comparing with patients from other ancestries, although subtle, if present in the early stage of the disease, can allow for more pronounced differences throughout TNBC's progression.

Only the contrast with premenopausal stage I disease patients (2 TNBC-AA vs 7 TNBC-White patients) did not resulted in statistically significant DEG (Chapter IV, Table 18). This may be due to the fact that, in the early development of TNBC, transcriptional program differences among African-ancestry and White patients are subtle, particularly in younger patients. On the other hand, DGEA of post-menopause stage I disease patients (3 TNBC-AA vs 11 TNBC-White patients) resulted in 10 upregulated genes in TNBC-AA patients (Chapter IV, Table 18). Interestingly, most of these overexpressed DEG, which are involved in detoxifying mechanisms and fatty acid metabolism (Chapter IV, Table 19), are common among the DGEA with stage I patients, independently of menopause status (Supplementary - Matching DEG file). Overexpressed genes include *GSTM5*, *GSTM2*, *GUSB*, *GSTT2B* or *FMO5* (Chapter IV – Table 19), which are normally involved in cellular detoxification of toxic compounds, such as products of oxidative stress, environmental pollutants and carcinogens, through neutralization of their electrophilic sites and production of stable and soluble compounds, which are easily excreted from the organism. Nonetheless, upregulation of these enzymes may also promote detoxification of cytotoxic chemotherapeutic agents and, consequently, may promote chemo-resistance (Coles and Kadlubar, 2003; Matejcic et al., 2011; Thodberg and Neilson, 2020). Furthermore,

overexpression of genes such as *ACSL3*, *IDI1*, *ELOVL5* or *AACS* (Chapter IV, Table 19) are involved in fatty acid metabolism and lipid biosynthesis. During tumorigenesis, cancer cells rely on de novo biosynthesis and fatty acid uptake to not only sustain their rapid proliferative rate, but also provide an essential energy source during conditions of metabolic stress (Koundouros and Poulogiannis, 2020). However, the role of each of these genes in carcinogenesis still needs further clarification. Interestingly, and in line with what was observed in stage I disease TNBC-AA patients, *ACSL3* is reportedly downregulated in MDA-MB-231 TNBC cell line (Wright et al., 2017), which belongs to a White women, and in a cohort of Korean TNBC patients (Jeong et al., 2017).

Interestingly, it should be noted that, of all these upregulated genes involved in cellular detoxification or lipid metabolism, only *GSTM5* and *GSTM2* belong to a chromosomic cluster. Hence, although these genes share similar functions, their joint upregulation is not due to an amplification of the same chromosomic region. Altogether, these results suggest that early stage TNBC-AA patients show increased expression of genes involved in drug resistance and lipid metabolism, which may contribute to cancer progression.

However, this hypothesis requires some caution, since there are only 3 TNBC-AA patients in this contrast (Chapter III, Table 8), and one patient in particular is highly expressing all of these 10 DEG (Chapter IV, Figure 28).

3.4. PIK3CA is overexpressed in premenopausal and stage III disease TNBC-AA patients

Among the contrasts with premenopausal and/or stage III disease patients, *PIK3CA* is the most statistically significant upregulated gene in TNBC-AA patients (FDR = 3.87E-08 in the pre-menopause contrast, 9 TNBC-AA vs 29 TNBC-White patients; FDR = 3.13E-05 in the stage III disease contrast, 10 TNBC-AA vs 12 TNBC-White patients; and FDR = 4.36E-08 in the pre-menopause stage III disease contrast, 2 TNBC-AA vs 4 TNBC-White patients; Supplementary - DGEA file, eTables 2, 6 and 9; Supplementary - Matching DEG file). In fact, only the contrast with postmenopausal stage III disease patients (7 TNBC-AA vs 6 TNBC-White patients), does not have PIK3CA as a DEG. In turn, this contrast has *KCND2* and *PLAU* (Chapter IV - Table 21, FDR = 1.75E-02 and 2.34E-02, respectively), which, to our knowledge, are uncorrelated to each other, as the only statistically significant downregulated genes.

PIK3CA and PI3K network has been widely studied in TNBC (Basho et al., 2017; Costa et al., 2018; De et al., 2014; Massihnia et al., 2016; Shah et al., 2012). PI3K-Akt-mTOR (PAM) uncontrolled activation represents a profound disturbance in the regulation of cell growth and survival, which ultimately leads to a competitive growth advantage, metastatic competence, angiogenesis, and therapy resistance (Porta et al., 2014). Thus, this complex pathway has been taken into consideration as one of the most attractive targets for the development of anticancer agents (Costa et al., 2018; De et al., 2014; Hennessy et al., 2005; Saini et al., 2013). In TNBC, overexpression or activating mutations of PI3K catalytic subunit α, *PIK3CA*, can lead to oncogenic activation of the PAM pathway (Cossu-Rocca et al., 2015). Remarkably, TCGA's BC-AA patients have fewer somatic mutations in *PIK3CA* when compared with BC White patients (20.0 % vs 33.9 %, p = 0.003 in Keenan et al. (Keenan et al., 2015), and 23 % vs 34 %, p = 0.021 in Ademuyiwa et al. (Ademuyiwa et al., 2017)). Nonetheless, this ancestry-associated discrepancy in *PIK3CA* mutational prevalence was not observed in TNBC patients, and authors did not investigated *PIK3CA* differential expression (Ademuyiwa et al., 2017; Keenan et al., 2015).

Although interesting, the observed *PIK3CA* overexpression in premenopausal and/or stage III disease TNBC-AA patients, must be interpreted with caution, as this result might reflect a particular overexpression of this gene in two TNBC-AA patients (Chapter IV, Figures 21b and 26b, respectively). Interestingly, the other overexpressed gene found in these two premenopausal stage III disease TNBC-AA patients is *ZMAT3* (Chapter IV - Table 20), a gene that is located in the same chromosomic region as *PIK3CA*, at 3q26.32. These observations suggest that these two premenopausal stage III disease patients must have such a large duplication in the chromosomic region covering *PIK3CA* and *ZMAT.*

To our best knowledge, *PIK3CA* overexpression was never associated with a younger age of onset nor with patients of African-ancestry, although PAM pathway activation has been reportedly associated with TNBC development and chemo-resistance (Costa et al., 2018; De et al., 2014; Hennessy et al., 2005; Porta et al., 2014; Saini et al., 2013). It would be interesting to validate *PIK3CA* overexpression in a larger cohort of young African-ancestry patients with advanced TNBC, even more considering that there is no ancestry disparity in the prevalence of somatic mutations in this gene in TNBC (Ademuyiwa et al., 2017; Keenan et al., 2015), which are known to promote activation of PAM pathway.

4. ALTERED PATHWAYS, CELLULAR MECHANISMS AND NETWORKS

Through DGEA results, gene set enrichment analysis (GSEA) and Ingenuity Pathway Analysis (IPA) network design feature, were performed to predict signaling pathways, cellular mechanisms and genes regulating altered networks involved in the observed ancestry-associated discrepancy of TNBC.

Although studies reported that although African-ancestry patients have a higher incidence of BC and TNBC at younger age than matching White patients (Carey et al., 2006; Huo et al., 2009; Lund et al., 2009; Zaky et al., 2009), patients from our study cohort are more frequently postmenopausal, as observed in other studies (Lund et al., 2009; Stead et al., 2009; Stewart et al., 2013), and diagnosed at the pathological stage II of the disease (Ademuyiwa et al., 2017; Amirikia et al., 2011; Huo et al., 2017; Keenan et al., 2015; Lund et al., 2009; Stewart et al., 2013). Thus, although GSEA and IPA's core analysis were performed using the complete list of genes obtained through DGEA from all the contrasts (Supplementary - DGEA file), here, most of the findings I highlight are regarding to postmenopausal and/or stage II AA patients since the DGEA results showed that in these conditions AA and White TNBC patients have more pronounced differences. Hence, further analysis of these contrasts has the greatest potential to result in important discoveries.

4.1. Hippo signaling pathway is decreased in TNBC-AA patients

Hippo signaling pathway was found to be the most enriched gene set, being downregulated in the contrasts with all TNBC patients, postmenopausal, stage II and postmenopausal stage II patients (Chapter IV, Figures 32 and 33, and Supplementary – GSEA file eTables 1, 3, 5 and 11, respectively).

The Hippo signaling pathway is not only considered an important player in organ growth and size maintenance as dysregulation of this pathway also contributes to cancer development (Johnson and Halder, 2014; Zhao et al., 2010, 2008). Simply put, the main function of the Hippo pathway is to negatively regulate the activity of Yes-associated protein (YAP) and transcriptional co-activator with PDZ-binding motif (TAZ), two homologous transcriptional co-activators that are the main downstream

mediators of the Hippo pathway (Hong and Guan, 2012). The Hippo core module, composed by the kinases large tumor suppressor homolog 1 (*LATS1)* and *LATS2*, and by the kinases mammalian STE20 like protein kinase 1 (*MST1*) and *MST2*, as well as by the adaptor proteins Salvador homolog 1 (*SAV1)*, MOB kinase activator 1A (*MOB1A)*, *MOB1B*, carries out an inhibitory phosphorylation of *TAZ* and *YAP* (Johnson and Halder, 2014). Thus, the Hippo pathway is considered to be in the active state when the MST and LATS kinases are active (Johnson and Halder, 2014), as showed in Figure 51a.

Contributing to the negative net direction of the Hippo signaling pathway gene set, *LATS1*, a tumor suppressor gene (St John et al., 1999), is downregulated in postmenopausal and/or stage II TNBC-AA patients (Chapter IV, Figure 33), comparing with their White counterparts. *LATS1* downregulation has been associated with more aggressive forms of cancer (Britschgi et al., 2017; Jiang et al., 2006; Steinmann et al., 2009; Takahashi et al., 2005), including basal-like BC and TNBC (Furth et al., 2018). Also, *LATS1* depletion in BC cell lines increases cancer cell plasticity and expression of basal-like features and, in BC patients, is associated with increased resistance to hormone therapy with tamoxifen (Furth et al., 2018).

Decreased expression of the Hippo pathway components, such as *LATS1*, as well as other pathway-extrinsic cues that may, activate YAP and TAZ, promoting the transcription of target genes involved in several biological processes, such as cell proliferation and competition, cell death inhibition, epithelial-to-mesenchymal transition (EMT), tumor metastasis, and tumorigenesis (Harvey et al., 2013; Lei et al., 2008; Marti et al., 2015; Piccolo et al., 2014; Zanconato et al., 2016; Zhao et al., 2010) (Figure 51b). The most recent studies also demonstrate that YAP/TAZ is required to sustain the self-renewal and tumor-initiation capacities of cancer stem cells (CSCs) (Hayashi et al., 2015; Lian et al., 2010) (Figure 51b).

Although *YAP* and *TAZ* are not differentially expressed in TNBC-AA patients, comparing with matching White patients (Supplementary - DGEA file, eTables 1, 3, 5 and 11), a decreased expression of the Hippo signaling pathway may promote an increased activation of YAP/TAZ. In TNBC and basal-like BC, high YAP/TAZ activity correlates with bioactivity of CSCs, high-grade histology, and metastasis (Bartucci et al., 2015; Cordenonsi et al., 2011; Díaz-Martín et al., 2015; Kim et al., 2015; Liu et al., 2018; Zhao et al., 2010). Thus, as expected, patients' prognosis is impaired with attenuation of Hippo signaling pathway and, accordingly, by a higher activity of YAP/TAZ. In a retrospective study where TAZ and YAP expression was assessed in TNBC samples, the combined expression of YAP in TNBC cells and in the surrounding stroma seems to be associated with a decreased likelihood of patients achieving pathological complete response (pCR). Conversely, the combined expression of TAZ and YAP in tumor cells conferred poor survival outcomes (Vici et al., 2016).

Therapeutic intervention for TNBC would involve reducing or inhibiting the oncogenic function of YAP and/or TAZ. However, to date, few small molecule inhibitors have been discovered that target the Hippo pathway, and the prevalent view is that most Hippo pathway signaling components are not conventional drug targets. Possibly, inhibiting YAP and TAZ function may require targeting protein-protein interactions (Li et al., 2018), since not only YAP and TAZ are transcriptional co-activators with no known catalytic activity as, to our knowledge, no known upstream regulators that specifically promote YAP and TAZ activity have demonstrated enzymatic activity (Harvey et al., 2013; Johnson and Halder, 2014; Yu and Guan, 2013). A further obstacle is that YAP and TAZ are essential for tissue repair and regeneration in some contexts (Brandão et al., 2019; Grusche et al., 2011; Karpowicz et al., 2010; Sun and Irvine, 2011; Xin et al., 2013), raising questions as to whether systemic and chronic manipulation of Hippo signaling might have potential deleterious side effects on normal tissue function and homeostasis. Thus, the development of small-molecule modulators of Hippo signaling,

pr **a** Activated Hippo pathway tly regulate **b** Inactivated Hippo pathway provide exciting new approaches for non-regenerative medicine ar JBC patients

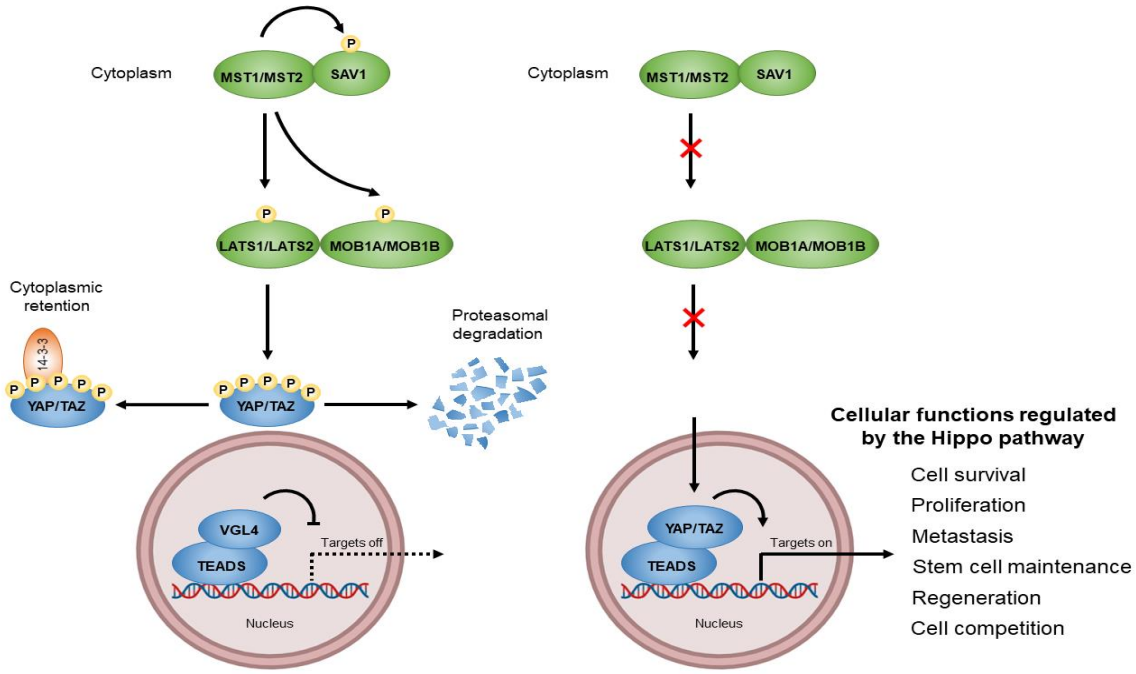


Figure 51 Schematic represenation of the Hippo signalling pathway components and how they interact. **a** Activated Hippo pathway - MST1 or MST2 phosphorylate SAV1 and, together, they phosphorylate and activate MOB1A, MOB1B, LATS1 and LATS2 kinases, which then phosphorylate YAP and transcriptional co-activator with TAZ. Phosphorylated YAP and TAZ are sequestered in the cytoplasm by the 14-3-3 protein and shunted for proteasomal degradation. As a result, the TEA domain-containing sequence-specific transcription factors (TEADs) associate with the transcription cofactor vestigial-like protein 4 (VGL4) and suppress target gene expression. **b** Inativated Hippo pathway - MST1, MST2, LATS1 and LATS2 kinases are inactive, so YAP and TAZ are not phosphorylated and instead accumulate in the nucleus where they relocate VGL4 and form a complex with TEADs, promoting the expression of target genes involved in cell survival, proliferation, metastasis, stem cell maintenance, regeneration and cell competiton. Created with Biorender. Adapted from Harvey et al., 2013 and Johnson et al., 2014.

4.2. Cellular community and cell motility gene sets are negatively regulated in TNBC-AA patients

We observed that all the enriched gene sets involved in cellular community and cell motility, from the comparisons between all TNBC patients, postmenopausal, stage II and postmenopausal stage II patients, are downregulated in TNBC-AA patients (Chapter IV, Figure 32 and Supplementary - GSEA eTables 1, 3, 5 and 11, respectively). Those gene sets include tight junctions, signaling pathways regulating pluripotency of stem cells, focal adhesion, gap junction, regulation of actin cytoskeleton and adherens junction (Chapter IV, Figure 34a-f, respectively). Interestingly, although being evenly downregulated in TNBC-AA patients, these enriched gene sets are composed by widely different DEG (Chapter IV, Figure 34).

The ability of epithelial cells to organize in a community through cell-cell contacts, involving reorganization of the actin cytoskeleton into a functioning epithelium, serves the purpose of a tight epithelial protective barrier. In epithelium-derived cancers, such as TNBC, intactness of the primary tumor mass is influenced by intercellular structures as well as cell-cell adhesion (Cavallaro and Christofori, 2004). Irregularities of these factors may lead to tumor dissociation and subsequent metastasis (Bogenrieder and Herlyn, 2003). Figure 52 displays a simplified schematic representation of the epithelial intercellular junctions and its components, which are fundamental pieces in organization, communication and motility of epithelial cells. These proteins mediate firm mechanical stability, assist as a gatekeeper for paracellular pathways, and help in preserving tissue homeostasis (Bhat et al., 2019).

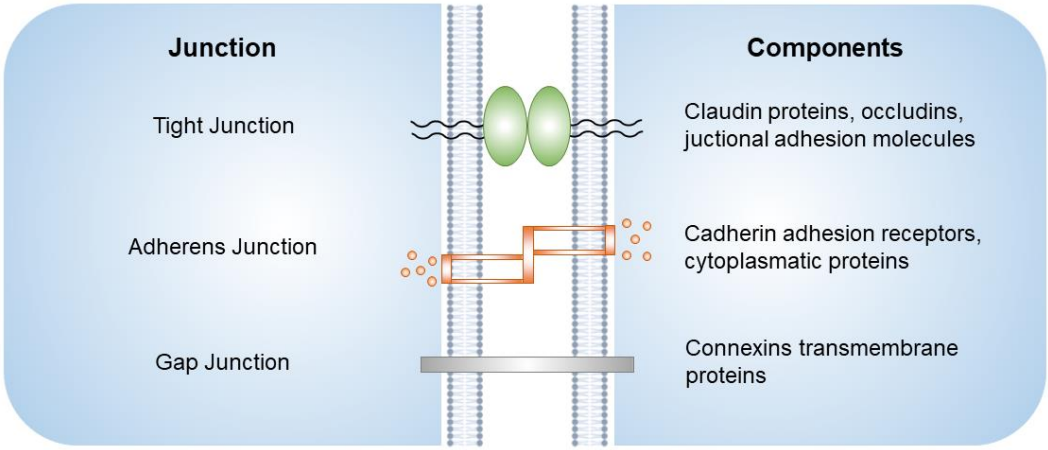


Figure 52 Schematic representation of epithelial intercellular junctions and its components in vertebrate cells. Adapted from Bhat et al., 2019.

Tight junctions (TJs) are among the structures that preserve cell adhesiveness of the tumor mass, suppressing cell proliferation (Matter and Balda, 2007). TJ proteins, such as occluding, claudin or junctional adhesion molecules (JAMs), are involved in maintaining cell polarity, in establishing organ-specific apical domains and also in recruiting signaling proteins involved in the regulation of various important cellular functions including proliferation, differentiation, and migration. The downregulation of TJ proteins leads to the loss of cell-cell associations and cell contact inhibition that results to uncontrolled growth, loss of adhesion to the basement as well as its degradation, essential for cancer cells to grow and metastasize (Martin and Jiang, 2009).

Similar to TJ, adherens junctions (AJ) complex is also involved in the formation and maturation of cell-cell contacts. While TJ regulate the paracellular pathway for the movement of ions and solutes in-between cells, AJ initiate cell-cell contacts and mediate the maturation and maintenance of the contact. AJ, which are usually more basal than TJ, are composed of cadherin receptors that bridge the neighboring plasma membranes via their homophilic interactions. Transmembrane cadherins associate with cytoplasmic proteins, catenins, which in turn bind to cytoskeletal components, such as actin filaments and microtubules (Hartsock and Nelson, 2008; Meng and Takeichi, 2009). Studies reported that loss of epithelial AJ, particularly E-cadherin, is involved in altered cell morphology and metastasis and tissue invasion in TNBC (Eades et al., 2015; Gwin et al., 2010; Lopes et al., 2012).

The third type of cell junctions, gap junctions (GJ), comprise arrays of intercellular channels formed by connexin proteins and provide for the direct communication between adjacent cells. This type of intercellular communication allows the coordination of cellular activities and plays key roles in

the control of cell growth and differentiation, and in the maintenance of tissue homoeostasis (Aasen et al., 2019; Leithe et al., 2006). Studies report that cancer cells have indeed downregulated levels of gap junctions, and several lines of evidence suggest that loss of gap junctional intercellular communication is an important step in carcinogenesis (Aasen et al., 2019; Leithe et al., 2006). In fact, several connexin knockout mouse models have supported the notion that connexins have tumor suppressive functions (Aasen et al., 2016). Connexin 43 (Cx43, *GJA1*) is the most frequently studied connexin in BC (Aasen et al., 2019), being often downregulated (Kazan et al., 2019; Teleki et al., 2014). *GJA1* is downregulated in TNBC-AA patients in all contrasts, however it is never significantly differentially expressed (i.e., FDR > 0.05). Nevertheless, *GJA1* downregulation and knock-down was reported to induced a mesenchymal phenotype with increased cell invasion leading to an enhanced metastatic phenotype (Kazan et al., 2019) and being associated with a poor prognosis in TNBC patients [111].

Interestingly, and contrary to the downregulation observed in the focal adhesion gene set (Chapter IV, Figure 32), focal adhesion kinases (FAKs) are usually significantly overexpressed in invasive and metastatic TNBC (Gari et al., 2016; Taliaferro-Smith et al., 2015). FAKs are non-receptor tyrosine kinases, being the most prominent signaling molecules in focal adhesions, hence its namesake. In order to metastasize, cancer cells must lose attachment with neighboring tumor cells and adopt the ability to migrate, attach to, and invade through the epithelial basement membrane. This is a complex process primarily orchestrated through the formation, stabilization and remodeling of focal adhesion complexes composed of FAK, integrins, ERK, Src and numerous adaptor proteins and downstream effectors, such as RhoGTPases, that collectively regulate EMT and trigger activation of cell migration and invasion programs (Burridge and Chrzanowska-Wodnicka, 1996; Wozniak et al., 2004; Zhao and Guan, 2009). FAKs activation by integrins in transformed cells in the absence of adhesion signal was proposed to be responsible for anchorage-independent growth of these cells, a hallmark of cancerous cells (Zhao and Guan, 2009). Furthermore, activation of FAK in cell adhesion protects cells from anoikis, a form of apoptosis induced by cell detachment from the extracellular matrix (ECM) (Frisch et al., 1996). Thus, dynamic regulation of FAKs activity contributes to tumorigenesis through promotion of cell survival, proliferation and metastasis. For these reasons, molecular targeting FAKs, promoting its downregulation, or upregulating its upstream regulators, such as the tumor suppressor PTEN (Zhang et al., 2014), are potential strategies for cancer therapy (Selvendiran et al., 2010). Having in mind the more aggressive behavior of TNBC development reported in African-ancestry women, including metastasis formation, the fact that TNBC-AA patients show a negative net direction of the focal adhesion gene set may be due to a number a reasons, such as the fact that genes included in the focal adhesion gene sets, which are not statistically significant (i.e., FDR > 0.05), are influencing the result. Another reason may be the effect of the downregulation of *EGFR*, observed in the gene sets with postmenopausal and/or stage II TNBC-AA patients (Chapter IV, Figure 34c and Supplementary - GSEA file, eTables 1, 3, 5 and 11). Mechanistically, EGF modulates adhesion in a dual fashion, by firstly redistributing focal adhesion elements to adhesion sites, but also by amplifying levels of activated RhoA antagonist p190RhoGAP, important for cell motility. Thus, EGF response underlies an EGFR-integrin cross-talk which involves the recruitment of receptor proximal FAK and Src, and MAP kinase and p190RhoGAP as receptor distal events, promoting cell mobility and metastasis (Eberwein et al., 2015). In turn, low *EGFR* expression was reported to promote inhibition of FAKs, causing detachment and apoptosis via pathways involving activation of caspase-3 and -8, cleavage of poly(ADP-ribose) polymerase, and caspase-3-dependent degradation of AKT, in BC cell lines (Golubovskaya et al., 2002).

Signaling pathways regulating pluripotency of stem cells gene set is also downregulated in TNBC-AA patients (Chapter IV, Figure 32). Very little is known regarding breast CSC patterns African-ancestry women (Jiagge et al., 2018), although some studies suggest that African-ancestry patients have BC cases enriched with breast CSC markers (Ginestier et al., 2007; Jiagge et al., 2018; Nalwoga et al., 2010; Schwartz et al., 2013) (Chapter I, Section 3.3.2.5.). Hence, the rather surprising observation that TNBC-AA patients have a negative regulation of the genes involved in pluripotency of stem cells. However, interestingly, studies suggested that downregulation of proteins involved in GJ, as observed in the downregulation of GJ gene set in TNBC-AA patients (Chapter IV, Figure 32), might be associated with CSC characteristics (Beckmann et al., 2019). Thus, given that gene sets from the cellular community and mobility KEGG collection are interconnected, although the signaling pathways regulating pluripotency of stem cells gene set has a negative net direction in TNBC-AA patients, when comparing with matching TNBC-White patients, this result is possibly being influenced by the genes also involved in other gene sets from this collection.

Similarly, regulation of actin regulation of actin cytoskeleton gene set with a negative net direction is expected, since both TJ and AJ complexes are associated with the actin cytoskeleton, and formation and maturation of cell-cell contacts involves reorganization of the actin cytoskeleton (Hartsock and Nelson, 2008).

All in all, the results observed regarding the cellular community and mobility gene sets in TNBC-AA patients suggest that TNBCs in this population might have a higher preponderance for the loss of cell-cell associations and cell contact inhibition and, consequently, for a higher cell proliferation, cell mobility and metastasis formation and tissue invasion, when comparing with matching TNBC-AA patients.

4.3. African-ancestry women have a higher activation of components and processes involved in BC development

Interestingly, the enriched Breast Cancer gene set is upregulated in TNBC-AA patients, when comparing with their White counterparts (Chapter IV, Figures 32 and 35). BC gene set is composed by well-known genes involved in the different subtypes of BC, including TNBC. According to KEGG's information regarding BC gene set "In patients suffering from TNBC, the deregulation of various signaling pathways (Notch and Wnt/β-catenin) and EGFR protein have been confirmed" (https://www.genome.jp/kegg-bin/show_pathway?hsa05224, as of June 8, 2020). Accordingly, genes involved in the Notch and Wnt/β-catenin pathways, such as *WNT7*, *WNT5B*, *FZD9*, *FZD8*, *NOTCH3*, *MYC* or *CTNNB1*, as well as *EGFR* itself, are among the DEG included in the BC gene set (Chapter IV, Figure 35).

Thus, the upregulation of the BC gene set in TNBC-AA patients, which includes genes and pathways that promote the activation of components and processes involved in tumorigenesis, may be translated into TNBC's faster progression, more aggressive behavior and lower survival probability observed in these patients, comparing with White patients.

4.4. MAPK signaling pathway is upregulated in stage II disease TNBC-AA patients

It was observed that MAPK signaling pathway gene set is exclusively upregulated in stage II disease AA patients, comparing with matching White patients (Chapter IV, Figure 32). Mitogen-activated protein kinase (MAPK) pathways are an extensive regulatory network consisting of a series of cross-talking and compensatory pathways affecting a variety of cellular processes, including proliferation, differentiation, apoptosis, stress responses, growth signals and metabolism (Guo et al., 2020; Plotnikov et al., 2011). Thus, deregulation of this highly interconnected signaling cascade is frequently involved in oncogenesis, tumor progression, and drug resistance (Braicu et al., 2019; Guo et al., 2020). In fact, due to a high degree of interactions and possible compensatory responses, most MAPK inhibitors trigger resistance due to the activation of compensatory feed-back loops in tumor cells and tumor microenvironment components (Braicu et al., 2019). Nonetheless, an insight into these compensatory mechanisms promoting chemoresistance is still lacking (Braicu et al., 2019). In TNBC patients, it was observed that an increased genomic or transcriptomic activation of the MAPK signaling pathway was correlated to a reduced amount of tumor-infiltrating lymphocytes (TIL) in the residual disease after neoadjuvant-chemotherapy (NACT), which resulted in a poorer prognosis and survival of these patients (Loi et al., 2016). Accordingly, high levels of TILs in the tumor are predictive of pathological complete response (pCR) to NACT, and increased disease-free survival and overall-survival (OS) of TNBC patients (Adams et al., 2014; Loi et al., 2014). Although the mechanism has not yet been elucidated, authors speculate that, if MAPK activity can suppress expression of MHC-I and MHC-II, both intrinsically and those induced by IFNβγ, according to their results (Loi et al., 2016), then tumor cells can circumvent antigen presentation pathways by activating the MAPK pathway, reducing TIL recruitment (Loi et al., 2016). Jiang et al. (Jiang et al., 2020) also observed an increased expression of MAPK in TNBC tissue. These authors observed that the positive expression rates of MAPK in patients with lymph node metastasis, advanced clinical stage, tumor recurrence and metastasis was higher than those without. Also, patients with positive expression of MAPK in TNBC tissues had poorer prognosis and lower overall survival times than those without expression (Jiang et al., 2020).

Thus, the increased activation of MAPK signaling pathway specifically in TNBC-AA patients presenting stage II disease at diagnosis may be contributing to the observed lower survival of this group of patients when comparing with their White counterparts.

4.5. EGFR expression is reduced in TNBC diagnosed in African-ancestry patients

We observed that EGFR expression is downregulated in TNBC-AA patients (Supplementary – DGEA file eTables 1, 3, 5 and 11). Remarkably, *EGFR* is frequently reported as being overexpressed in TNBC, and its high expression and copy number variation recognized as a factor of poor prognosis and associated with shorter disease-free survival and aggressive biological properties of TNBC (Ali and Wendt, 2017; Foidart et al., 2019; Nakai et al., 2016; Nielsen et al., 2004; Park et al., 2014; Viale et al., 2009).

Furthermore, *EGFR* is involved in a number of significantly enriched KEGG gene sets (Chapter IV, Figure 32), such as MAPK, Ras, Phospholipase D, PI3K-Akt, Rap1, FoxO, ErbB2, Jak-STAT and HIF-1 signaling pathways from the Signaling transduction collection; all Cellular community and motility gene sets except for TJ and Signaling pathways regulating pluripotency of stem cells gene sets; Oxytocin and Estrogen signaling pathways from the Endocrine system collection; Pathways in cancer; Breast cancer,

Proteoglycans in cancer, MicroRNAs and EGRF tyrosine kinase inhibitor resistance from the conjoint collection of Cancer: overview and Drug resistance: antineoplastic. Interestingly, with few exceptions, almost all of these enriched gene sets follow the same negative net direction as *EGFR* expression in TNBC-AA patients.

IPA's core analysis also revealed EGFR as one of the top central network regulators in the groups with all TNBC-AA patients, postmenopausal, stage II disease and postmenopausal stage II disease TNBC-AA patients (Chapter IV, Figure 37 and Supplementary - Networks file eTables 1, 3, 5 and 7, respectively). Interesting, complementing the suggested *EGFR* influence in the KEGG collection of Cellular community and mobility gene sets, "Cellular Movement" was one of the top diseases and functions associated with the EGFR network design in the contrasts with all TNBC patients, stage II disease patients and postmenopausal stage II disease patients (Chapter IV, Figure 37, a, c and d, respectively).

Given these contrasting results between the low expression of *EGRF* observed in TNBC-AA patients from the TCGA and the reportedly EGFR activation and overexpression paradigm in cancer and TNBC (Ali and Wendt, 2017; Nakai et al., 2016) (Chapter I, Section 3.2.4.), EGFR protein expression was further validated through immunohistochemistry (IHC) of formalin-fixed paraffin-embedded (FFPE) tissue samples from TNBC patients followed in Portugal. We observed that EGFR protein was significantly less expressed in African-ancestry patients than in White patients (Chapter IV, Figures 46 and 47).

Furthermore, *EGFR* is not differentially expressed in normal-adjacent BC tissue from AA patients when comparing with matching White patients (Supplementary - DGEA file, eTable 13), nor in TNBC-AA when comparing with normal-AA cases (Supplementary - DGEA file, eTable 14). Additionally, *EGFR-AS1*, the EGFR Antisense RNA 1, is also downregulated in TNBC-AA patients, in the same groups of patients as *EGFR* (Supplementary - DGEA file, eTables 1, 3, 5 and 11), suggesting that, indeed, there is no amplification of the EGFR chromosomic location in TNBC-AA patients. Furthermore, a recent TCGA-based work (Omilian et al., 2020) confirmed that White-TNBC patients tend to have a higher copy number variation (CNV) in the *EGFR* region, when comparing with TNBC-AA patients. Nonetheless, those results were not statistically significant (Omilian et al., 2020).

EGFR belongs to the ErbB family of receptor tyrosine kinases (RTKs) and exerts critical functions in epithelial cell physiology (Schlessinger, 2014). EGFR is a transmembrane protein comprising an extracellular ligand binding domain, transmembrane domain, and cytoplasmic tyrosine kinase domain (Burgess, 2008; Hynes and MacDonald, 2009; Yarden, 2001). EGFR transphosphorylation, which usually occurs in response to ligand stimulation, such as EGF, leads to the activation of the intracellular signaling cascade. Briefly, when EGFR is transphosphorylated it forms a dimer, turning on its kinase activity, followed by autophosphorylation at multiple tyrosine residues in the intracellular region to recruit various substrates. EGFR activation promotes crucial cellular functions, such as cell proliferation, motility, and survival via activation of a myriad of downstream signaling pathways, including Ras-Raf-MEK-ERK (MAPK), PI3K-AKT-mTOR, or Src-STAT3 (Schlessinger, 2014; Yarden and Sliwkowski, 2001). Ligand-activated EGFR molecules are then ubiquitinated, internalized, and isolated in endosomes (Sigismund et al., 2018). Furthermore, EGFR can also translocate into the nucleus, being involved in transcriptional regulation, DNA replication, and DNA repair (Lee et al., 2015; Sigismund et al., 2018). Figure 53 displays a simple schematic representation of the current knowledge regarding the downstream pathways' activators, potential inhibitors and outcomes at cellular level of EGFR signaling.

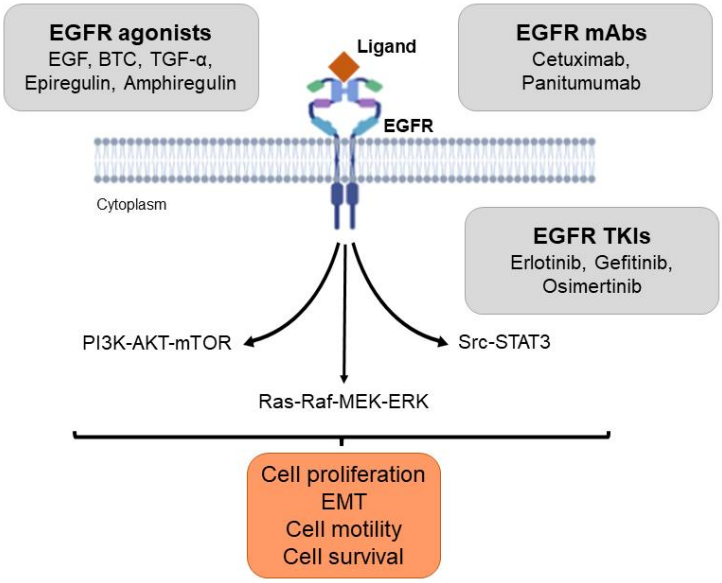


Figure 53 Schematic representation of the downstream activators, inhibitors and outcomes of EGFR signaling. EGFR is capable of binding several different extracellular ligands that agonize the receptor leading to activation of several downstream signaling cascade events, including Ras-Raf-MEK-ERK, PI3K-AKT-mTOR, or Src-STAT3, among others. Activation of such pathways may lead to several cellular events, including cell proliferation, epithelial-to-mesenchymal transition (EMT), cell motility and cell survival. Several therapeutics have been developed to antagonize EGFR including monoclonal antibodies (mAbs) that block ligand binding, as well as several different tyrosine kinase inhibitors (TKIs). EGFR agonists have been also recently studied due to their potential therapeutic role in cancer. Created with Biorender. Adapted from Ali et al., 2017.

As previously mentioned, EGFR signaling is frequently altered in several human cancers due to EGFR gene amplification and/or protein overexpression, mutations or in-frame deletions (Sigismund et al., 2018). Also, it has been reported that impaired EGFR degradation through endocytosis is critical for further upregulation of EGFR protein in some types of cancer cells, including BC cells (Vivanco et al., 2010; Zhang et al., 2013). Additionally, inhibition of BRCA1 has been shown to induce upregulation of EGFR mRNA and protein in breast and ovarian cancer cells (Burga et al., 2011; Li et al., 2013). Nonetheless, overall, molecular mechanisms regarding EGFR mRNA and protein role in tumorigenesis, particularly in BC, remain mostly unknown.

Also, few studies reported *EGFR* underexpression as being associated with BC progression and metastasis formation (Choong et al., 2007; Wendt et al., 2015), and patients' poor prognosis and worse disease outcome (Kreike et al., 2010). Interestingly, Kreike et al. study (Kreike et al., 2010) suggests that *EGFR*-expression in BC patients has a nonlinear relation with disease outcome, with both lower and higher expression of *EGFR* being associated with poor prognosis and a worse outcome compared to intermediate expression levels.

In line with these findings, our results also challenge the most conventional wisdom in which *EGFR* amplification and overexpression is the norm in BC and TNBC and that is associated with patients' poor prognosis. Furthermore, since most TNBC studies are performed in countries with predominantly White populations, our results also suggest that the role of *EGFR* overexpression in cancer development and poor prognosis, particularly in TNBC, should not be generalized. Since the massively studied small-molecule inhibitors and monoclonal antibodies are being developed and tested under

the supposition that most cancers present *EGFR* overexpression, such generalization might negatively impact African-ancestry patients' treatment, as well as patients from other ancestries (Nakajima et al., 2014). The lack or lower *EGFR* expression when using target therapies may potentially enhance secondary effects, such as impairment of essential related networks and pathways.

4.6. MYC oncogene has an increased expression in TNBCs from African-ancestry patients

Here we highlight MYC mRNA and protein overexpression in African-ancestry patients in comparison with their White counterparts. This result is in line with the fact that MYC is a well-known player in carcinogenesis and associated with patients' poor prognosis.

MYC oncogene encodes a transcription factor, Myc, whose broad effects make its precise oncogenic role puzzlingly elusive. Myc expression and activity are tightly regulated in normal cells by multiple mechanisms, involving a number of signaling pathways, transcription factors, and cis regulatory elements, as well as an dependence upon growth factor stimulation (Xu et al., 2010). Wnt/β-catenin signaling, Notch, MAPK and NF-κB pathways are, among others, regulators of the *MYC* promoter (Xu et al., 2010). Myc activation triggers selective gene expression amplification to promote cell growth and proliferation, including key genes involved in ribosomal and mitochondrial biogenesis, glucose and glutamine metabolism, lipid synthesis, and cell cycle progression. To robustly activate its targets, Myc also coordinates the acquisition of bioenergetics substrates to produce ATP and key cellular building blocks that increase cell mass and trigger DNA replication and cell division (Stine et al., 2015).

In the cancer context, unrestraint *MYC* expression and loss of checkpoint components, such as p53, allow Myc to drive malignant transformation through cancer cell growth and proliferation, EMT and angiogenesis (Dang, 2012; Xu et al., 2010). Also, one of the mechanisms underlying the pro-tumor properties of oncogenic Myc is likely its function in antagonizing activation of induced senescence (Wu et al., 2007). Thus, downregulation of Myc is a critical event for growth inhibition through cell cycle arrestment, which can be induced by TGF-β (Frederick et al., 2004). Furthermore, Myc amplification is a frequent event in BC from *BRCA1* germ-line mutation carriers and in sporadic tumors with *BRCA1* inactivation owing to *BRCA1* promoter hypermethylation (Grushko et al., 2004; Wei et al., 2005), suggesting that Myc overexpression occurs after loss of *BRCA1* during the progression of tumorigenesis (Xu et al., 2010). Nonetheless, how Myc regulates BRCA1 remains unclear. Figure 54 displays a schematic representation of the pathways involved in Myc regulation, as well as downstream effects of Myc deregulation.

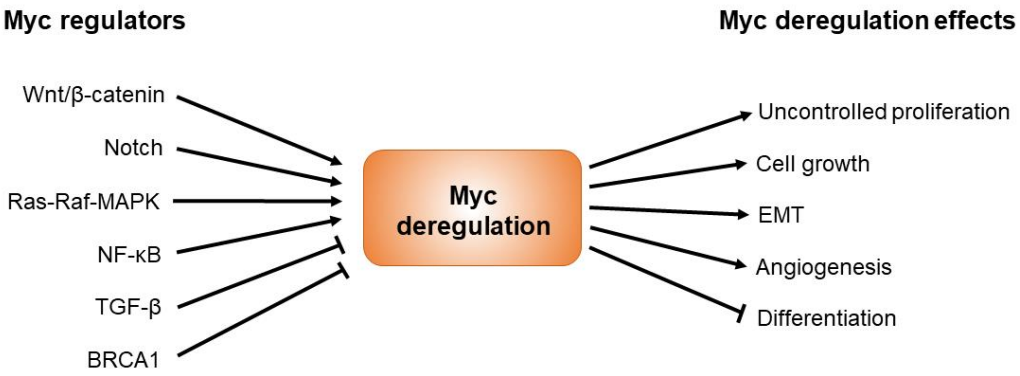


Figure 54 Schematic representation of Myc regulatory network. Myc is regulated in multiple levels by different pathways, including Wnt/β-catenin signaling, Notch pathway, Ras-Raf-MAPK, NF-κB pathways. Yet, loss of tumor suppressors and activation of these oncogenic pathways can affect Myc expression or stability, leading to its deregulation. TGF-β and BRCA1, which can reportedly repress MYC gene expression and its transcriptional activity, respectively, are frequently inactivated by mutation or epigenetic mechanisms in breast cancer. Deregulated Myc promotes tumor growth, epithelial-to-mesenchymal transition (EMT), cell undifferentiation and angiogenesis. Adapted from Xu et al., 2010.

Previous studies have demonstrated that various cancers are Myc-dependent and that its inactivation leads to tumor regression in multiple preclinical tumor models (Wu et al., 2007). Also, due to the reliance of Myc-driven cancers on specific metabolic pathways, synthetic lethal interactions between Myc increased expression and specific enzyme inhibitors may provide novel cancer therapeutic opportunities (Stine et al., 2015). In addition to its role in tumorigenesis, Myc was also identified as one of four genes, including Sox2, Oct4, and KLF4, that could collectively reprogram fibroblasts to a pluripotent stem cell state (Laurenti et al., 2009; Singh and Dalton, 2009; Takahashi and Yamanaka, 2006).

Expression of the oncogenic *MYC* is described as being disproportionately elevated in TNBC, as compared other BC subtypes (Carey et al., 2018; Horiuchi et al., 2012; Koboldt et al., 2012; Lee et al., 2017). Furthermore, those studies show evidence that overexpression of *MYC* contributes to CT resistance in TNBC (Carey et al., 2018; Lee et al., 2017). Here we report that *MYC*, which is already stated as being overexpressed in TNBC in studies mostly composed by patients of White-ancestry or cell lines (Carey et al., 2018; Horiuchi et al., 2012; Koboldt et al., 2012; Lee et al., 2017), as having an increased overexpression in TNBC-AA patients. This increased *MYC* overexpression was observed in the contrasts with all TNBC patients, postmenopausal, stage II disease, and postmenopausal stage II disease patients (Supplementary - DGEA eTables 1, 3, 5 and 11, respectively). In line with this observation, one of the previous TCGA-based studies exploring TNBC's racial discrepancy reported that TNBC-AA patients have a higher proportion of *MYC*/8q24.21 amplification when compared with TNBC-White patients (30.9% vs 20.4%), independently of other variables (Huo et al., 2017).

Remarkably, with few exceptions, most of the enriched gene sets including *MYC* have a negative net direction in TNBC-AA patients, such as Hippo, MAPK, Wnt, PI3K-Akt, ErbB, Jak-STAT and TGA-beta signaling pathways from the Signaling transduction collection; Cellular senescence and Cell cycle from the Cell cycle and death collection; Signaling pathways regulating pluripotency of stem cells from the Cell community and motility collection; and Pathways in cancer, Breast cancer, Proteoglycans in

cancer, MicroRNAs in cancer and Transcriptional misregulation in cancer from the Cancer collection (Chapter IV, Figure 32). Although downregulated genes with larger fold-change (FC) and/or lower FDR might be affecting the observed net direction, true role and influence of *MYC* overexpression in the aforementioned signaling pathways, cellular mechanisms and diseases should be further investigated.

Myc as an interesting marker in TNBC-AA patients is also highlighted by IPAs' network design, in which Myc is a central network regulator in the contrasts with all TNBC patients, postmenopausal patients and postmenopausal stage II disease patients (Chapter IV, Figure 38 and Supplementary - Networks file eTables 1, 3 and 7, respectively).

Myc protein expression was then investigated through IHC of FFPE samples from TNBC patients followed in Portugal. Myc H-score was significantly higher in African-ancestry patients than in White patients ($p = 4.01E-04$, Chapter IV, Figures 46 and 48), consistent with the mRNA overexpression observed in TNBC-AA patients in the study cohort.

MYC is not differentially expressed in normal-adjacent BC tissue between AA patients and matching White patients (Supplementary – DGEA file eTable 13), nor is differentially expressed in TNBC-AA patients compared with normal-AA cases (Supplementary – DGEA file eTable 14). Overall, these results support the hypothesis that Myc expression may be, indeed, more upregulated in TNBC patients from African-ancestry, hence contributing to an increased chemoresistance and poor prognosis (Carey et al., 2018; Lee et al., 2017) in this population.

4.7. BCL2 downregulation is characteristic of TNBC development in premenopausal African-ancestry patients

We observed that *BCL2* gene expression was downregulated in premenopausal TNBC-AA patients when comparing with their White counterparts and with normal-adjacent BC tissue from AA patients (Supplementary - DGEA file, eTables 2 and 14, respectively), suggesting that the downregulation of *BCL2* is an acquired trait in TNBC development in AA patients, particularly in those diagnosed at a younger age. In fact, African-ancestry patients are reported as being diagnosed with TNBC at a significantly younger age than White patients, frequently in pre-menopause, as observed in our group of patients from the validation study (Chapter IV, Table 24) and in other epidemiological studies (Amirikia et al., 2011; P. Boyle, 2012; Carey et al., 2006; Clarke et al., 2012; Fregene and Newman, 2005; Newman et al., 2019; Stark et al., 2010). Although the number of DEG obtained in the premenopausal contrasts was not enough to generated significantly enriched gene sets (Supplementary - GSEA file, eTable 2), IPAs' core analysis confirms BCL2 as a core network regulator (Chapter IV, Figure 39), influencing "Cancer", "Organismal Injury and Abnormalities" and "Skeletal and Muscular System Development and Function" diseases and functions, and having a network score of 47 (Supplementary - Networks file, eTable 2), a score which is only second to the network score (49) of EGFR as a core regulator in the post-menopause stage II disease contrast. Thus, we decided to include this candidate gene for validation through IHC, where we confirmed an under-expression of Bcl2 in African-Ancestry patients (Chapter IV, Figures 46 and 49), most of which were premenopausal (Chapter IV, Table 24). Hence, here we highlight Bcl2 as an ancestry-associated marker in TNBC cases diagnosed in premenopausal African-ancestry women.

Bcl2-family proteins play central roles in cell death regulation and are able of regulating diverse cell death mechanisms that encompass apoptosis, necrosis and autophagy (Cory and Adams, 2002; Yip and Reed, 2008). These family of proteins act in response to diverse intracellular damage signals,

including those evoked by cancer therapy. The damage signals are transduced by the following "BH3-only" proteins: Bcl2, Bcl-xL, Bcl-w, Mcl-1 and A1. These proteins are distinguished by the BH3 domain used to engage their pro-survival relatives (Cory and Adams, 2002). Cell's decision to undergo apoptosis is determined by interactions between different members of the Bcl2 protein family. Bcl2 and several pro-survival relatives associate with the mitochondrial outer membrane and the endoplasmic reticulum/nuclear membrane and maintain their integrity. Initiation of apoptosis requires not only pro-apoptotic family members such as Bax and Bak, but also distant Bcl2-family cousins, such as Bid, being related only by the small BH3 protein-interaction domain (Cory and Adams, 2002). Essentially, the BH3-only proteins are sentinels that detect developmental death cues or intracellular damage. In healthy cells, they are restrained in diverse ways, including sequestration on the cytoskeleton. When unleashed by death signals, they switch off survival function by inserting their BH3 domain into a channel on their pro-survival relatives. During apoptosis, Bax and Bak oligomerize in the mitochondrial outer membrane, breaching its integrity and freeing proapoptotic proteins such as cytochrome c, which allows activation of caspase-9, as well as other "killer" molecules, including Omi/HtrA2 and Diablo/Smac, which antagonize the inhibitor of apoptosis proteins which, in turn, would inhibit processed caspases (Cory and Adams, 2002). Bcl2 can prevent cytochrome c release and, consequently, caspase-9 activation and apoptosis. Thus, the balance between the pro-survival proteins and their BH3 ligands regulates tissue homeostasis (Reed, 1998; Yip and Reed, 2008). Figure 55 shows the intracellular signal mediated by the Bcl2 family upon cellular stress.

Thus, Bcl2 acts by promoting cell survival rather than by driving cell proliferation (Vaux et al., 1988). Bcl2 suppresses apoptosis indirectly by blocking the mitochondrial release of cytochrome c and inhibiting the activation of caspases by the cytoplasmic scaffolding protein Apaf1 (apoptosis activating factor-1) (Brichese et al., 2002; Kaufmann and Gores, 2000; Zörnig et al., 2001). Bcl2 may also protect against chromosomal damage by promoting cell-cycle arrest under suboptimal conditions. Expression of Bcl2 is cell-cycle regulated, peaking during mid-G1 and contributing to S-phase delay in response to some stimuli (Zinkel et al., 2006). Maintenance of microtubule integrity may also depend on Bcl2 because drugs that affect tubulin structure, such as taxanes, phosphorylate and inactivate the Bcl2 protein (Haldar et al., 1996). Indeed, Bcl2-overexpressing tumors often show gross chromosomal aberrations (Petrini et al., 2012).

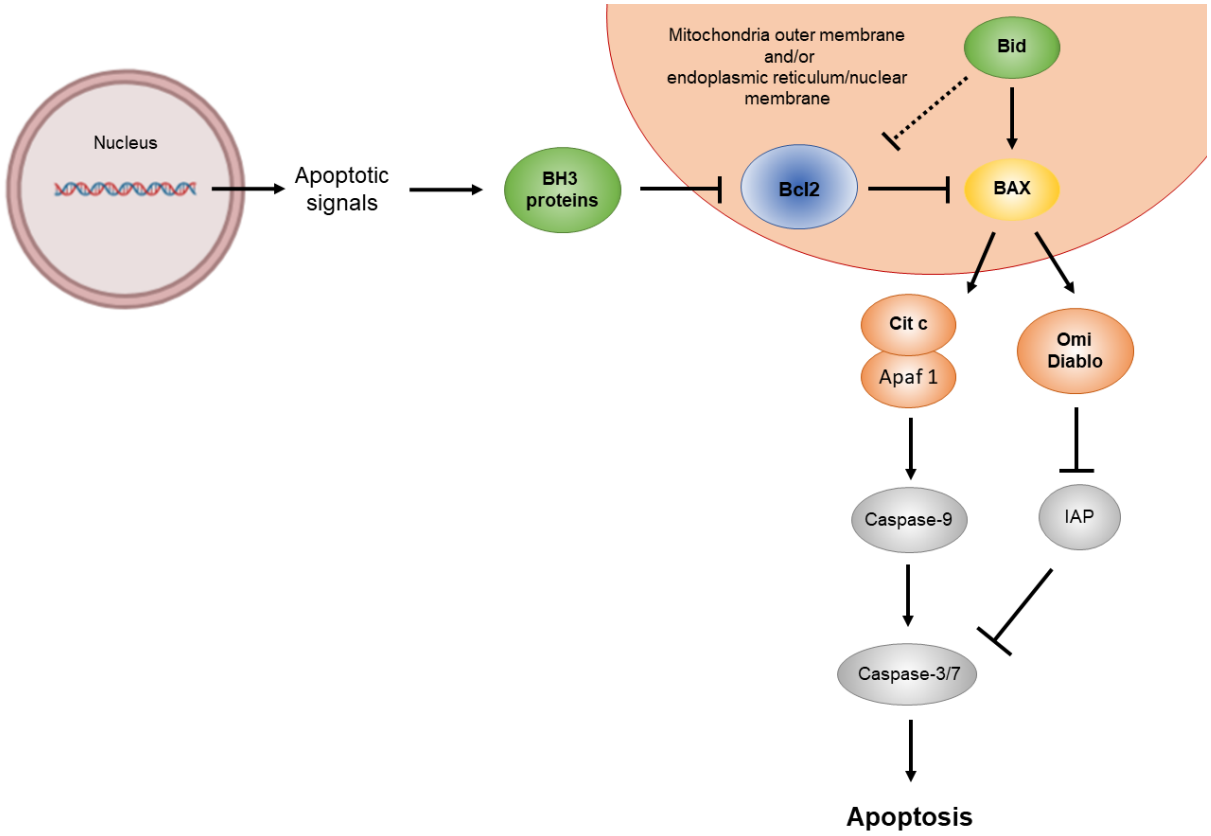


Figure 55 Intracellular stress signals are mediated through the Bcl2 family. Activity of caspase-9 and caspase-3 is restrained by inhibitor of apoptosis proteins (IAPs), but the IAPs can be countermanded by Diablo/Smac and Omi/HtrA2, which are released from the mitochondria. Apaf1, apoptotic protease-activating factor 1. cyt c, cytochrome c. Adapted from Cory et al., 2002, and Adams et al., 2007.

Impaired apoptosis promoted by, *inter alia*, overexpressed Bcl2 is, in fact, a crucial step in tumorigenesis since a defective suicide program bestows emerging neoplastic cells with multiple selective advantages. Such advantages include survival capacity in hostile niches (for instance, where cytokines or oxygen are limiting). This death evasion mechanism allows cells to evolve into more-aggressive derivatives. Thus, defective apoptosis contributes to metastasis formation, as the cells can ignore restraining signals from neighbors and survive detachment from the extracellular matrix (Adams and Cory, 2007; Cory and Adams, 2002). Impaired apoptosis also significantly affects cytotoxic therapy since the mutations that favored tumor development dampen the response to CT and radiation, and treatment might select more refractory clones (Johnstone et al., 2002). Thus, overall, targeting Bcl2-expressing tumors seems to be indeed an interesting approach to enhance the effect of chemotherapeutic agents (Beh et al., 2009; Douglas Hanahan and Robert A. Weinberg, 2000; Lima et al., 2004; Papadopoulos, 2006). Nonetheless, the effects in developing cancers presenting low or absent expression of Bcl2, as well as implications in patients' prognosis and treatment are, comparably, poorly explored and the existing reports are contradictory.

In earlier studies, *BCL2* over- and under-expression effect on BC patients' prognosis and treatment has had mixed results according to the BC subtype. *BCL2* is upregulated by estrogens in BC, through a direct consequence of transcriptional induction (Leung and Wang, 1999) and, thus, *BCL2*-positive expression in BC may be considered a sign of ER functional activity and is associated with more favorable outcomes (Aleskandarany et al., 2015; Bouchalova et al., 2015; Choi et al., 2014; Eom et al.,

2016; Leung and Wang, 1999). In fact, studies have suggested *BCL2* as a reliable prognostic marker for HR-positive BCs (Choi et al., 2014), since HR-positive BC patients expressing *BCL2* have a better prognosis considering the OS and relapse-free survival (Eom et al., 2016; Kim et al., 2012). Additionally, *BCL2*-positive expression was associated with better outcomes in metastatic and early BC treated with either hormone therapy or chemotherapy (Eom et al., 2016; Gasparini et al., 1995; Tsutsui et al., 2006). Furthermore, Eom et al. (Eom et al., 2016) observed that Bcl2-expressing BC patients (comprising 53.8% of all 1356 BC cases) had lower histological grade ($p < 0.001$), low Ki-67 level (< 14%, $p < 0.001$), low relapse rate ($p = 0.016$) and, importantly, had more frequently luminal A BC ($p < 0.001$) (Eom et al., 2016).

However, studies investigating Bcl2 levels in TNBC cases and its association with clinicopathological factors are not so straightforward as with HR+ BCs. A study reported an association between Bcl2 expression and better survival in patients with TNBC (Hwang et al., 2017). Also, in Abdel-Fatah et al. study (Abdel-Fatah et al., 2013), when controlling for CT and other prognostic factors, absence or low Bcl2 expression was associated with 2-fold increased risk of death ($p = 0.006$) and recurrence ($p = 0.0004$) in early primary TNBC cases. Remarkably, when Bcl2-negative patients received adjuvant-anthracycline combination-CT, both BC-specific survival ($p = 0.002$) and disease-free survival ($p = 0.003$) was improved (Abdel-Fatah et al., 2013). However, on contrary, Tawfik et al. (Tawfik et al., 2012) and Ozretic et al. (Ozretic et al., 2018) observed that the increased expression of Bcl-2 was an independent prognostic factor for poorer OS in TNBC and, as such, a significant marker for tumor aggressiveness (Ozretic et al., 2018). The reason for different results among these studies is not clear but different patient sample size and origin (Abdel-Fatah et al. study (Abdel-Fatah et al., 2013) encompasses 736 TNBC patients diagnosed in Nottingham, UK, whereas Tawfik et al. (Tawfik et al., 2012) and Ozretic et al. (Ozretic et al., 2018) included 124 patients diagnosed in Kansas, USA, and 64 patients from Zagreb, Croatia, respectively), as well as differences in the IHC protocols or other factors in the studies may play a role.

Abdel-Fatah et al. (Abdel-Fatah et al., 2013) also observed that Bcl2- patients treated with NACT anthracycline combination-CT are more likely to have a pCR than Bcl2+ patients. Given these results, authors suggested that, since Bcl2 protein acts as an anti-apoptotic factor and is also able to promote cell cycle arrest at G0, Bcl2+ cells are likely to recover from damage caused by CT (Abdel-Fatah et al., 2013), which may contribute to the poor prognosis observed in patients presenting Bcl2+ TNBC in other studies (Ozretic et al., 2018; Tawfik et al., 2012). On the other hand, Bcl2- TNBCs respond to treatment through accumulation of DNA damage, abnormal mitoses and subsequent mitotic catastrophe (Castedo et al., 2004). Importantly, Bcl2- TNBCs that do not respond to CT may be able to escape through accumulation of genetic abnormalities that would not lead to a mitotic catastrophe but rather to aneuploidy and subsequent growth advantage. In fact, these authors (Abdel-Fatah et al., 2013) observed that loss of Bcl2 in TNBC was associated with higher mitotic index, low levels of p27 (G0/G1 check protein), MDM4 (p53-inhibitory factor) and SPAG5 (essential for cell cycle progression and fidelity of chromosomal segregation) (Abdel-Fatah et al., 2013).

Explanations regarding our results, specifically about the lower/negative expression of Bcl2 in African-ancestry TNBC patients, are likely to be complex and involve a myriad of Bcl2 functions that require further clarifications. Overall, Bcl2 may be indeed a promising prognosis and treatment marker for TNBC patients. Although our results are not completely in line with previous studies (Abdel-Fatah et al., 2013; Ozretic et al., 2018; Tawfik et al., 2012), our observations suggest that loss of expression of *BCL2*/Bcl2 must have a substantial role in TNBC development in African-ancestry patients, particularly in those diagnosed in during pre-menopause state.

4.8. β-Catenin is downregulated in TNBCs from African-ancestry patients

We observed that *CTNNB1* is downregulated only in postmenopausal stage II disease TNBC-AA patients, comparing with matching TBBC-White patients (log_2FC = -0.935, FDR = 1.63E-02, Supplementary – DGEA file, eTable 11). We also observed that Wnt signaling pathway is indeed one of the most enriched gene sets observed in TNBC-AA patients, having a negative net direction (Chapter IV, Figure 32). Besides Wnt signaling gene set, *CTNNB1* is also downregulated in the Hippo signaling pathway gene set from the Signal transduction collection; Focal adhesion and Adherens junctions gene sets from the Cell community and motility collection; and in the and Breast cancer (Chapter IV, Figures 33, 34c, 34f and 35, respectively) and Pathways in cancer gene sets from the Cancer collection.

Through IHC, we observed that, as in the group of postmenopausal stage II disease patients, African-ancestry patients also express less β-catenin comparing with White patients, which show a significantly higher H-score values (p = 1.44E-04, Chapter IV, Figures 46 and 50). Interestingly, all White patients expressed β-catenin, with few expressing less β-catenin in some areas of the tissue.

β-catenin is an important intermediate in several signal transduction pathways, playing a central role in canonical Wnt signaling pathway, as well as an intermediate in many other signal transduction pathways, such as the PI3K-AKT pathway. β-catenin, which is encoded by the oncogene *CTNNB1*, was first identified as one of several proteins that were tightly bound to the C-terminus of classical cadherins (Ozawa et al., 1989). As its name suggests (Latin *catena*, meaning chain), it provides a critical link in a chain of proteins that strengthen cell-cell adhesion by coupling cadherins to the actin cytoskeleton. In the absence of Wnt ligands, β-catenin is kept at a low level through the ubiquitin proteasome system (Shang et al., 2017). In the breast, in addition to creating and stabilizing the mammary gland architecture, these catenin-cadherin complexes also exert effects on cell fate, polarity, motility and survival. Also, β-catenin together with T-cell factor/lymphoid enhancer factor (TCF/Lef) DNA binding factors act as a switch to determine cell fate and promote cell survival and proliferation at several stages during mammary gland development (Hatsell et al., 2003).

In the past two decades, the role and impact of *CTNNB1*/β-catenin expression and misregulation in BC have been massively studied (Bànkfalvi et al., 1999; Chung et al., 2004; Dillon et al., 1998; Dolled-Filhart et al., 2006; Fanelli et al., 2008; Karayiannakis et al., 2001; Lin et al., 2000; López-Knowles et al., 2010; Nakopoulou et al., 2000; Prasad et al., 2008), following the crescent number of studies involving the targeting Wnt/β-catenin signaling pathway in many tumors, especially colorectal carcinomas (Cheng et al., 2019; Krishnamurthy and Kurzrock, 2018; O'Toole et al., 2013; Shang et al., 2017; van Schie and van Amerongen, 2020). Nuclear accumulation of β-catenin resulting from aberrant Wnt signaling or mutation of the β-catenin gene leads to BC through upregulation of the expression of pro-invasive proteins and triggering EMT (Hatsell et al., 2003; Sánchez-Tilló et al., 2011). However, the associations with clinicopathologic variables and disease outcome reported in different studies are not always concordant. Although mutations in *CTNNB1* are rare in BC, unlike in colon or hepatocellular cancers (López-Knowles et al., 2010), altered β-catenin expression has previously been shown to be a prognostic marker in BC (Khramtsov et al., 2010; López-Knowles et al., 2010). Cytoplasmic (Dolled-Filhart et al., 2006; Fanelli et al., 2008; Khramtsov et al., 2010; Lin et al., 2000; López-Knowles et al., 2010) and nuclear (Lin et al., 2000; Prasad et al., 2008) β-catenin expression were reported to be associated with worse outcome and reduced OS in BC patients. Also, studies have reported an association between cytoplasmic expression and ER-positive status (Fanelli et al., 2008; Nakopoulou et al., 2000).

In TNBC, β-catenin overexpression is reportedly required for TNBC development by controlling numerous tumor-associated properties, such as migration, stemness, anchorage-independent growth and chemosensitivity (Xu et al., 2015). Furthermore, β-catenin signaling activation has being associated with a poor clinical outcome (Dey et al., 2013b, 2013a; Geyer et al., 2011; Khramtsov et al., 2010; López-Knowles et al., 2010). Among the many β-catenin downstream effectors (Xu et al., 2015), here we highlight Myc (He et al., 1998). Nevertheless, how β-catenin regulates Myc and other downstream targets in BC, as well as its biological significance, particularly in BC, remain poorly explored. In basal-like cell lines, β-catenin was observed to mediate the balance between Myc and *CDKN1A*/p21, a G1-S phase cell cycle regulator, therefore regulating cell death and proliferation in by TCF-independent mechanisms (Xu et al., 2016). Remarkably, upon β-catenin silencing, Myc expression was reduced in non-basal-like BC cell lines but, on contrary, both Myc mRNA and protein expression were significantly upregulated in the basal-like BC cell lines ($p < 0.001$) (Xu et al., 2016). Also, instead of promoting apoptosis, reduced β-catenin expression promoted an increase number of cells in the G1 phase of the cell cycle and decrease of cells in S phase (Xu et al., 2016). Overall, this study suggests that basal-like BC cell lines with downregulated β-catenin are arrested in G1 phase, presenting a slower growth or quiescent state. Cells with these features are usually more resistant to chemotherapeutic agents, such as doxorubicin and cisplatin, promoting future relapse of the disease (Xu et al., 2016).

Thus, here we report that Myc expression is enhanced in TNBCs from African-ancestry patients which, contrary to what is of current knowledge, also present lower expression of β-catenin when comparing with White patients. In fact, to our knowledge, and although all basal-like BC cell lines belonged to White patients (ATCC Catalog, 2013), the study performed by Xu et al. (Xu et al., 2016) is the only aligned with our results. Those authors also showed *CDKN1A*/p21 upregulation upon β-catenin silencing; however, we do not observe significantly differential expression of *CDKN1A* in TNBC-AA patients, comparing with matching White patients. Furthermore, no differential expression of *CTNNB1* and *CDKN1A* was observed when comparing normal-adjacent BC between AA and White patients, nor when comparing TNBC expression profile with normal-adjacent BC tissue expression among AA patients (Supplementary - DGEA file, eTables 13 and 14).

Overall, our findings suggest that Wnt signaling pathway and β-catenin are less expressed in African-ancestry patients, when comparing with their TNBC-White counterparts. Thus, considering Wnt/β-catenin tremendous potential value as a therapeutic target in cancer treatment (Cheng et al., 2019), including BC and TNBC, and the fact that such pre-clinical and clinical studies are developed in countries whose population is mostly of White-ancestry, here we highlight the importance of not generalizing therapeutic findings and applications to other populations.

5. EGFR, MYC, BCL2 AND B-CATENIN AS POTENTIAL ANCESTRY-ASSOCIATED MARKERS IN AFRICAN-ANCESTRY PATIENTS DIAGNOSED WITH TNBC

Importantly, given the exploratory nature of this study, other relevant DEG identified in African-ancestry patients, when comparing with matching White patients, which may be potentially involved in TNBC development, prognosis and treatment of these patients, were certainly overlooked. Additionally, due to the limited amount of biological material available after extensive routine pathological assessment, we were unable to consider further potential markers to be validated through IHC. Remarkably, among the DEG validated through IHC in FFPE tissue samples, protein

expression of EGFR, Myc, Bcl2 and β-catenin are in line with the FC direction of the DEG from which are encoded, even though TCGA and HFF patients are geographically and culturally distinct.

Unfortunately, validation of gene expression through isolation of RNA from FFPE samples from the validation cohort followed by quantitative RT-PCR was not possible due to the poor quality of the RNA isolated. Thus, the validation of the results obtained in the study cohort was exclusively ascertained through IHC. Of note the fact that IHC is actually the most used method of cancer diagnosis. It is a more mature, relatively inexpensive technology, being accessible to the majority of pathology laboratories. Immunohistochemical markers are often used to guide treatment decisions and to classify BC into subtypes that are biologically distinct and behave differently. Thus, such markers can be both prognostic and predictive factors. Although other techniques, such as quantitative RT-PCR, are more sensitive, they lack specificity in distinguishing different cells. In addition, mRNA profile analysis is complex and is more inconvenient for routine clinical use.

In the absence of advanced molecular biological techniques, IHC is used to identify BC histologic subtype, molecular phenotype and to predict patients' response to treatment (Zaha, 2014). The best approach in an immunohistochemical procedure is to combine panels of markers with the examination of standard hematoxylin-eosin slides. In BC cases, only ER, PR and HER2 expression are routinely assessed in the patients' biopsies or surgical pieces, with most anatomic pathology laboratories also performing IHC staining with the proliferation marker Ki-67. However, other biomarkers can be further assessed to predict patients' outcome and response to targeted therapies (Zaha, 2014), including those analyzed through IHC in this project.

Regarding EGFR, to date, most clinical trials of EGFR inhibitors in BC and TNBC patients have failed due to low response rates (Ali and Wendt, 2017; Nabholtz et al., 2016; Nakai et al., 2016). Nonetheless, numerous factors, including lack of proper patient selection, may have contributed to the failure of these trials. Wendt et al. (Wendt et al., 2015) showed that BC cells suffer a progressive loss of EGFR throughout the EMT process that gives rise to metastasis. Authors also suggest that primary breast tumors might be composed by a preexisting EGFR-low subpopulation (Wendt et al., 2015), however, to our best knowledge, such claim remains to be confirmed. Interestingly, this hypothesis would be in line with our results, in which TNBCs diagnosed in African-ancestry patients, which frequently present faster progression and relapse after systemic CT, have lower EGFR mRNA and protein expression when comparing with their White counterparts. Thus, we suggest EGFR as a complementary IHC marker for African-ancestry patients after TNBC diagnosis, even in patients that are not postmenopausal or do not present pathological stage II disease, as the analysis with clinical samples hints that EGFR downregulation might not be exclusive of patients with these characteristics. As previously stated (Kreike et al., 2010), low EGFR downregulation should be considered as a poor prognosis marker in the BC setting and such patients should be carefully followed in the clinical context, as they may be more prone to develop metastasis (Ali and Wendt, 2017; Wendt et al., 2015). Furthermore, such patients should be carefully considered to be included in clinical trials involving small-molecule EGFR inhibitors and monoclonal antibodies, as the low expression of EGFR could promote severe secondary effects upon these treatments (Ali and Wendt, 2017). Thus, we suggest EGFR agonists (Ali and Wendt, 2017; Wendt et al., 2015), such as EGF, as a potential adjuvant therapy for low EGFR expressing African-ancestry patients diagnosed with TNBC.

We observed that both Myc mRNA and protein expression are increased in TNBC cases from African-ancestry patients, when comparing with matching White patients. Given that Myc overexpression is a characteristic of aggressive forms of cancer and patients' poor prognosis, including in TNBC cases, is interesting that we observed that African-ancestry patients have a higher Myc

expression than White patients. As EGFR, Myc protein overexpression may not be specific to patients presenting a menopause status, as suggested by IHC results. Myc augmented expression in African-ancestry might be an interesting potential prognosis marker. However, to our best knowledge, although Myc is an obvious target given that its deregulation is perhaps one of the most important events in cancer development, no approved therapeutic approaches targeting Myc are currently implemented, mostly due to aggressive side effects in normal tissues. Nevertheless, recently, Omomyc, a Myc dominant-negative gene product that was usually used to inhibit *MYC* for research purposes, was showed to be able to penetrate the nuclei of cancer cells in vitro and in vivo in various experimental models of non-small-cell lung carcinoma harboring different oncogenic mutation profiles. Clinical trials with Omomyc peptides to assess its safety and efficacy were expected to be initiated in 2020 (Beaulieu et al., 2019; Soucek et al., 2002).

Also, we observed a lower Bcl2 mRNA and protein expression in TNBCs from premenopausal African-ancestry patients, when comparing with matching White patients. Additionally, gene downregulation was also observed in TNBCs when comparing with normal-adjacent BC tissue from African-ancestry patients, not controlling for menopause status. Thus, low Bcl2 is acquired during TNBC development in African-ancestry patients, and our results suggest that this downregulation is more specifically found in younger patients. Bcl2 downregulation and its role in cancer development and prognosis is still controversial. Previous studies suggest that Bcl2 downregulation in TNBC is associated with faster progression and poor prognosis, such as increased risk of death and relapse (Abdel-Fatah et al., 2013; Eom et al., 2016). However, due to the high mitotic rate, authors observed that Bcl2- patients treated with NACT anthracycline combination-CT were more likely to have a pCR than Bcl2+ patients. However, Bcl2- patients who did not achieve a pCR have a poor prognosis, with a faster relapse (Abdel-Fatah et al., 2013). Thus, as Abdel-Fatah et al. (Abdel-Fatah et al., 2013), we also highlight the importance of adding Bcl2 to a panel of markers used in current clinical practice, since Bcl2 might provide both prognostic and predictive information in TNBC cases from premenopausal African-ancestry patients.

Finally, we observed that Wnt signaling and β-Catenin, which partially regulates Myc expression (Cowling and Cole, 2007), are downregulated in African-ancestry patients in both cohorts, comparing with TNBC-White patients. We suggest that Wnt signaling and β-Catenin might be, in fact, more activated in TNBCs diagnosed in White patients and that, although Myc has an increased expression in African-ancestry patients, Myc is not positively regulating Wnt signaling pathway and vice-versa. According to a myriad of studies, aberrant Wnt/β-Catenin is a trigger for BC (Chung et al., 2004; Dillon et al., 1998; Dolled-Filhart et al., 2006; Fanelli et al., 2008; Karayiannakis et al., 2001; Lin et al., 2000; López-Knowles et al., 2010; Nakopoulou et al., 2000; Prasad et al., 2008; van Schie and van Amerongen, 2020) and TNBC development (Dey et al., 2013b, 2013a; Geyer et al., 2011; Khramtsov et al., 2010; López-Knowles et al., 2010; Xu et al., 2015). Nonetheless, many questions about the role of the reported Wnt/β-Catenin overexpression in BC remain, including the lack of basic understanding of the function of Wnt signaling in normal human breast development and physiology, the effect and precise role of the different β-Catenin effectors, and which underlying molecular and cellular mechanisms are disrupted (van Schie and van Amerongen, 2020). Importantly in the context of this study, we observed downregulation of Wnt/β-Catenin expression in African-ancestry TNBC patients, comparing with matching TNBC-White patients. Since the current status of targeted therapeutics are aiming at interfering with Wnt/β-Catenin overexpression in BC patients (van Schie and van Amerongen, 2020), here we highlight the importance and complexity of selecting the subset of patients that may benefit from treatment.

Although studies with a larger and heterogeneous group of TNBC patients are essential to validation and further implementation of our results in the clinical routine, we suggest EGFR, Myc, Bcl2 and β-Catenin expression potential ancestry-associated markers in African-ancestry patients diagnosed with TNBC, to be appreciated through IHC.

6. THE CONCEPT OF ANCESTRY AND RACE IN A CLINICAL PRESPETIVE

Over the past two decades, with the completion of the human genome in the early 2000s (Craig Venter et al., 2001), a debate arose over the use of a racial classification in medicine and biomedical research. Oddly, the use of race as a biological category increased in the postgenomic age (Chow-White and Green, 2013), although genome pioneers and social scientists alike called for an end to the use of race as a variable in genetic research (Collins, 2004; Foster and Sharp, 2004).

Here we highlight the importance of distinguishing "ancestry" from "race". Ancestry is a process-based concept, a statement about an individual's relationship to other individuals in their genealogical history, being a personal understanding of one's genomic heritage. Race, on the other hand, is a pattern-based concept that has led scientists and laypersons alike to draw conclusions about hierarchical organization of humans, which connect an individual to a larger preconceived geographically circumscribed or socially constructed group (Yudell et al., 2016). Accordingly, phylogenetic and population genetic methods do not support a priori classifications of race, as expected for an interbreeding species like Homo sapiens (Serre and Pääbo, 2004).

In the scientific, biomedical and clinic context, the contemporary practice to identify genetic risk factors of diseases associated to the patient's "race", and to help determine the best course of medical treatments is, in fact, based on the individual's self-identified ancestry (Yudell et al., 2016). Such practice was used in the patients' data gathered in the TCGA consortium and in the medical records from the patients' followed in HFF. Nonetheless, the term "race" and/or "ethnicity" is used in both of those records, instead of ancestry.

Importantly, these ancestry assumptions should not be considered biological guideposts, as commonly defined racial groups are genetically heterogeneous and lack clear-cut genetic boundaries (Serre and Pääbo, 2004; Tishkoff et al., 2009). For instance, cystic fibrosis is underdiagnosed in populations of African-ancestry, because it is thought of as a "White disease" (Stewart and Pepper, 2016), or hemoglobinopathies being misdiagnosed because of the identification of sickle-cell as a "African-ancestry disease" and thalassemia as a "Mediterranean disease" (Amato and Giordano, 2009; Khelil et al., 2010; Solovieff et al., 2011). Thus, although we suggest EGFR, Myc, Bcl2 and β-Catenin as potential ancestry-associated markers, we cannot exclude the possibility of individuals whose genome is largely from African-ancestry having those chromosomic regions, or the chromosomic regions of genes regulating these markers, as being of European origin instead of African (Bryc et al., 2010). The reverse is also possible for patients self-identifying as being from other ancestries.

Remarkably, all four candidate DEG identified in the comparison between TNBC-AA and TNBC-White patients had corresponding protein expression differences in the FFPE tissues from the patients followed in Portugal. While AA population is predominantly of Niger-Kordofanian/western Africa ancestry (Bryc et al., 2010; Tishkoff et al., 2009), most of the African-ancestry patients followed in Portugal emigrated from former Portuguese colonies, including Cape Verde, São Tomé, Guinea-Bissau, Brazil and Mozambique (Supplementary - Patients file, eTable 3). Among the African-ancestry patients from HFF, 8/12 are indeed from western Africa countries, namely Cape Verde, São Tomé and Guinea-

Bissau. These observations suggest that at least patients with Western Africa ancestry, besides having an higher incidence of TNBC (Huo et al., 2009; Jiagge et al., 2016; Newman et al., 2019; Siddharth and Sharma, 2018), also share some TNBC-associated features that are distinct from patients of European-ancestry.

Although our results are promising, given the tremendous human genetic diversity found in the human population (Serre and Pääbo, 2004) and, particularly, in the African continent (Tishkoff et al., 2009), whenever possible, patients' ancestry inference should be confirmed with resource to human ancestry markers, instead of self-described ancestry. This is particularly relevant in populations composed by people with diverse biogeographical ancestries, such as in Portugal or in the USA. Remarkably, to our best knowledge, in all studies cited in this thesis, patients' ancestry was self-described. One of the major reasons for so few studies to investigate ancestry-associated disease markers resourcing to SNP panels for ancestry inference is the fact that such panels still have limitations in their generality and efficiency, such as the fact that some only have a very limited amount of populations (Kidd et al., 2014), or are based in a limited number of SNPs or chromosomes (Toma et al., 2018). Nonetheless, SNP panels for ancestry inference are constantly developing (Lan et al., 2020; Pakstis et al., 2019) and soon will have important implications in the clinical context.

Thus, currently, patients' self-described ancestry-based predictions in clinical settings are helpful in predicting patients' prognosis and the most appropriate treatment, being an easy and inexpensive guiding method. However, whenever possible, self-described ancestry should not substitute the specific etiology of the individual patient following a true personalized therapeutic approach.

7. CONCLUDING REMARKS

Biological differences between TNBC incidence, development and prognosis in African and European-ancestry women are influenced by differences in the transcriptomic profile and regulation of pathways associated to carcinogenesis. Identification of altered mRNA and protein expression profile are expected to contribute to patient's prognosis, and interventions to improve cancer risk assessment and optimal use of more effective targeted therapies have the potential to close the widening mortality gap between these populations. Our study is one of the few to systematically examine ancestry differences in mRNA and protein expression specifically in patients with TNBC, stratifying patients according to their menopause status and/or pathological stage.

One of the major strengths of this PhD thesis regards the stringency of the multiple testing procedures taken. This stringency allowed for more robust results regarding the ancestry-associated transcriptomic differences in TNBC in the different contrasts of the computational analysis, either in DGEA, GSEA and IPA's core analysis. Additionally, to our best knowledge, this study is pioneer in assessing those differences according to menopause status and pathological stage at TNBC diagnosis, variables that affect further disease development and patient's prognosis and treatment options. Moreover, we also validated four candidate genes identified in the computational analysis through IHC, the most common method used for cancer diagnosis and prognosis in the clinic, in a group of patients composed by sociocultural and geographically distinct women.

Nonetheless, despite the fact that we identified potential ancestry-associated markers for African-ancestry patients diagnosed with TNBC, we faced some limitations. The major limitation was indeed the number of patients. The group of TCGA patients is comprised by African-ancestry and White

women diagnosed with TNBC with available RNA-seq and clinicopathological data. Other databases did not meet our requirements of having ancestry information and African-ancestry patients, RNA-seq data, and other relevant clinicopathological data, including menopause status or age, pathological stage of the disease and patients' follow-up and survival status. Thus, for the computational analysis, only TCGA data was included, from which we also removed perimenopausal and pathological stage IV disease patients, given the small number of patients presenting those characteristics. Furthermore, by stratifying patients according to menopause status and pathological stage, the number of patients included in each group would be further reduced. Also, concerning the group of patients from HFF, with available FFPE TNBC tissue, we gathered an unexpectedly small number of samples. Initially, we identified many more African-ancestry patients having a medical record stating them as a TNBC patient, from 2012. Nonetheless, most of those patients do not live in Portugal, being followed in our collaboration hospital only after diagnosis and surgery in their country of residence, mostly to proceed to further adjuvant treatments. Thus, FFPE samples from such patients was not available. Finally, we also faced a surprisingly shortage in available molecular data from reference normal breast tissue from African-ancestry women, either from TCGA or from other databases. We also found this limitation in our collaboration hospital, where tissue from health donors was not available and normal-adjacent BC FFPE samples were in poor conditions or have a very reduced number of epithelial cells. Also, some of the FFPE samples were biopsies, which do not have normal breast tissue surrounding the tumor. Thus, for these reasons, we could not validate EGFR, Myc, Bcl2 and β-catenin protein expression in a normal breast tissue from patients gathered in Portugal. Other major limitation of this work was the poor quality of the RNA isolated from the FFPE samples, which impaired the validation of gene expression through quantitative RT-PCR, prior the IHC essays. Other limitations included an unequal and/or short follow-up time of patients, as well as an unclear definition of the cause of death of TCGA patients, which may impair the real effect of the ancestry in the probability of survival after TNBC diagnosis. Furthermore, socioeconomic and environmental factors were not available nor in TCGA nor in most of the medical records in the hospital with which we collaborate, hence, their contribution to ancestry disparity in TNBC development and prognosis could not be questioned.

The major findings of this study include the identification of hundreds of DEG between TNBC patients of African and European-ancestry, according to patients' menopause status and stage of the disease at diagnosis, that are TNBC related and not ancestry alone-related (the lists of DEG, according to patients' characteristics, can be consulted in the Supplementary - DGEA file). Interestingly, although young African-ancestry patients are more frequently diagnosed with TNBC than their White counterparts, the expression profile of the two groups of patients is more distant according to the stage of the disease than with the menopause status. We also highlight the fact that among the DEG validated through IHC in FFPE tissue samples, remarkably, protein expression of EGFR, Myc, Bcl2 and β-catenin are in line with the FC direction of the DEG from which are encoded, even though the TCGA cases and clinical samples are composed by geographically and culturally distinct groups of patients. This consistency between the computational analysis and IHC results may be due to most patients from both groups of African-ancestry sharing western Africa ancestry. Regarding the selected candidate ancestry-associated markers, to summarize, we highlight 1) the under-expression of EGFR in African-ancestry patients, comparing with TNBC-White patients; 2) the increased expression of Myc; 3) Bcl2 downregulation involved in TNBC development in premenopausal African-ancestry patients; and 4) β-Catenin under-expression in African-ancestry patients.

These results highlight the critical need for researchers investigating the underlying pathogenesis and exploiting therapeutically the distinct biology of TNBC, particularly in African-

ancestry patients. Ultimately, this PhD thesis underscores the potential for individualized tumor molecular analysis to mitigate the ancestry divide in TNBC outcome. While this work was primarily focused on finding differential expression profile and potential ancestry-associated markers in African-ancestry patients diagnosed with TNBC, as well as potentially altered pathways and cellular mechanisms considering patient's menopause condition and pathological stage at diagnosis, additional research is urgently needed to evaluate these findings, particularly in the pre-clinical setting, including functional assays using human cell lines and animal models, to provide a deeper understanding of the complex association between tumor transcriptomics and outcome among patients with different ancestries.

Overall, future studies with larger and diverse TNBC cohorts are urgently needed to better elucidate ancestry differences in TNBC transcriptomics and proteomics, adjusting for patients' characteristics, such as age and pathological stage at diagnosis, germline, therapeutic and socioeconomic influences.

Chapter VI

Bibliography

Aasen, T., Leithe, E., et al, 2019. Connexins in cancer: bridging the gap to the clinic. Oncogene 38, 4429–4451.

Aasen, T., Mesnil, M., Naus, C.C., Lampe, P.D., Laird, D.W., 2016. Gap junctions and cancer: Communicating for 50 years. Nat. Rev. Cancer 16, 775–788.

Abd El-Rehim, D.M., Ball, G., Finder, S.E., Rakha, E., Paish, C., Robertson, J.F.R., Macmillan, D., Blamey, R.W., Ellis, I.O., 2005. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int. J. Cancer 116, 340–350.

Abdel-Fatah, T.M.A., Perry, C., Dickinson, P., Ball, G., Moseley, P., Madhusudan, S., Ellis, I.O., Chan, S.Y.T., 2013. Bcl2 is an independent prognostic marker of triple negative breast cancer (TNBC) and predicts response to anthracycline combination (ATC) chemotherapy (CT) in adjuvant and neoadjuvant settings. Ann. Oncol. 24, 2801–2807.

Abe, O., Abe, R., Enomoto, K., Kikuchi, K., Koyama, H., Nomura, Y., Sakai, K., Sugimachi, K., Tominaga, T., Uchino, J., Yoshida, M., van de Velde, A.O., van Dongen, J.A., Vermorken, J.B., Giokas, G., Lissaios, B., Harvey, V.J., Holdaway, T.M., Kay, R.G., Mason, B.H., Coates, A., Forbes, J.F., Focan, C., Lobelle, J.P., Peek, U., Oates, G.D., Powell, J., Durand, M., Mauriac, L., Bartholomeus, S., Piccart, M.J., Gelman, R.S., Harris, J.R., Shapiro, C.L., Hancock, A.K., Masood, M.B., Parker, D., Price, J.J., Jackson, S., Ragaz, J., Delozier, T., Macé-Lesec'h, J., Haybittle, J.L., Cirrincione, C., Henderson, I.C., Korzun, A., Weiss, R.B., Wood, W.C., Baum, M., Houghton, J., Riley, D., Dent, D.M., Gudgeon, C.A., Hacking, A., Horgan, K., Hughes, L., Stewart, H.J., Gordon, N.H., Davis, H.L., Romestaing, P., Lehingue, Y., Owen, J.R., Meier, P., Howell, A., Ribeiro, G.G., Swindell, R., Albano, J., de Oliveira, C.F., Gervásio, H., Gordilho, J., Carstensen, B., Palshof, T., Johansen, H., Korzeniowski, S., Skolyszewski, J., Portnoj, S.M., Andersen, K.W., Axelsson, C.K., Blichert-Toft, M., Mouridsen, H.T., Overgaard, M., Rose, C., Corcoran, N., Trampisch, H.J., Comis, R.L., Davidson, N.E., Gray, R., Robert, N., Tormey, D.C., Wood, W., Rossbach, J., Bijnens, L., van de Velde, C., Cunningham, M.P., Bastert, G., Rauschecker, H., Sauer, R., Sauerbrei, W., Schauer, A., Schumacher, M., de Schryver, A., Belfiglio, M., Mari, E., Nicolucci, A., Scorpiglione, N., Yosef, H.M.A., McArdle, C.S., Smith, D.C., Lara, P.C., Boccardo, F., Rubagotti, A., Erazo, A., Medina, J.Y., Izuo, M., Morishita, Y., Bentley, A., Doran, Z., Fentiman, I.S., Hayward, J.L., Rubens, R.D., Kaufmann, M., Jonat, W., von Fournier, D., Kaufmann, M., Fountzilas, G., Klefstrom, P., Blomqvist, C., Cuzick, J., Margreiter, R., Castiglione, M., Cavalli, F., Collins, J., Forbes, J., Gelber, R.D., Goldhirsch, A., Lindtner, J., Price, K.N., Rudenstam, C.M., Senn, H.J., Bliss, J.M., Coombes, R.C., Marty, M., Borovik, R., Brufman, G., Hayat, H., Robinson, E., Wigler, N., Pannuti, F., Takashima, S., Yasutomi, T., Sonoo, H., Yamashita, J., Ogawa, M., Nomura, Y., Hupperets, P.S.G.J., Bonte, J., Tengrup, I., Tennvall-Nittby, L., Martin, P., Romain, S., Ahmann, D., Schaid, D.J., Buzdar, A.U., Smith, T., Hakes, T., Norton, L., Wittes, R., de la Huerta, R., Sainz, M.G., Bonadonna, G., del Vecchio, M., Valagussa, P., Veronesi, U., Dubois, J.B., Bianco, A.R., Lippman, M.E., Pierce, L.J., Simon, R., Steinberg, S.M., Myles, J.D., Pater, J.L., Pritchard, K.I., Anderson, S., Brown, A., Bryant, J., Costantino, J., Dignam, J., Fisher, B., Redmond, C., Wieand, S., Wolmark, N., Baum, M., Jackson, I.M., Palmer, M.K., Ingle, J.N., Suman, V.J., Bengtsson, N.O., Larsson, L.G., Lythgoe, J.P., Swindell, R., Kissin, M., Hannisdal, E., Varhaug, J.E., Nissen-Meyer, R., Blamey, R.W., Mitchell, A.K., Robertson, J.F.R., Nakamura, Y., Mathé, G., Misset, J.L., Abu-Zahra, H.T., Clarke, E.A., McLaughlin, J.R., Clark, R.M., Levine, M., Morimoto, K., Sawa, K., Takatsuka, Y., Gundersen, S., Hauer-Jensen, M., Host, H., Crossley, E., Harris, A., Baker, C., Beighton, A., Clarke, M., Collins, R., Davies, C., Elphinstone, T., Evans, V., Godwin, J., Gray, R., Greaves, E., Harwood, C., Headon, A., Hicks, C., Jackson, D., James, S., Lau, E., McGale, P., Mead, G., Monaghan, H., Mozley, S., Naughten, A., Peto, R., Tooth, A., Wheatley, K., Rambert, P., Asselain, B., Salmon, R.J., Vilcoq, J.R., Arriagada, R., Hill, C., Laplanche, A., Lê, M.G., Spielmann, M., Cocconi, G., di Blasio, B., Catalano, R., Creech, R.H., Brockschmidt, J., Cooper, M.R., Andrysek, O., Barkmanova, J., Falkson, C.I., Abraham, M., Klijn, J.G.M., Treurniet-Donker, A.D., van Putten, W.L.J., Easton, D., Powles, T.J., Gazet, J.C.,

Semiglazov, V., Businico, A., Sardinia, S., Deshpande, N., di Martino, L., Douglas, P., Hacking, A., Host, H., Lindtner, A., Notter, G., Bryant, A.J.S., Ewing, G.H., Krushen-Kosloski, J.L., Nissen-Meyer, R., Forrest, A.P.M., Jack, W., McDonald, C., Stewart, H.J., Möller, T.R., Rydén, S., Carstensen, J., Hatschek, T., Söderberg, M., Carpenter, J.T., Albain, K., Crowley, J., Green, S., Martino, S., Osborne, C.K., Ravdin, P.M., Rutqvist, L.E., Wallgren, A., Holm, L.E., Castiglione, M., Goldhirsch, A., Senn, H.J., Thürlimann, B., Brenner, H., Hercbergs, A., Yoshimoto, M., DeBoer, G., Paterson, A.H.G., Meakin, J.W., Panzarella, T., Naja, A., Bahi, J., Reid, M., Spittle, M., Senanayake, F., Love, R.R., de Mets, D.L., Bergh, J., Holmberg, L., Sevelda, P., Zielinsky, C.C., Gnant, M., Jakesz, R., Buchanan, R.B., Royle, G.T., Dunn, J.A., Gillespie, W.M., Kelly, K., Morrison, J.M., Litton, A., Chlebowski, R.T., Bezwoda, W.R., Caffier, H., 1998. Tamoxifen for early breast cancer: An overview of the randomised trials. Lancet 351, 1451–1467.

Abramson, V.G., Lehmann, B.D., Ballinger, T.J., Pietenpol, J.A., 2015. Subtyping of Triple-Negative Breast Cancer: Implications for Therapy.

Adams, J.M., Cory, S., 2007. The Bcl-2 apoptotic switch in cancer development and therapy. Oncogene 26, 1324–1337.

Adams, S., Gray, R.J., Demaria, S., Goldstein, L., Perez, E.A., Shulman, L.N., Martino, S., Wang, M., Jones, V.E., Saphner, T.J., Wolff, A.C., Wood, W.C., Davidson, N.E., Sledge, G.W., Sparano, J.A., Badve, S.S., 2014. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. J. Clin. Oncol. 32, 2959–2966.

Ademuyiwa, F.O., Tao, Y., Luo, J., Weilbaecher, K., Ma, C.X., 2017. Differences in the mutational landscape of triple-negative breast cancer in African Americans and Caucasians. Breast Cancer Res. Treat. 161, 491–499.

Ades, F., Zardavas, D., Bozovic-spasojevic, I., Pugliano, L., Fumagalli, D., 2014. Luminal B Breast Cancer : Molecular Characterization, Clinical Management, and Future Perspectives 1–12.

Aebi, S., Davidson, T., Gruber, G., Cardoso, F., 2011. clinical practice guidelines Primary breast cancer : ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up clinical practice guidelines 22, 12–24.

Al-Ejeh, F., Shi, W., Miranda, M., Simpson, P.T., Vargas, A.C., Song, S., Wiegmans, A.P., Swarbrick, A., Welm, A.L., Brown, M.P., Chenevix-Trench, G., Lakhani, S.R., Khanna, K.K., 2013. Treatment of triple-negative breast cancer using anti-EGFR-directed radioimmunotherapy combined with radiosensitizing chemotherapy and PARP inhibitor. J. Nucl. Med. 54, 913–921.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., Clarke, M.F., 2003. Prospective identification of tumorigenic prostate cancer stem cells. PNAS 100, 3983–3988.

Albain, K., Anderson, S., Arriagada, R., Barlow, W., Bergh, J., Bliss, J., Buyse, M., Cameron, D., Carrasco, E., Clarke, M., Correa, C., Coates, A., Collins, R., Costantino, J., Cutter, D., Cuzick, J., Darby, S., Davidson, N., Davies, C., Davies, K., Delmestri, A., Di Leo, A., Dowsett, M., Elphinstone, P., Evans, V., Ewertz, M., Gelber, R., Gettins, L., Geyer, C., Goldhirsch, A., Godwin, J., Gray, R., Gregory, C., Hayes, D., Hill, C., Ingle, J., Jakesz, R., James, S., Kaufmann, M., Kerr, A., MacKinnon, E., McGale, P., McHugh, T., Norton, L., Ohashi, Y., Paik, S., Pan, H.C., Perez, E., Peto, R., Piccart, M., Pierce, L., Pritchard, K., Pruneri, G., Raina, V., Ravdin, P., Robertson, J., Rutgers, E., Shao, Y.F., Swain, S., Taylor, C., Valagussa, P., Viale, G., Whelan, T., Winer, E., Wang, Y., Wood, W., 2012. Comparisons between different polychemotherapy regimens for early breast cancer: Meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. Lancet 379, 432–444.

Albain, K.S., Unger, J.M., Crowley, J.J., Coltman, C.A., Hershman, D.L., 2009. Racial disparities in cancer survival among randomized clinical trials patients of the southwest oncology group. J. Natl. Cancer Inst. 101, 984–992.

Aleskandarany, M.A., Rakha, E.A., Ahmed, M.A.H., Powe, D.G., Paish, E.C., MacMillan, R.D., Ellis, I.O., Green, A.R., 2010. PIK3CA expression in invasive breast cancer: A biomarker of poor prognosis. Breast Cancer Res. Treat. 122, 45–53.

Aleskandarany, M.A., Soria, D., Green, A.R., Nolan, C., Diez-Rodriguez, M., Ellis, I.O., Rakha, E.A., 2015. Markers of progression in early-stage invasive breast cancer: a predictive immunohistochemical panel algorithm for distant recurrence risk stratification. Breast Cancer Res. Treat. 151, 325–333.

Ali, R., Wendt, M.K., 2017. The paradoxical functions of EGFR during breast cancer progression. Signal Transduct. Target. Ther. 2, 1–7.

Althuis, M.D., Fergenbaum, J.H., Brinton, L.A., Madigan, M.P., Sherman, M.E., 2004. Etiology of Hormone Receptor – Defined Breast Cancer : A Systematic Review of the Literature 13, 1558–1569.

Amato, A., Giordano, P.C., 2009. Screening and Genetic Diagnosis of Hemoglobinopathies in Southern and Northern Europe: Two Examples. Mediterr. J. Hematol. Infect. Dis.

Amirikia, K.C., Mills, P., Bush, J., Newman, L.A., 2011. Higher population-based incidence rates of triple-negative breast cancer among young African-American women. Cancer 117, 2747–2753.

Anampa, J., Chen, A., Wright, J., Patel, M., Pellegrino, C., Fehn, K., Sparano, J.A., Andreopoulou, E., 2018. Phase I Trial of Veliparib, a Poly ADP Ribose Polymerase Inhibitor, Plus Metronomic Cyclophosphamide in Metastatic HER2-negative Breast Cancer. Clin. Breast Cancer 18, e135–e142.

Anders, C.K., Fan, C., Parker, J.S., Carey, L.A., Blackwell, K.L., Klauber-DeMore, N., Perou, C.M., 2011. Breast carcinomas arising at a young age: Unique biology or a surrogate for aggressive intrinsic subtypes? J. Clin. Oncol. 29, 2010–2012.

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

Andre, F., Job, B., Dessen, P., Tordai, A., Michiels, S., Liedtke, C., Richon, C., Yan, K., Wang, B., Vassal, G., Delaloge, S., Hortobagyi, G.N., Symmans, W.F., Lazar, V., Pusztai, L., 2009. Human Cancer Biology Molecular Characterization of Breast Cancer with High-Resolution Oligonucleotide Comparative Genomic Hybridization Array 15, 441–452.

André, F., Zielinski, C.C., 2012. Optimal strategies for the treatment of metastatic triple-negative breast cancer with currently approved agents. Ann. Oncol. 23, vi46–vi51. 5

Arnold, K.M., Pohlig, R.T., Sims-Mourtada, J., 2017. Co-activation of hedgehog and Wnt signaling pathways is associated with poor outcomes in triple negative breast cancer. Oncol. Lett. 14, 5285–5292.

ATCC Catalog, 2013. Breast Cancer and Normal Cell Lines.

Audebert, M., Salles, B., Calsou, P., 2004. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. J. Biol. Chem. 279, 55117–55126.

Auguie, B., 2015. gridExtra: Miscellaneous Functions for “Grid” Graphics. R Packag. version 2.0.0.

Ayanian, J.Z., Kohler, B.A., Abe, T., Epstein, A.M., 1993. The relation between health insurance coverage and clinical outcomes among women with breast cancer. N Engl J Med 329, 326–331.

Balko, J.M., Giltnane, J.M., Wang, K., Schwarz, L.J., Young, C.D., Cook, R.S., Owens, P., Sanders, M.E., Kuba, M.G., Sánchez, V., Kurupi, R., Moore, P.D., Pinto, J.A., Doimi, F.D., Gómez, H., Horiuchi, D., Goga, A., Lehmann, B.D., Bauer, J.A., Pietenpol, J.A., Ross, J.S., Palmer, G.A., Yelensky, R., Cronin, M., Miller, V.A., Stephens, P.J., Arteaga, C.L., 2014. Molecular profiling of the residual disease of

triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets. Cancer Discov. 4, 232–245.

Bandera, E. V., Chandran, U., Hong, C.C., Troester, M.A., Bethea, T.N., Adams-Campbell, L.L., Haiman, C.A., Park, S.Y., Olshan, A.F., Ambrosone, C.B., Palmer, J.R., Rosenberg, L., 2015. Obesity, body fat distribution, and risk of breast cancer subtypes in African American women participating in the AMBER Consortium. Breast Cancer Res. Treat. 150, 655–666.

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K.K., Carter, S.L., Frederick, A.M., Lawrence, M.S., Sivachenko, A.Y., Sougnez, C., Zou, L., Cortes, M.L., Fernandez-Lopez, J.C., Peng, S., Ardlie, K.G., Auclair, D., Bautista-Piña, V., Duke, F., Francis, J., Jung, J., Maffuz-Aziz, A., Onofrio, R.C., Parkin, M., Pho, N.H., Quintanar-Jurado, V., Ramos, A.H., Rebollar-Vega, R., Rodriguez-Cuevas, S., Romero-Cordoba, S.L., Schumacher, S.E., Stransky, N., Thompson, K.M., Uribe-Figueroa, L., Baselga, J., Beroukhim, R., Polyak, K., Sgroi, D.C., Richardson, A.L., Jimenez-Sanchez, G., Lander, E.S., Gabriel, S.B., Garraway, L.A., Golub, T.R., Melendez-Zajgla, J., Toker, A., Getz, G., Hidalgo-Miranda, A., Meyerson, M., 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 486, 405–409.

Bànkfalvi, À., Terpe, H.J., Breukelmann, D., Bier, B., Rempe, D., Pschadka, G., Krech, R., Lellè, R.J., Boecker, W., 1999. Immunophenotypic and prognostic analysis of E-cadherin and β-catenin expression during breast carcinogenesis and tumour progression: A comparative study with CD44. Histopathology 34, 25–34.

Bareche, Y., Venet, D., Ignatiadis, M., Aftimos, P., Piccart, M., Rothe, F., Sotiriou, C., 2018. Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. Ann. Oncol. 29, 895–902.

Barker, N., Clevers, H., 2006. Mining the Wnt pathway for cancer therapeutics. Nat. Rev. Drug Discov. 5, 997–1014.

Barlow, W.E., White, E., Ballard-Barbash, R., Vacek, P.M., Titus-Ernstoff, L., Carney, P.A., Tice, J.A., Buist, D.S.M., Geller, B.M., Rosenberg, R., Yankaskas, B.C., Kerlikowske, K., 2006. Prospective breast cancer risk prediction model for women undergoing screening mammography. J. Natl. Cancer Inst. 98, 1204–1214.

Bartucci, M., Dattilo, R., Moriconi, C., Pagliuca, A., Mottolese, M., Federici, G., Di Benedetto, A., Todaro, M., Stassi, G., Sperati, F., Amabile, M.I., Pilozzi, E., Patrizii, M., Biffoni, M., Maugeri-Saccà, M., Piccolo, S., De Maria, R., 2015. TAZ is required for metastatic activity and chemoresistance of breast cancer stem cells. Oncogene 34, 681–690.

Baselga, J., Gómez, P., Greil, R., Braga, S., Climent, M.A., Wardley, A.M., Kaufman, B., Stemmer, S.M., Pego, A., Chan, A., Goeminne, J.C., Graas, M.P., Kennedy, M.J., Ciruelos Gil, E.M., Schneeweiss, A., Zubel, A., Groos, J., Melezínkova, H., Awada, A., 2013. Randomized phase II study of the anti-epidermal growth factor receptor monoclonal antibody cetuximab with cisplatin versus cisplatin alone in patients with metastatic triple-negative breast cancer. J. Clin. Oncol. 31, 2586–2592.

Baselga, J., Swain, S.M., 2009. Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. Nat. Publ. Gr. 9, 463–475.

Basho, R.K., Gilcrease, M., Murthy, R.K., Helgason, T., Karp, D.D., Meric-Bernstam, F., Hess, K.R., Herbrich, S.M., Valero, V., Albarracin, C., Litton, J.K., Chavez-MacGregor, M., Ibrahim, N.K., Murray, J.L., Koenig, K.B., Hong, D., Subbiah, V., Kurzrock, R., Janku, F., Moulder, S.L., 2017. Targeting the PI3K/AKT/mTOR pathway for the treatment of mesenchymal triple-negative breast cancer: Evidence from a phase 1 trial of mTOR inhibition in combination with liposomal doxorubicin and bevacizumab. JAMA Oncol. 3, 509–515.

Bates, G.J., Fox, S.B., Han, C., Leek, R.D., Garcia, J.F., Harris, A.L., Banham, A.H., 2006. Quantification of regulatory T cells enables the identification of high-risk breast cancer patients and those at risk

of late relapse. J. Clin. Oncol. 24, 5373–5380.

Bauer, K.R., Brown, M., Cress, R.D., Parise, C.A., Caggiano, V., 2007. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: A population-based study from the California Cancer Registry. Cancer 109, 1721–1728.

Beaulieu, M.E., Jauset, T., Massó-Vallés, D., Martínez-Martín, S., Rahl, P., Maltais, L., Zacarias-Fluck, M.F., Casacuberta-Serra, S., Del Pozo, E.S., Fiore, C., Foradada, L., Cano, V.C., Sánchez-Hervás, M., Guenther, M., Sanz, E.R., Oteo, M., Tremblay, C., Martín, G., Letourneau, D., Montagne, M., Alonso, M.Á.M., Whitfield, J.R., Lavigne, P., Soucek, L., 2019. Intrinsic cell-penetrating activity propels omomyc from proof of concept to viable anti-myc therapy. Sci. Transl. Med. 11, 1–14.

Beckers, R.K., Selinger, C.I., Vilain, R., Madore, J., Wilmott, J.S., Harvey, K., Holliday, A., Cooper, C.L., Robbins, E., Gillett, D., Kennedy, C.W., Gluch, L., Carmalt, H., Mak, C., Warrier, S., Gee, H.E., Chan, C., McLean, A., Walker, E., McNeil, C.M., Beith, J.M., Swarbrick, A., Scolyer, R.A., O'Toole, S.A., 2016. Programmed death ligand 1 expression in triple-negative breast cancer is associated with tumour-infiltrating lymphocytes and improved outcome. Histopathology 69, 25–34.

Beckmann, A., Hainz, N., Tschernig, T., Meier, C., 2019. Facets of communication: Gap junction ultrastructure and function in cancer stem cells and tumor cells. Cancers (Basel). 11.

Beh, C.W., Seow, W.Y., Wang, Y., Zhang, Y., Ong, Z.Y., Ee, P.L.R., Yang, Y.Y., 2009. Efficient delivery of bcl-2-targeted siRNA using cationic polymer nanoparticles: Downregulating mRNA expression level and sensitizing cancer cells to anticancer drug. Biomacromolecules 10, 41–48.

Berns, K., Horlings, H.M., Hennessy, B.T., Madiredjo, M., Hijmans, E.M., Beelen, K., Linn, S.C., Gonzalez-Angulo, A.M., Stemke-Hale, K., Hauptmann, M., Beijersbergen, R.L., Mills, G.B., van de Vijver, M.J., Bernards, R., 2007. A Functional Genetic Approach Identifies the PI3K Pathway as a Major Determinant of Trastuzumab Resistance in Breast Cancer. Cancer Cell 12, 395–402.

Bhat, A.A., Uppada, S., Achkar, I.W., Hashem, S., Yadav, S.K., Shanmugakonar, M., Al-Naemi, H.A., Haris, M., Uddin, S., 2019. Tight junction proteins and signaling pathways in cancer and inflammation: A functional crosstalk. Front. Physiol. 10, 1–19.

Bhattacharyya, A., Ear, U.S., Koller, B.H., Weichselbaum, R.R., Bishop, D.K., 2000. The Breast Cancer susceptibility gene BRCA1 is required for subnuclear assembly of Rad51 and survival following treatment with the DNA cross-linking agent cisplatin. J. Biol. Chem. 275, 23899–23903.

Bhatti, S., Heldstab, J., Lehn, C., Tawfik, O., Ash, R.M., Hout, D.R., Seitz, R.S., Bailey, D.B., O'Dea, A.P., Jensen, R.A., Fan, F., Khan, Q.J., Godwin, A.K., Sharma, P., 2017. Clinical Activity of Pembrolizumab in a Patient With Metastatic Triple-Negative Breast Cancer Without Tumor Programmed Death-Ligand 1 Expression: A Case Report and Correlative Biomarker Analysis. JCO Precis. Oncol. 1–6.

Bhola, N.E., Balko, J.M., Dugger, T.C., Kuba, M.G., Sánchez, V., Sanders, M., Stanford, J., Cook, R.S., Arteaga, C.L., 2013. TGF-β inhibition enhances chemotherapy action against triple-negative breast cancer. J. Clin. Invstigation 123, 1348–1358.

Bianchini, G., Balko, J.M., Mayer, I.A., Sanders, M.E., Gianni, L., 2016. Triple-negative breast cancer: Challenges and opportunities of a heterogeneous disease. Nat. Rev. Clin. Oncol. 13, 674–690.

Bianchini, G., Qi, Y., Alvarez, R.H., Iwamoto, T., Coutant, C., Ibrahim, N.K., Valero, V., Cristofanilli, M., Green, M.C., Radvanyi, L., Hatzis, C., Hortobagyi, G.N., Andre, F., Gianni, L., Fraser Symmans, W., Pusztai, L., 2010. Molecular anatomy of breast cancer stroma and its prognostic value in estrogen receptor-positive and -negative cancers. J. Clin. Oncol. 28, 4316–4323.

Birnbaum, J.K., Duggan, C., Anderson, B.O., Etzioni, R., 2018. Early detection and treatment strategies for breast cancer in low- income and upper middle-income countries: a modelling study. Lancet Glob Heal. 6, e885–e893.

Black, S.A., 2002. Diabetes, diversity, and disparity: What do we do with the evidence? Am. J. Public Health 92, 543–548.

Bogenrieder, T., Herlyn, M., 2003. Axis of evil: Molecular mechanisms of cancer metastasis. Oncogene 22, 6524–6536.

Bouchalova, K., Svoboda, M., Kharaishvili, G., Vrbkova, J., Bouchal, J., Trojanec, R., Koudelakova, V., Radova, L., Cwiertka, K., Hajduch, M., Kolar, Z., 2015. BCL2 is an independent predictor of outcome in basal-like triple-negative breast cancers treated with adjuvant anthracycline-based chemotherapy. Tumor Biol. 36, 4243–4252.

Boyd, N.F., Li, Q., Melnichouk, O., Huszti, E., Martin, L.J., Gunasekara, A., Mawdsley, G., Yaffe, M.J., Minkin, S., 2014. Evidence that breast tissue stiffness is associated with risk of breast cancer. PLoS One 9, 1–8.

Boyle, P, 2012. research article Triple-negative breast cancer : epidemiological considerations and recommendations research article 23, 8–13.

Boyle, P., 2012. Triple-negative breast cancer: Epidemiological considerations and recommendations. Ann. Oncol. 23, vi7–vi12.

Boyle, P., Boniol, M., Koechlin, A., Robertson, C., Valentini, F., Coppens, K., Fairley, L.L., Boniol, M., Zheng, T., Zhang, Y., Pasterk, M., Smans, M., Curado, M.P., Mullie, P., Gandini, S., Bota, M., Bolli, G.B., Rosenstock, J., Autier, P., 2012. Diabetes and breast cancer risk: A meta-analysis. Br. J. Cancer 107, 1608–1617.

Bradley, C.J., 2002. Race, Socioeconomic Status, and Breast Cancer Treatment and Survival. CancerSpectrum Knowl. Environ. 94, 490–496.

Braicu, C., Buse, M., Busuioc, C., Drula, R., Gulei, D., Raduly, L., Rusu, A., Irimie, A., Atanasov, A.G., 2019. A Comprehensive Review on MAPK: A Promising Therapeutic Target in Cancer. Cancers (Basel). 11.

Brand, T.M., Iida, M., Luthar, N., Starr, M.M., Huppert, E.J., Wheeler, D.L., 2013. Nuclear EGFR as a molecular target in cancer. Radiother. Oncol. 108, 370–377.

Brandão, A.S., Bensimon-Brito, A., Lourenço, R., Borbinha, J., Soares, A.R., Mateus, R., Jacinto, A., 2019. Yap induces osteoblast differentiation by modulating Bmp signalling during zebrafish caudal fin regeneration. J. Cell Sci. 132.

Bray, F., Jemal, A., Grey, N., Ferlay, J., Forman, D., 2012. Global cancer transitions according to the Human Development Index (2008-2030): A population-based study. Lancet Oncol. 13, 790–801.

Brenton, J.D., Carey, L.A., Ahmed, A., Caldas, C., 2005. Molecular classification and molecular forecasting of breast cancer: Ready for clinical application? J. Clin. Oncol. 23, 7350–7360.

Brewster, A.M., Chavez-MacGregor, M., Brown, P., 2014. Epidemiology, biology, and treatment of triple-negative breast cancer in women of African ancestry. Lancet Oncol. 15, e625–e634.

Brichese, L., Barboule, N., Heliez, C., Valette, A., 2002. Bcl-2 phosphorylation and proteasome-dependent degradation induced by paclitaxel treatment: Consequences on sensitivity of isolated mitochondria to Bid. Exp. Cell Res. 278, 101–111.

Britschgi, A., Duss, S., Kim, S., Couto, J.P., Brinkhaus, H., Koren, S., De Silva, D., Mertz, K.D., Kaup, D., Varga, Z., Voshol, H., Vissieres, A., Leroy, C., Roloff, T., Stadler, M.B., Scheel, C.H., Miraglia, L.J., Orth, A.P., Bonamy, G.M.C., Reddy, V.A., Bentires-Alj, M., 2017. The Hippo kinases LATS1 and 2 control human breast cell fate via crosstalk with ERα. Nature 541, 541–545.

Bruna, A., Greenwood, W., Le Quesne, J., Teschendorff, A., Miranda-Saavedra, D., Rueda, O.M., Sandoval, J.L., Vidakovic, A.T., Saadi, A., Pharoah, P., Stingl, J., Caldas, C., 2012. TGFβ induces the

formation of tumour-initiating cells in claudin low breast cancer. Nat. Commun. 3.

Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., Helleday, T., 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase.[erratum appears in Nature. 2007 May 17;447(7142):346]. Nature 434, 913–917.

Bryc, K., Auton, A., Nelson, M.R., Oksenberg, J.R., Hauser, S.L., Williams, S., Froment, A., Bodo, J.M., Wambebe, C., Tishkoff, S.A., Bustamante, C.D., 2010. Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc. Natl. Acad. Sci. U. S. A. 107, 786–791.

Burga, L.N., Hu, H., Juvekar, A., Tung, N.M., Troyan, S.L., Hofstatter, E.W., Wulf, G.M., 2011. Loss of BRCA1 leads to an increase in epidermal growth factor receptor expression in mammary epithelial cells, and epidermal growth factor receptor inhibition prevents estrogen receptor-negative cancers in BRCA1-mutant mice. Breast Cancer Res. 13, 1–18.

Burgess, A.W., 2008. EGFR family: Structure physiology signalling and therapeutic target. Growth Factors 26, 263–274.

Bürkle, A., 2001. Poly(APD-ribosyl)ation, a DNA damage-driven protein modification and regulator of genomic instability. Cancer Lett. 163, 1–5.

Burridge, K., Chrzanowska-Wodnicka, M., 1996. Focal Adhesions, Contractility, and Signaling. Annu. Rev. Cell Dev. Biol. 12, 463–519.

Burstein, M.D., Tsimelzon, A., Poage, G.M., Covington, K.R., Contreras, A., Fuqua, S.A.W., Savage, M.I., Osborne, C.K., Hilsenbeck, S.G., Chang, J.C., Mills, G.B., Lau, C.C., Brown, P.H., 2015. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin. Cancer Res. 21, 1688–1698.

Byrski, T., Huzarski, T., Dent, R., Gronwald, J., Zuziak, D., Cybulski, C., Kladny, J., Gorski, B., Lubinski, J., Narod, S.A., 2009. Response to neoadjuvant therapy with cisplatin in BRCA1-positive breast cancer patients. Breast Cancer Res. Treat. 115, 359–363.

Byrski, T., Huzarski, T., Dent, R., Marczyk, E., Jasiowka, M., Gronwald, J., Jakubowicz, J., Cybulski, C., Wisniowski, R., Godlewski, D., Lubinski, J., Narod, S.A., 2014. Pathologic complete response to neoadjuvant cisplatin in BRCA1-positive breast cancer patients. Breast Cancer Res. Treat. 147, 401–405.

Calle, E.E., Kaaks, R., 2004. Overweight, obesity and cancer: Epidemiological evidence and proposed mechanisms. Nat. Rev. Cancer 4, 579–591.

Camarda, R., Zhou, A.Y., Kohnz, R.A., Balakrishnan, S., Mahieu, C., Anderton, B., Eyob, H., Kajimura, S., Tward, A., Krings, G., Nomura, D.K., Goga, A., 2016. Inhibition of fatty acid oxidation as a therapy for MYC-overexpressing triple-negative breast cancer. Nat. Med. 22, 427–432.

Carey, J.P.W., Karakas, C., Bui, T., Chen, X., Vijayaraghavan, S., Zhao, Y., Wang, J., Mikule, K., Litton, J.K., Hunt, K.K., Keyomarsi, K., 2018. Synthetic lethality of PARP inhibitors in combination with MYC blockade is independent of BRCA status in triple-negative breast cancer. Cancer Res. 78, 742–757.

Carey, L.A., Dees, E.C., Sawyer, L., Gatti, L., Moore, D.T., Collichio, F., Ollila, D.W., Sartor, C.I., Graham, M.L., Perou, C.M., 2007. The triple negative paradox: Primary tumor chemosensitivity of breast cancer subtypes. Clin. Cancer Res. 13, 2329–2334.

Carey, L.A., Perou, C.M., Livasy, C.A., Dressler, L.G., Cowan, D., Conway, K., Troester, M.A., Tse, C.K., Edmiston, S., Deming, S.L., Geradts, J., Cheang, M.C.U., Moorman, P.G., Earp, H.S., 2006. Race , Breast Cancer Subtypes , and Survival in the Carolina Breast Cancer Study. JAMA 295, 2492–2502.

Carey, L.A., Rugo, H.S., Marcom, P.K., Mayer, E.L., Esteva, F.J., Ma, C.X., Liu, M.C., Storniolo, A.M., Rimawi, M.F., Forero-Torres, A., Wolff, A.C., Hobday, T.J., Ivanova, A., Chiu, W.K., Ferraro, M., Burrows, E., Bernard, P.S., Hoadley, K.A., Perou, C.M., Winer, E.P., 2012. TBCRC 001: Randomized phase II study of cetuximab in combination with carboplatin in stage IV triple-negative breast cancer. J. Clin. Oncol. 30, 2615–2623.

Carlo M. Croce, 2008. Oncogenes and Cancer. N. Engl. J. Med. 358, 502–11.

Carlson M, 2019. org.Hs.eg.db: Genome wide annotation for Human. R Packag. version 3.8.2.

Casella, G., 1992. Illustrating empirical Bayes methods. Chemom. Intell. Lab. Syst. 16, 107–125.

Castedo, M., Perfettini, J.L., Roumier, T., Andreau, K., Medema, R., Kroemer, G., 2004. Cell death by mitotic catastrophe: A molecular definition. Oncogene 23, 2825–2837.

Cavallaro, U., Christofori, G., 2004. Multitasking in tumor progression: Signaling functions of cell adhesion molecules. Ann. N. Y. Acad. Sci. 1014, 58–66.

Chang, C.S., Kitamura, E., Johnson, J., Bollag, R., Hawthorn, L., 2018. Genomic analysis of racial differences in triple negative breast cancer. Genomics 1–14.

Cheang, M.C.U., Martin, M., Nielsen, T.O., Prat, A., Voduc, D., Rodriguez-Lescure, A., Ruiz, A., Chia, S., Shepherd, L., Ruiz-Borrego, M., Calvo, L., Alba, E., Carrasco, E., Caballero, R., Tu, D., Pritchard, K.I., Levine, M.N., Bramwell, V.H., Parker, J., Bernard, P.S., Ellis, M.J., Perou, C.M., Di Leo, A., Carey, L.A., 2015. Defining Breast Cancer Intrinsic Subtypes by Quantitative Receptor Expression. Oncologist 20, 474–482.

Chen, J., Liu, J., Jin, R., Shen, J., Liang, Y., Ma, R., Lin, H., Liang, X., Yu, H., Cai, X., 2014. Cytoplasmic and/or nuclear expression of β-catenin correlate with poor prognosis and unfavorable clinicopathological factors in hepatocellular carcinoma: A meta-analysis. PLoS One 9.

Chen, L., Li, C.I., 2015. Racial disparities in breast cancer diagnosis and treatment by hormone receptor and HER2 status. Cancer Epidemiol. Biomarkers Prev. 24, 1666–1672.

Chen, X., Li, J., Gray, W.H., Lehmann, B.D., Bauer, J.A., Shyr, Y., Pietenpol, J.A., 2012. TNBCtype: A subtyping tool for triple-negative breast cancer. Cancer Inform. 11, 147–156.

Chen, Y., Lun, A.T.L., Smyth, G.K., 2016. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Research 5, 1438.

Chen, Y., McCarthy, D., Ritchie, M., Robinson, M., Smyth, G., 2018. edgeR: differential expression analysis of digital gene expression data. Bioconductor.Fhcrc.Org 1–76.

Cheng, X., Xu, X., Chen, D., Zhao, F., Wang, W., 2019. Therapeutic potential of targeting the Wnt/β-catenin signaling pathway in colorectal cancer. Biomed. Pharmacother. 110, 473–481.

Chlebowski, R.T., Anderson, G.L., Aragaki, A.K., Prentice, R., 2016. Breast Cancer and Menopausal Hormone Therapy by Race/Ethnicity and Body Mass Index. J. Natl. Cancer Inst. 108, 2014–2017.

Chlebowski, R.T., Anderson, G.L., Gass, M., Lane, D.S., Aragaki, A.K., Kuller, L.H., Wactawski-Wende, J., Manson, J.E., Stefanick, M.L., Ockene, J., Sarto, G.E., Johnson, K.C., Ravdin, P.M., Schenken, R., Hendrix, S.L., Rajkovic, A., Rohan, T.E., Yasmeen, S., Prentice, R.L., WHI Investigators, 2010. Estrogen plus progestin and breast cancer incidence and mortality in postmenopausal women. JAMA 304, 1684–1692.

Chlebowski, R.T., Chen, Z., Anderson, G.L., Rohan, T., Aragaki, A., Lane, D., Dolan, N.C., Paskett, E.D., McTiernan, A., Hubbell, F.A., Adams-Campbell, L.L., Prentice, R., 2005. Ethnicity and breast cancer: Factors influencing differences in incidence and outcome. J. Natl. Cancer Inst. 97, 439–

Choi, J.E., Kang, S.H., Lee, S.J., Bae, Y.K., 2014. Prognostic significance of Bcl-2 expression in non-basal

triple-negative breast cancer patients treated with anthracycline-based chemotherapy. Tumor Biol. 35, 12255–12263.

Choong, L.Y., Lim, S., Loh, M.C.S., Man, X., Chen, Y., Toy, W., Pan, M., Chen, C.S., Poonepalli, A., Hande, M.P., Tan, P.H., Salto-Tellez, M., Wong, C.Y., Shah, N., Druker, B.J., Lim, Y.P., 2007. Progressive loss of epidermal growth factor receptor in a subpopulation of breast cancers: Implications in target-directed therapeutics. Mol. Cancer Ther. 6, 2828–2842.

Chow-White, P.A., Green, S.E., 2013. Data mining difference in the age of big data: Communication and the social shaping of genome technologies from 1998 to 2007. Int. J. Commun. 7, 556–583.

Chung, G.G., Zerkowski, M.P., Ocal, I.T., Dolled-Filhart, M., Kang, J.Y., Psyrri, A., Camp, R.L., Rimm, D.L., 2004. β-Catenin and p53 Analyses of a Breast Carcinoma Tissue Microarray. Cancer 100, 2084–2092.

Cimino-Mathews, A., Thompson, E., Taube, J.M., Ye, X., Lu, Y., Meeker, A., Xu, H., Sharma, R., Lecksell, K., Cornish, T.C., Cuka, N., Argani, P., Emens, L.A., 2016. PD-L1 (B7-H1) expression and the immune tumor microenvironment in primary and metastatic breast carcinomas. Hum. Pathol. 47, 52–63.

Ciriello, G., Miller, M.L., Aksoy, B.A., Senbabaoglu, Y., Schultz, N., Sander, C., 2013. Emerging landscape of oncogenic signatures across human cancers. Nat. Publ. Gr. 45, 1127–1135.

Clarke, C.A., Keegan, T.H.M., Yang, J., Press, D.J., Kurian, A.W., Patel, A.H., Lacey, J. V., 2012. Age-specific incidence of breast cancer subtypes: Understanding the black-white crossover. J. Natl. Cancer Inst. 104, 1094–1101.

Coates, R.J., Clark, W.S., Eley, J.W., Greenberg, R.S., Huguly, C.M., Brown, R.L., 1990. Race, nutritional status, and survival from breast cancer. J. Natl. Cancer Inst. 82, 1684–1692.

Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., Ceccarelli, M., Bontempi, G., Noushmehr, H., 2016. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 44, e71.

Coles, B.F., Kadlubar, F.F., 2003. Detoxification of electrophilic compounds by glutathione S-transferase catalysis: Determinants of individual response to chemical carcinogens and chemotherapeutic drugs? BioFactors 17, 115–130.

Collaborative Group on Hormonal Factors in Breast Cancer, 1997. Breast cancer and hormone replacement therapy: Collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer (The Lancet (1997) Oct 11 (1047)). Lancet 350, 1484.

Collett, K., Stefansson, I.M., Eide, J., Braaten, A., Wang, H., Eide, G.E., Thoresen, S.Ø., Foulkes, W.D., Akslen, L.A., 2005. A basal epithelial phenotype is more frequent in interval breast cancers compared with screen detected tumors. Cancer Epidemiol. Biomarkers Prev. 14, 1108–1112.

Collins, F.S., 2004. What we do and don't know about "race", "ethnicity", genetics and health at the dawn of the genome era. Nat. Genet. 36, 13–15.

Conklin, M.W., Eickhoff, J.C., Riching, K.M., Pehlke, C.A., Eliceiri, K.W., Provenzano, P.P., Friedl, A., Keely, P.J., 2011. Aligned collagen is a prognostic signature for survival in human breast carcinoma. Am. J. Pathol. 178, 1221–1232.

Conway, J.R., Lex, A., Gehlenborg, N., 2017. Genome analysis UpSetR : an R package for the visualization of intersecting sets and their properties 33, 2938–2940.

Copson, E., Maishman, T., Gerty, S., Eccles, B., Stanton, L., Cutress, R.I., Altman, D.G., Durcan, L., Simmonds, P., Jones, L., Tapper, W., Eccles, D., 2014. Ethnicity and outcome of young breast cancer patients in the United Kingdom: The POSH study. Br. J. Cancer 110, 230–241.

Cordenonsi, M., Zanconato, F., Azzolin, L., Forcato, M., Rosato, A., Frasson, C., Inui, M., Montagner, M., Parenti, A.R., Poletti, A., Daidone, M.G., Dupont, S., Basso, G., Bicciato, S., Piccolo, S., 2011. The hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. Cell 147, 759–772.

Cortazar, P., Zhang, L., Untch, M., Mehta, K., Costantino, J.P., Wolmark, N., Bonnefoi, H., Cameron, D., Gianni, L., Valagussa, P., Swain, S.M., Prowell, T., Loibl, S., Wickerham, D.L., Bogaerts, J., Baselga, J., Perou, C., Blumenthal, G., Blohmer, J., Mamounas, E.P., Bergh, J., Semiglazov, V., Justice, R., Eidtmann, H., Paik, S., Piccart, M., Sridhara, R., Fasching, P.A., Slaets, L., Tang, S., Gerber, B., Geyer, C.E., Pazdur, R., Ditsch, N., Rastogi, P., Eiermann, W., Von Minckwitz, G., 2014. Pathological complete response and long-term clinical benefit in breast cancer: The CTNeoBC pooled analysis. Lancet 384, 164–172.

Cory, S., Adams, J.M., 2002. The BCL2 family: Regulators of the cellular life-or-death switch. Nat. Rev. Cancer 2, 647–656.

Cossu-Rocca, P., Orrù, S., Muroni, M.R., Sanges, F., Sotgiu, G., Ena, S., Pira, G., Murgia, L., Manca, A., Uras, M.G., Sarobba, M.G., Urru, S., De Miglio, M.R., 2015. Analysis of PIK3CA mutations and activation pathways in triple negative breast cancer. PLoS One 10, 1–14.

Costa, R.L.B., Han, H.S., Gradishar, W.J., 2018. Targeting the PI3K/AKT/mTOR pathway in triple-negative breast cancer: a review. Breast Cancer Res. Treat. 169, 397–406.

Cowling, V.H., Cole, M.D., 2007. Turning the tables: Myc activates Wnt in breast cancer. Cell Cycle 6, 2625–2627.

Cox, D.R., Reid, N., 1987. Parameter Orthogonality and Approximate Conditional Inference. J. R. Stat. Soc. Ser. B 49, 1–18.

Craig Venter, J., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Yuan Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M.H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S.C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Lai Cheng, M., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Ni Tint, N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Deslattes Mays, A.,

Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The sequence of the human genome. Science (80-.). 291, 1304–1351.

Creighton, C.J., Li, X., Landis, M., Dixon, J.M., Neumeister, V.M., Sjolund, A., Rimm, D.L., Wong, H., Rodriguez, A., Herschkowitz, J.I., Fan, C., Zhang, X., He, X., Pavlick, A., Gutierrez, M.C., Renshaw, L., Larionov, A.A., Faratian, D., Hilsenbeck, S.G., Perou, C.M., Lewis, M.T., Rosen, J.M., Chang, J.C., 2009. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc. Natl. Acad. Sci. U. S. A. 106, 13820–13825.

Criscitiello, C., Esposito, A., Trapani, D., Curigliano, G., 2016. Prognostic and predictive value of tumor infiltrating lymphocytes in early breast cancer. Cancer Treat. Rev. 50, 205–207.

Crowe, J.P., Gordon, N.H., Hubay, C.A., Pearson, O.H., Marshall, J.S., McGuire, W.L., 1986. The interaction of estrogen receptor status and race in predicting prognosis for stage I breast cancer patients. Surgery 100, 599–605.

Curtin, J.A., Fridlyand, J., Kageshita, T., Patel, H.N., Busam, K.J., Kutzner, H., Cho, K.H., Aiba, S., Bröcker, E.B., LeBoit, P.E., Pinkel, D., Bastian, B.C., 2005. Distinct sets of genetic alterations in melanoma. N. Engl. J. Med. 353, 2135–2147.

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Caldas, C., Aparicio, S., Curtis†, C., Shah, S.P., Caldas, C., Aparicio, S., Brenton, J.D., Ellis, I., Huntsman, D., Pinder, S., Purushotham, A., Murphy, L., Caldas, C., Aparicio, S., Caldas, C., Bardwell, H., Chin, S.-F., Curtis, C., Ding, Z., Gräf, S., Jones, L., Liu, B., Lynch, A.G., Papatheodorou, I., Sammut, S.J., Wishart, G., Aparicio, S., Chia, S., Gelmon, K., Huntsman, D., McKinney, S., Speers, C., Turashvili, G., Watson, P., Ellis, I., Blamey, R., Green, A., Macmillan, D., Rakha, E., Purushotham, A., Gillett, C., Grigoriadis, A., Pinder, S., di Rinaldis, E., Tutt, A., Murphy, L., Parisien, M., Troup, S., Caldas, C., Chin, S.-F., Chan, D., Fielding, C., Maia, A.-T., McGuire, S., Osborne, M., Sayalero, S.M., Spiteri, I., Hadfield, J., Aparicio, S., Turashvili, G., Bell, L., Chow, K., Gale, N., Huntsman, D., Kovalik, M., Ng, Y., Prentice, L., Caldas, C., Tavaré, S., Curtis, C., Dunning, M.J., Gräf, S., Lynch, A.G., Rueda, O.M., Russell, R., Samarajiwa, S., Speed, D., Markowetz, F., Yuan, Y., Brenton, J.D., Aparicio, S., Shah, S.P., Bashashati, A., Ha, G., Haffari, G., McKinney, S., Langerød, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A.-L., Brenton, J.D., Tavaré, S., Caldas, C., Aparicio, S., 2012a. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 1–7.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Aparicio, S., Brenton, J.D., Ellis, I., Huntsman, D., Pinder, S., Murphy, L., Bardwell, H., Ding, Z., Jones, L., Liu, B., Papatheodorou, I., Sammut, S.J., Wishart, G., Chia, S., Gelmon, K., Speers, C., Watson, P., Blamey, R., Green, A., MacMillan, D., Rakha, E., Gillett, C., Grigoriadis, A., De Rinaldis, E., Tutt, A., Parisien, M., Troup, S., Chan, D., Fielding, C., Maia, A.T., McGuire, S., Osborne, M., Sayalero, S.M., Spiteri, I., Hadfield, J., Bell, L., Chow, K., Gale, N., Kovalik, M., Ng, Y., Prentice, L., Tavaré, S., Markowetz, F., Langerød, A., Provenzano, E., Purushotham, A., Børresen-Dale, A.L., Caldas, C., 2012b. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346–352.

D'Arcy, M., Fleming, J., Robinson, W.R., Kirk, E.L., Perou, C.M., Troester, M.A., 2015. Race-associated biological differences among Luminal A breast tumors. Breast Cancer Res. Treat. 152, 437–448.

Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., Shi, B., 2015. Breast cancer intrinsic subtype classification, clinical use and future trends. Am. J. Cancer Res. 5, 2929–2943.

Daly, B., Olopade, O.I., 2015. Race, Ethnicity, and the Diagnosis of Breast Cancer. JAMA - J. Am. Med. Assoc. 313, 141–142.

Daly, M.B., Pal, T., Berry, M.P., Buys, S.S., Dickson, P., Domchek, S.M., Elkhanany, A., Friedman, S., Goggins, M., Hutton, M.L., Karlan, B.Y., Khan, S., Klein, C., Kohlmann, W., Kurian, A.W., Laronga, C., Litton, J.K., Mak, J.S., Menendez, C.S., Merajver, S.D., Norquist, B.S., Offit, K., Pederson, H.J., Reiser, G., Senter-Jamieson, L., Shannon, K.M., Shatsky, R., Visvanathan, K., Weitzel, J.N., Wick, M.J., Wisinski, K.B., Yurgelun, M.B., Darlow, S.D., Dwyer, M.A., 2021. Genetic/familial high-risk assessment: Breast, ovarian, and pancreatic, version 2.2021. JNCCN J. Natl. Compr. Cancer Netw. 19, 77–102. 1

Danforth, D.N., 2013. Disparities in breast cancer outcomes between Caucasian and African American women: A model for describing the relationship of biological and nonbiological factors. Breast Cancer Res. 15, 11–16.

Dang, C. V., 2012. MYC on the path to cancer. Cell 149, 22–35.

Dang, C. V., O'Donnell, K.A., Zeller, K.I., Nguyen, T., Osthus, R.C., Li, F., 2006. The c-Myc target gene network. Semin. Cancer Biol. 16, 253–264.

Davies, B.R., Greenwood, H., Dudley, P., Crafter, C., Yu, D.H., Zhang, J., Li, J., Gao, B., Ji, Q., Maynard, J., Ricketts, S.A., Cross, D., Cosulich, S., Chresta, C.C., Page, K., Yates, J., Lane, C., Watson, R., Luke, R., Ogilvie, D., Pass, M., 2012. Preclinical pharmacology of AZD5363, an inhibitor of AKT: Pharmacodynamics, antitumor activity, and correlation of monotherapy activity with genetic background. Mol. Cancer Ther. 11, 873–887.

Davis, M.B., Walens, A., Hire, R., Mumin, K., Brown, A.M., Ford, D.J., Howerth, E.W., Monteil, M., 2015. Distinct transcript isoforms of the atypical chemokine receptor 1 (ACKR1) / Duffy Antigen Receptor for Chemokines (DARC) gene are expressed in lymphoblasts and altered isoform levels are associated with genetic ancestry and the duffy-null allele. PLoS One 10, 1–16.

Dawson, S., Makretsov, N., Blows, F.M., Driver, K.E., Provenzano, E., Quesne, J. Le, Baglietto, L., 2010. BCL2 in breast cancer : a favourable prognostic marker across molecular subtypes and independent of adjuvant therapy received. Br. J. Cancer 668–675.

de Bono, J., Ramanathan, R.K., Mina, L., Chugh, R., Glaspy, J., Rafii, S., Kaye, S., Sachdev, J., Heymach, J., Smith, D.C., Henshaw, J.W., Herriott, A., Patterson, M., Curtin, N.J., Byers, L.A., Wainberg, Z.A., 2017. Phase I, dose-escalation, two-part trial of the PARP inhibitor talazoparib in patients with advanced germline BRCA1/2 mutations and selected sporadic cancers. Cancer Discov. 7, 620–629.

De, P., Sun, Y., Carlson, J.H., Friedman, L.S., Leyland-Jones, B.R., Dey, N., 2014. Doubling down on the PI3k-AKT-mTOR pathway enhances the antitumor efficacy of PARP inhibitor in triple negative breast cancer model beyond BRCA-ness. Neoplasia (United States) 16, 43–72.

Denduluri, N., Somerfield, M.R., Eisen, A., Holloway, J.N., Hurria, A., King, T.A., Lyman, G.H., Partridge, A.H., Telli, M.L., Trudeau, M.E., Wolff, A.C., 2016. Selection of optimal adjuvant chemotherapy regimens for human epidermal growth factor receptor 2 (HER2) -negative and adjuvant targeted therapy for HER2-positive breast cancers: An American Society of Clinical Oncology Guideline Adaptation of the Cancer C. J. Clin. Oncol. 34, 2416–2427.

Denkert, C., Loibl, S., Noske, A., Roller, M., Müller, B.M., Komor, M., Budczies, J., Darb-Esfahani, S., Kronenwett, R., Hanusch, C., Von Törne, C., Weichert, W., Engels, K., Solbach, C., Schrader, I., Dietel, M., Von Minckwitz, G., 2010. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J. Clin. Oncol. 28, 105–113.

Denkert, C., Von Minckwitz, G., Brase, J.C., Sinn, B. V., Gade, S., Kronenwett, R., Pfitzner, B.M., Salat, C., Loi, S., Schmitt, W.D., Schem, C., Fisch, K., Darb-Esfahan, S., Mehta, K., Sotiriou, C., Wienert, S., Klare, P., André, F., Klauschen, F., Blohmer, J.U., Krappmann, K., Schmidt, M., Tesch, H., Kümmel, S., Sinn, P., Jackisch, C., Dietel, M., Reimer, T., Untch, M., Loibl, S., 2015. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. J. Clin. Oncol. 33, 983–991.

Dent, R., Trudeau, M., Pritchard, K.I., Hanna, W.M., Kahn, H.K., Sawka, C.A., Lickley, L.A., Rawlinson, E., Sun, P., Narod, S.A., 2007. Triple-negative breast cancer: Clinical features and patterns of recurrence. Clin. Cancer Res. 13, 4429–4434.

Desantis, C.E., Fedewa, S.A., Sauer, A.G., Kramer, J.L., Smith, R.A., Jemal, A., 2016. Breast Cancer Statistics, 2015: Convergence of Incidence Rates Between Black and White Women. CA Cancer 66, 31–42.

DeSantis, C.E., Ma, J., Gaudet, M.M., Newman, L.A., Miller, K.D., Goding Sauer, A., Jemal, A., Siegel, R.L., 2019. Breast cancer statistics, 2019. CA. Cancer J. Clin. 69, 438–451.

DeVita, V.T., Rosenberg, S.A., 2012. Two hundred years of cancer research. N. Engl. J. Med. 366, 2207–2214.

Dey, N., Barwick, B.G., Moreno, C.S., Ordanic-Kodani, M., Chen, Z., Oprea-Ilies, G., Tang, W., Catzavelos, C., Kerstann, K.F., Sledge, G.W., Abramovitz, M., Bouzyk, M., De, P., Leyland-Jones, B.R., 2013a. Wnt signaling in triple negative breast cancer is associated with metastasis. BMC Cancer 13.

Dey, N., Young, B., Abramovitz, M., Bouzyk, M., Barwick, B., De, P., Leyland-Jones, B., 2013b. Differential Activation of Wnt-β-Catenin Pathway in Triple Negative Breast Cancer Increases MMP7 in a PTEN Dependent Manner. PLoS One 8, 1–17.

Díaz-Martín, J., López-García, M.Á., Romero-Pérez, L., Atienza-Amores, M.R., Pecero, M.L., Castilla, M.Á., Biscuola, M., Santón, A., Palacios, J., 2015. Nuclear TAZ expression associates with the triple-negative phenotype in breast cancer. Endocr. Relat. Cancer 22, 443–454.

Dieci, M. V., Mathieu, M.C., Guarneri, V., Conte, P., Delaloge, S., Andre, F., Goubar, A., 2015. Prognostic and predictive value of tumor-infiltrating lymphocytes in two phase III randomized adjuvant breast cancer trials. Ann. Oncol. 26, 1698–1704.

Dieci, M. V., Orvieto, E., Dominici, M., Conte, P., Guarneri, V., 2014. Rare Breast Cancer Subtypes: Histological, Molecular, and Clinical Peculiarities. Oncologist 19, 805–813.

Dietze, E.C., Chavez, T.A., Seewaldt, V.L., 2018. Obesity and Triple-Negative Breast Cancer Disparities, Controversies, and Biology. Am. J. Pathol. 188, 280–290.

Dietze, E.C., Sistrunk, C., Miranda-Carboni, G., O'Regan, R., Seewaldt, V.L., 2015a. Triple-negative breast cancer in African-American women: disparities versus biology. Nat. Rev. Cancer 15, 248–254.

Dietze, E.C., Sistrunk, C., Miranda-Carboni, G., O'Regan, R., Seewaldt, V.L., 2015b. Triple-negative breast cancer in African-American women: Disparities versus biology. Nat. Rev. Cancer 15, 248–254.

Dillon, D.A., D'Aquila, T., Reynolds, A.B., Fearon, E.R., Rimm, D.L., 1998. The expression of p120ctn protein in breast cancer is independent of α- and β-catenin and E-cadherin. Am. J. Pathol. 152, 75–82.

Ding, L., Ellis, M.J., Li, S., Larson, D.E., Chen, K., Wallis, J.W., Harris, C.C., McLellan, M.D., Fulton, R.S., Fulton, L.L., Abbott, R.M., Hoog, J., Dooling, D.J., Koboldt, D.C., Schmidt, H., Kalicki, J., Zhang, Q., Chen, L., Lin, Ling, Wendl, M.C., McMichael, J.F., Magrini, V.J., Cook, L., McGrath, S.D., Vickery,

T.L., Appelbaum, E., DeSchryver, K., Davies, S., Guintoli, T., Lin, Li, Crowder, R., Tao, Y., Snider, J.E., Smith, S.M., Dukes, A.F., Sanderson, G.E., Pohl, C.S., Delehaunty, K.D., Fronick, C.C., Pape, K.A., Reed, J.S., Robinson, J.S., Hodges, J.S., Schierding, W., Dees, N.D., Shen, D., Locke, D.P., Wiechert, M.E., Eldred, J.M., Peck, J.B., Oberkfell, B.J., Lolofie, J.T., Du, F., Hawkins, A.E., O'Laughlin, M.D., Bernard, K.E., Cunningham, M., Elliott, G., Mason, M.D., Thompson Jr, D.M., Ivanovich, J.L., Goodfellow, P.J., Perou, C.M., Weinstock, G.M., Aft, R., Watson, M., Ley, T.J., Wilson, R.K., Mardis, E.R., 2010. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 464, 999–1005.

Dolled-Filhart, M., McCabe, A., Giltnane, J., Cregger, M., Camp, R.L., Rimm, D.L., 2006. Quantitative in situ analysis of β-catenin expression in breast cancer shows decreased expression is associated with poor outcome. Cancer Res. 66, 5487–5494.

Douglas Hanahan and Robert A. Weinberg, 2000. The Hallmarks of Cancer. Cell 100, 57–70.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., Huber, W., 2005. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. Bioinformatics 21, 3439–3440.

Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor package biomaRt. Nat Protoc 4, 1184–1191.

E. Goldberg, J., L. Schwertfeger, K., 2010. Proinflammatory Cytokines in Breast Cancer: Mechanisms of Action and Potential Targets for Therapeutics. Curr. Drug Targets 11, 1133–1146.

Eades, G., Wolfson, B., Zhang, Y., Li, Q., Yao, Y., Zhou, Q., 2015. LincRNA-RoR and miR-145 regulate invasion in triple-negative breast cancer via targeting Arf6. Mol. Cancer Res. 13, 330–338.

Eberwein, P., Laird, D., Schulz, S., Reinhard, T., Steinberg, T., Tomakidi, P., 2015. Modulation of focal adhesion constituents and their down-stream events by EGF: On the cross-talk of integrins and growth factor receptors. Biochim. Biophys. Acta - Mol. Cell Res. 1853, 2183–2198.

Echavarria, I., Lopez-Tarruella, S., Picornell, A., García-Saenz, J.A., Jerez, Y., Hoadley, K., Gomez, H.L., Moreno, F., Monte-Millan, M. Del, Marquez-Rodas, I., Alvarez, E., Ramos-Medina, R., Gayarre, J., Massarrah, T., Ocaña, I., Cebollero, M., Fuentes, H., Barnadas, A., Ballesteros, A.I., Bohn, U., Perou, C.M., Martin, M., 2018. Pathological response in a triple-negative breast cancer cohort treated with neoadjuvant carboplatin and docetaxel according to Lehmann's refined classification. Clin. Cancer Res. 24, 1845–1852.

El Guerrab, A., Bamdad, M., Bignon, Y.J., Penault-Llorca, F., Aubel, C., 2017. Anti-EGFR monoclonal antibodies enhance sensitivity to DNA-damaging agents in BRCA1-mutated and PTEN-wild-type triple-negative breast cancer cells. Mol. Carcinog. 56, 1383–1394. https://doi.org/10.1002/mc.22596

Engebraaten, O., Vollan, H.K.M., Børresen-Dale, A.L., 2013. Triple-negative breast cancer and the need for new therapeutic targets. Am. J. Pathol. 183, 1064–1074.

Eom, Y.H., Kim, H.S., Lee, A., Song, B.J., Chae, B.J., 2016. BCL2 as a subtype-specific prognostic marker for breast cancer. J. Breast Cancer 19, 252–260.

Fackenthal, J.D., Sveen, L., Gao, Q., Kohlmeir, E.K., Adebamowo, C., Ogundiran, T.O., Adenipekun, A.A., Oyesegun, R., Campbell, O., Rotimi, C., Akang, E.E.U., Das, S., Olopade, O.I., 2005. Complete allelic analysis of BRCA1 and BRCA2 variants in young Nigerian breast cancer patients. J. Med. Genet. 42, 276–281.

Fallah, Y., Brundage, J., Allegakoen, P., Shajahan-Haq, A.N., 2017. MYC-Driven pathways in breast cancer subtypes. Biomolecules 7, 1–6.

Fan, C., Oh, D.S., Wessels, L., Weigelt, B., Nuyten, D.S.A., Nobel, A.B., Van't Veer, L.J., Perou, C.M.,

2006. Concordance among gene-expression-based predictors for breast cancer. N. Engl. J. Med. 355, 560–569.

Fanelli, M.A., Montt-Guevara, M., Diblasi, A.M., Gago, F.E., Tello, O., Cuello-Carrión, F.D., Callegari, E., Bausero, M.A., Ciocca, D.R., 2008. P-Cadherin and β-catenin are useful prognostic markers in breast cancer patients; β-catenin interacts with heat shock protein Hsp27. Cell Stress Chaperones 13, 207–220.

Farhat, G.N., Cummings, S.R., Chlebowski, R.T., Parimi, N., Cauley, J.A., Rohan, T.E., Huang, A.J., Vitolins, M., Hubbell, F.A., Manson, J.E., Cochrane, B.B., Lane, D.S., Lee, J.S., 2011. Sex hormone levels and risks of estrogen receptor-negative and estrogen receptor-positive breast cancers. J. Natl. Cancer Inst. 103, 562–570.

Farmer, H., McCabe, H., Lord, C.J., Tutt, A.H.J., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., Martin, N.M.B., Jackson, S.P., Smith, G.C.M., Ashworth, A., 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 434, 917–921.

Ferroni, P., Riondino, S., Buonomo, O., Palmirotta, R., Guadagni, F., Roselli, M., 2015. Type 2 diabetes and breast cancer: The interplay between impaired glucose metabolism and oxidant stress. Oxid. Med. Cell. Longev. 2015.

Field, L.A., Love, B., Deyarmin, B., Hooke, J.A., Shriver, C.D., Ellsworth, R.E., 2012. Identification of differentially expressed genes in breast tumors from African American compared with Caucasian women. Cancer 118, 1334–1344.

Filho, O.M., Giobbie-Hurder, A., Mallon, E., Gusterson, B., Viale, G., Winer, E.P., Thürlimann, B., Gelber, R.D., Colleoni, M., Ejlertsen, B., Debled, M., Price, K.N., Regan, M.M., Coates, A.S., Goldhirsch, A., 2015. Relative effectiveness of letrozole compared with tamoxifen for patients with lobular carcinoma in the BIG 1-98 Trial. J. Clin. Oncol. 33, 2772–2778.

Fillmore, C., Kuperwasser, C., 2007. Human breast cancer stem cell markers CD44 and CD24: Enriching for cells with functional properties in mice or in man? Breast Cancer Res. 9, 5–7.

Fioretti, F.M., Sita-Lumsden, A., Bevan, C.L., Brooke, G.N., 2014. Revising the role of the androgen receptor in breast cancer. J. Mol. Endocrinol. 52.

Foidart, P., Yip, C., Radermacher, J., Blacher, S., Lienard, M., Montero-Ruiz, L., Maquoi, E., Montaudon, E., Château-Joubert, S., Collignon, J., Coibion, M., Jossa, V., Marangoni, E., Noel, A., Sounni, N.E., Jerusalem, G., 2019. Expression of MT4-MMP, EGFR, and Rb in triple-negative breast cancer strongly sensitizes tumors to erlotinib and palbociclib combination therapy. Clin. Cancer Res. 25, 1838–1850.

Foster, M.W., Sharp, R.R., 2004. Beyond race: Towards a whole-genome perspective on human populations and genetic variation. Nat. Rev. Genet. 5, 790–796.

Fouad, Y.A., Aanei, C., 2017. Revisiting the hallmarks of cancer. Am. J. Cancer Res. 7, 1016–1036.

Foulkes, W.D., 2003. Germline BRCA1 Mutations and a Basal Epithelial Phenotype in Breast Cancer. CancerSpectrum Knowl. Environ. 95, 1482–1485.

Frasci, G., Comella, P., Rinaldo, M., Iodice, G., Di Bonito, M., D'Aiuto, M., Petrillo, A., Lastoria, S., Siani, C., Comella, G., D'Aiuto, G., 2009. Preoperative weekly cisplatin-epirubicin-paclitaxel with G-CSF support in triple-negative large operable breast cancer. Ann. Oncol. 20, 1185–1192.

Frederick, J.P., Liberati, N.T., Waddell, D.S., Shi, Y., Wang, X.-F., 2004. Transforming Growth Factor β-Mediated Transcriptional Repression of c-myc Is Dependent on Direct Binding of Smad3 to a Novel Repressive Smad Binding Element. Mol. Cell. Biol. 24, 2546–2559.

Fregene, A., Newman, L.A., 2005. Breast cancer in sub-Saharan Africa: How does it relate to breast

cancer in African-American women? Cancer 103, 1540–1550.

Frisch, S.M., Vuori, K., Ruoslahti, E., Chan-Hui, P.Y., 1996. Control of adhesion-dependent cell survival by focal adhesion kinase. J. Cell Biol. 134, 793–799.

Furberg, H., Millikan, R., Dressler, L., Newman, B., Geradts, J., 2001. Tumor characteristics in African American and white women. Breast Cancer Res. Treat. 68, 33–43.

Furth, N., Pateras, I.S., Rotkopf, R., Vlachou, V., Rivkin, I., Schmitt, I., Bakaev, D., Gershoni, A., Ainbinder, E., Leshkowitz, D., Johnson, R.L., Gorgoulis, V.G., Oren, M., Aylon, Y., 2018. LATS1 and LATS2 suppress breast cancer progression by maintaining cell identity and metabolic state. Life Sci. Alliance 1, 1–18.

Gala, K., Chandarlapaty, S., 2014. Molecular Pathways: HER3 targeted therapy. Clin Cancer Res 20, 1410–1416.

Gao, Q., Adebamowo, C.A., Fackenthal, J., Das, S., Sveen, L., Falusi, A.G., Olopade, O.I., 2000. Protein truncating BRCA1 and BRCA2 mutations in African women with pre-menopausal breast cancer. Hum. Genet. 107, 192–194.

Gari, H.H., DeGala, G.D., Ray, R., Lucia, M.S., Lambert, J.R., 2016. PRL-3 engages the focal adhesion pathway in triple-negative breast cancer cells to alter actin structure and substrate adhesion properties critical for cell migration and invasion. Cancer Lett. 380, 505–512.

Garrido-Castro, A.C., Lin, N.U., Polyak, K., 2019. Insights into molecular classifications of triple-negative breast cancer: Improving patient selection for treatment. Cancer Discov. 9, 176–198.

Gasparini, G., Boracchi, P., Barbareschi, M., Palma, P.D., Caffo, O., Morelli, L., Doglioni, C., Mauri, F.A., Bevilacqua, P., Verderio, P., Pezzella, F., Harris, A.L., 1995. Expression of bcl-2 Protein Predicts Efficacy of Adjuvant Treatments in Operable Node-Positive Breast Cancer. Clin. Cancer Res. 1, 189–198.

Getz, J.E., Teoh, D.B., Nasser, S., Waibhav, T., Christophe, L.R., Yellapantula, V., Ahearn, M.E., Gomez, C.R., Jorda, M., Pegram, M.D., Carpten, J.D., Baumbach, L.L., 2015. Abstract A74: Differential Wnt signaling in African American and Caucasian women with triple-negative breast cancer. Cancer Epidemiol. Biomark. Prev 24, e74.

Gewinner, C., Wang, Z.C., Richardson, A., Teruya-feldstein, J., Etemadmoghadam, D., Bowtell, D., Barretina, J., Lin, W.M., Rameh, L., Salmena, L., Pandolfi, P.P., 2009. Article Evidence that Inositol Polyphosphate 4-Phosphatase Type II Is a Tumor Suppressor that Inhibits PI3K Signaling. Cancer Cell 16, 115–125.

Geyer, F.C., Lacroix-Triki, M., Savage, K., Arnedos, M., Lambros, M.B., MacKay, A., Natrajan, R., Reis-Filho, J.S., 2011. B-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. Mod. Pathol. 24, 209–231.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., Schott, A., Hayes, D., Birnbaum, D., Wicha, M.S., Dontu, G., 2007. ALDH1 Is a Marker of Normal and Malignant Human Mammary Stem Cells and a Predictor of Poor Clinical Outcome. Cell Stem Cell 1, 555–567.

Giovannucci, E., Harlan, D.M., Archer, M.C., Bergenstal, R.M., Gapstur, S.M., Habel, L.A., Pollak, M., Regensteiner, J.G., Yee, D., 2010. Diabetes and cancer: A consensus report. Diabetes Care 33, 1674–1685.

Giuliano, A.E., Connolly, J.L., Edge, S.B., Mittendorf, E.A., Rugo, H.S., Solin, L.J., Weaver, D.L., Winchester, D.J., Hortobagyi, G.N., 2017. Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA. Cancer J. Clin. 67, 290–303.

Gnant, M., Thomssen, C., Harbeck, N., 2015. St. Gallen/Vienna 2015: A brief summary of the consensus

discussion. Breast Care 10, 124–130.

Golubovskaya, V., Beviglia, L., Xu, L.H., Shelton Earp, H., Craven, R., Cance, W., 2002. Dual inhibition of focal adhesion kinase and epidermal growth factor receptor pathways cooperatively induces death receptor-mediated apoptosis in human breast cancer cells. J. Biol. Chem. 277, 38978–38987.

Gong, Y., Yan, K., Lin, F., Anderson, K., Sotiriou, C., Andre, F., Holmes, F.A., Valero, V., Booser, D., Pippen, J.E., Vukelja, S., Gomez, H., Mejia, J., Barajas, L.J., Hess, K.R., Sneige, N., Hortobagyi, G.N., Pusztai, L., Symmans, W.F., 2007. Determination of oestrogen-receptor status and ERBB2 status of breast carcinoma: a gene-expression profiling study. Lancet Oncol. 8, 203–211.

Gong, Y., Zhang, X., Chen, R., Wei, Y., Zou, Z., Chen, X., 2017. Cytoplasmic expression of C-MYC protein is associated with risk stratification of mantle cell lymphoma. PeerJ 2017, e3457.

Gonzalez, M.E., DuPrie, M.L., Krueger, H., Merajver, S.D., Ventura, A.C., Toy, K.A., Kleer, C.G., 2011. Histone methyltransferase EZH2 induces Akt-dependent genomic instability and BRCA1 inhibition in breast cancer. Cancer Res. 71, 2360–2370.

Goodwin, P.J., Phillips, K.A., West, D.W., Ennis, M., Hopper, J.L., John, E.M., O'Malley, F.P., Milne, R.L., Andrulis, I.L., Friedlander, M.L., Southey, M.C., Apicella, C., Giles, G.G., Longacre, T.A., 2012. Breast cancer prognosis in BRCA1 and BRCA2 mutation carriers: An international prospective breast cancer family registry population-based cohort study. J. Clin. Oncol. 30, 19–26.

Gopurappilly, R., Bhonde, R., 2015. Transcriptional profiling and functional network analyses of islet-like clusters (ILCs) generated from pancreatic stem cells in vitro. Genomics 105, 211–219.

Greenup, R., Buchanan, A., Lorizio, W., Rhoads, K., Chan, S., Leedom, T., King, R., McLennan, J., Crawford, B., Kelly Marcom, P., Shelley Hwang, E., 2013. Prevalence of BRCA mutations among women with triple-negative breast cancer (TNBC) in a genetic counseling cohort. Ann. Surg. Oncol. 20, 3254–3258.

Grivennikov, S.I., Greten, F.R., Karin, M., 2010. Immunity, Inflammation, and Cancer. Cell 140, 883–899.

Grossman, R.L., Heath, A.P., Ferretti, V., Varmus, H.E., Lowy, Do.R., Kibbe, W.A., Staudt, L.M., 2016. Toward a Shared Vision for Cancer Genomic Data. N. Engl. J. Med. 375, 1109–1112.

Grunda, J.M., Steg, A.D., He, Q., Steciuk, M.R., Byan-Parker, S., Johnson, M.R., Grizzle, W.E., 2012. Differential expression of breast cancer-associated genes between stage- and age-matched tumor specimens from African- and Caucasian-American Women diagnosed with breast cancer. BMC Res. Notes 5.

Grusche, F.A., Degoutin, J.L., Richardson, H.E., Harvey, K.F., 2011. The Salvador/Warts/Hippo pathway controls regenerative tissue growth in Drosophila melanogaster. Dev. Biol. 350, 255–266.

Grushko, T.A., Dignam, J.J., Das, S., Blackwood, A.M., Perou, C.M., Ridderstråle, K.K., Anderson, K.N., Wei, M.J., Adams, A.J., Hagos, F.G., Sveen, L., Lynch, H.T., Weber, B.L., Olopade, O.I., 2004. MYC Is Amplified in BRCA1-Associated Breast Cancers. Clin. Cancer Res. 10, 499–507.

Gu-trantien, C., Sotiriou, C., Willard-, K., Gu-trantien, C., Loi, S., Garaud, S., Equeter, C., Libin, M., Wind, A. De, Ravoet, M., Buanec, H. Le, Sibille, C., Singhal, S.K., Michiels, S., Rothé, F., Salgado, R., Larsimont, D., Piccart, M., Sotiriou, C., Willard-gallo, K., 2013. CD4 + follicular helper T cell infiltration predicts breast cancer survival Find the latest version : CD4 + follicular helper T cell infiltration predicts breast cancer survival. J. Clin. invest. 123, 2873–2892.

Guerrab, A. El, Bamdad, M., Kwiatkowski, F., Bignon, Y.J., Penault-Llorca, F., Aubel, C., 2016. Anti-EGFR monoclonal antibodies and EGFR tyrosine kinase inhibitors as combination therapy for triple-negative breast cancer. Oncotarget 7, 73618–73637.

Gunter, M.J., Wang, T., Cushman, M., Xue, X., Wassertheil-Smoller, S., Strickler, H.D., Rohan, T.E., Manson, J.E., McTiernan, A., Kaplan, R.C., Scherer, P.E., Chlebowski, R.T., Snetselaar, L., Wang, D., Ho, G.Y.F., 2015. Circulating Adipokines and Inflammatory Markers and Postmenopausal Breast Cancer Risk. J. Natl. Cancer Inst. 107, 1–10.

Guo, Y.A.N.J.U.N., Pan, W.E.I.W.E.I., Liu, S.B., Shen, Z.F.E.I., Xu, Y., Hu, L.L., 2020. ERK / MAPK signalling pathway and tumorigenesis (Review) 1997–2007.

Gwin, K., Buell-Gutbrod, R., Tretiakova, M., Montag, A., 2010. Epithelial-to-mesenchymal transition in metaplastic breast carcinomas with chondroid differentiation: Expression of the E-cadherin repressor snail. Appl. Immunohistochem. Mol. Morphol. 18, 526–531.

Haiman, C.A., Chen, G.K., Vachon, C.M., Canzian, F., Dunning, A., Millikan, R.C., Wang, X., Ademuyiwa, F., Ahmed, S., Ambrosone, C.B., Baglietto, L., Balleine, R., Bandera, E. V., Beckmann, M.W., Berg, C.D., Bernstein, L., Blomqvist, C., Blot, W.J., Brauch, H., Buring, J.E., Carey, L.A., Carpenter, J.E., Chang-Claude, J., Chanock, S.J., Chasman, D.I., Clarke, C.L., Cox, A., Cross, S.S., Deming, S.L., Diasio, R.B., Dimopoulos, A.M., Driver, W.R., Dünnebier, T., Durcan, L., Eccles, D., Edlund, C.K., Ekici, A.B., Fasching, P.A., Feigelson, H.S., Flesch-Janys, D., Fostira, F., Försti, A., Fountzilas, G., Gerty, S.M., Giles, G.G., Godwin, A.K., Goodfellow, P., Graham, N., Greco, D., Hamann, U., Hankinson, S.E., Hartmann, A., Hein, R., Heinz, J., Holbrook, A., Hoover, R.N., Hu, J.J., Hunter, D.J., Ingles, S.A., Irwanto, A., Ivanovich, J., John, E.M., Johnson, N., Jukkola-Vuorinen, A., Kaaks, R., Ko, Y.D., Kolonel, L.N., Konstantopoulou, I., Kosma, V.M., Kulkarni, S., Lambrechts, D., Lee, A.M., Marchand, L. Le, Lesnick, T., Liu, J., Lindstrom, S., Mannermaa, A., Margolin, S., Martin, N.G., Miron, P., Montgomery, G.W., Nevanlinna, H., Nickels, S., Nyante, S., Olswold, C., Palmer, J., Pathak, H., Pectasides, D., Perou, C.M., Peto, J., Pharoah, P.D.P., Pooler, L.C., Press, M.F., Pylkäs, K., Rebbeck, T.R., Rodriguez-Gil, J.L., Rosenberg, L., Ross, E., Rüdiger, T., Silva, I.D.S., Sawyer, E., Schmidt, M.K., Schulz-Wendtland, R., Schumacher, F., Severi, G., Sheng, X., Signorello, L.B., Sinn, H.P., Stevens, K.N., Southey, M.C., Tapper, W.J., Tomlinson, I., Hogervorst, F.B.L., Wauters, E., Weaver, J., Wildiers, H., Winqvist, R., Berg, D. Van Den, Wan, P., Xia, L.Y., Yannoukakos, D., Zheng, W., Ziegler, R.G., Siddiq, A., Slager, S.L., Stram, D.O., Easton, D., Kraft, P., Henderson, B.E., Couch, F.J., 2011. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptorĝ€"negative breast cancer. Nat. Genet. 43, 1210–1214.

Haldar, S., Chintapalli, J., Croce, C.M., 1996. Taxol induces bcl-2 phosphorylation and death of prostate cancer cells. Cancer Res. 56, 1253–1255.

Hall, I.J., Moorman, P.G., Millikan, R.C., Newman, B., 2005. Comparative analysis of breast cancer risk factors among African-American women and white women. Am. J. Epidemiol. 161, 40–51.

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: The next generation. Cell 144, 646–674.

Hartman, A.R., Kaldate, R.R., Sailer, L.M., Painter, L., Grier, C.E., Endsley, R.R., Griffin, M., Hamilton, S.A., Frye, C.A., Silberman, M.A., Wenstrup, R.J., Sandbach, J.F., 2012. Prevalence of BRCA mutations in an unselected population of triple-negative breast cancer. Cancer 118, 2787–2795.

Hartman, Z.C., Poage, G.M., Den Hollander, P., Tsimelzon, A., Hill, J., Panupinthu, N., Zhang, Y., Mazumdar, A., Hilsenbeck, S.G., Mills, G.B., Brown, P.H., 2013. Growth of triple-negative breast cancer cells relies upon coordinate autocrine expression of the proinflammatory cytokines IL-6 and IL-8. Cancer Res. 73, 3470–3480.

Hartsock, A., Nelson, W.J., 2008. Adherens and tight junctions: Structure, function and connections to the actin cytoskeleton. Biochim. Biophys. Acta - Biomembr. 1778, 660–669.

Harvey, K.F., Zhang, X., Thomas, D.M., 2013. The Hippo pathway and human cancer. Nat. Rev. Cancer 13, 246–257.

Hatsell, S., Rowlands, T., Hiremath, M., Cowin, P., 2003. β-Catenin and Tcfs in Mammary Development and Cancer. J. Mammary Gland Biol. Neoplasia 8, 145–158.

Hayashi, H., Higashi, T., Yokoyama, N., Kaida, T., Sakamoto, K., Fukushima, Y., Ishimoto, T., Kuroki, H., Nitta, H., Hashimoto, D., Chikamoto, A., Oki, E., Beppu, T., Baba, H., 2015. An imbalance in TAZ and YAP expression in hepatocellular carcinoma confers cancer stem cell-like behaviors contributing to disease progression. Cancer Res. 75, 4985–4997.

Haynes-Maslow, L., Godley, P., Dimartino, L., White, B., Odom, J., Richmond, A., Carpenter, W., 2014. African American women's perceptions of cancer clinical trials. Cancer Med. 3, 1430–1439.

He, T.C., Sparks, A.B., Rago, C., Hermeking, H., Zawel, L., Da Costa, L.T., Morin, P.J., Vogelstein, B., Kinzler, K.W., 1998. Identification of c-MYC as a target of the APC pathway. Science (80-.). 281, 1509–1512.

Henderson, I.C., Patek, A.J., 1998. The relationship between prognostic and predictive factors in the management of breast cancer. Breast Cancer Res. Treat.

Hennessy, B.T., Smith, D.L., Ram, P.T., Lu, Y., Mills, G.B., 2005. Exploiting the PI3K/AKT pathway for cancer drug discovery. Nat. Rev. Drug Discov. 4, 988–1004.

Henry, N.L., Shah, P.D., Haider, I., Freer, P.E., Jagsi, R., Sabel, M.S., 2020. Chapter 88: Cancer of the Breast. In Abeloff's Clinical Oncology, Sixth Edit. ed, New England Journal of Medicine. Elsevier Inc.

Hoeflich, K.P., Brien, C.O., Boyd, Z., Cavet, G., Guerrero, S., Jung, K., Januario, T., Savage, H., Punnoose, E., Truong, T., Zhou, W., Berry, L., Murray, L., Amler, L., Belvin, M., Friedman, L.S., Lackner, M.R., 2009. Cancer Therapy : Preclinical In vivo Antitumor Activity of MEK and Phosphatidylinositol 3-Kinase Inhibitors in Basal-Like Breast Cancer Models 15, 4649–4665.

Hong, W., Guan, K.L., 2012. The YAP and TAZ transcription co-activators: Key downstream effectors of the mammalian Hippo pathway. Semin. Cell Dev. Biol. 23, 785–793.

Honório, M., Guerra-Pereira, N., Silva, J., Alves, J., Filipa, A., Braga, S., 2016. Decreased Survival in African Patients with Triple Negative Breast Cancer. J Palliat Care Med 6, 4.

Hope S. Rugo, Olopade, O.I., DeMichele, A., Yau, C., Veer, L.J. van 't, Buxton, M.B., Hogarth, M., Hylton, N.M., Paoloni, M., Perlmutter, J., Symmans, W.F., Yee, D., Chien, A.J., Wallace, A.M., Kaplan, H.G., Boughey, J.C., Haddad, T.C., Albain, K.S., Liu, M.C., Isaacs, C., Khan, Q.J., Lang, J.E., Viscusi, R.K., Pusztai, L., Moulder, S.L., Chui, S.Y., Kemmer, K.A., Elias, A.D., Edmiston, K.K., Euhus, D.M., Haley, B.B., Nanda, R., Northfelt, D.W., Tripathy, D., Wood, W.C., Lyandres, J., Davis, S.E., Hirst, G.L., Sanil, A., Berry, D.A., Esserman, L.J. on behalf of the I.-S.I., 2016. Adaptive Randomization of Veliparib–Carboplatin Treatment in Breast Cancer. N Engl J Med 375, 139–148.

Horiuchi, D., Kusdra, L., Huskey, N.E., Chandriani, S., Lenburg, M.E., Gonzalez-Angulo, A.M., Creasman, K.J., Bazarov, A. V., Smyth, J.W., Davis, S.E., Yaswen, P., Mills, G.B., Esserman, L.J., Goga, A., 2012. MYC pathway activation in triple-negative breast cancer is synthetic lethal with CDK inhibition. J. Exp. Med. 209, 679–696.

Hudis, C., Swanton, C., Janjigian, Y.Y., Lee, R., Sutherland, S., Lehman, R., Chandarlapaty, S., Hamilton, N., Gajria, D., Knowles, J., Shah, J., Shannon, K., Tetteh, E., Sullivan, D.M., Moreno, C., Yan, L., Han, H.S., 2013. A phase 1 study evaluating the combination of an allosteric AKT inhibitor (MK-2206) and trastuzumab in patients with HER2-positive solid tumors. Breast Cancer Res. 15, 1–10.

Hudis, C.A., 2007. Trastuzumab — Mechanism of Action and Use in Clinical Practice. N. Engl. J. Med. 357, 39–51.

Hunt, S.E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A., Armean, I.M., Trevanion, S.J., Flicek, P., Cunningham, F., 2018. Ensembl variation resources. Database (Oxford). 2018, 1–12.

Huo, D., Hu, H., Rhie, S.K., Gamazon, E.R., Cherniack, A.D., Liu, J., Yoshimatsu, T.F., Pitt, J.J., Hoadley,

K.A., Troester, M., Ru, Y., Lichtenberg, T., Sturtz, L.A., Shelley, C.S., Benz, C.C., Mills, G.B., Laird, P.W., Shriver, C.D., Perou, C.M., Olopade, O.I., 2017. Comparison of Breast Cancer Molecular Features and Survival by African and European Ancestry in The Cancer Genome Atlas. JAMA Oncol. 3, 1654–1662.

Huo, D., Ikpatt, F., Khramtsov, A., Dangou, J.M., Nanda, R., Dignam, J., Zhang, B., Grushko, T., Zhang, C., Oluwasola, O., Malaka, D., Malami, S., Odetunde, A., Adeoye, A.O., Iyare, F., Falusi, A., Perou, C.M., Olopade, O.I., 2009. Population differences in breast cancer: Survey in indigenous african women reveals over-representation of triple-negative breast cancer. J. Clin. Oncol. 27, 4515–4521.

Hvidtfeldt, U.A., Gunter, M.J., Lange, T., Chlebowski, R.T., Lane, D., Farhat, G.N., Freiberg, M.S., Keiding, N., Lee, J.S., Prentice, R., Tjønneland, A., Vitolins, M.Z., Wassertheil-Smoller, S., Strickler, H.D., Rod, N.H., 2012. Quantifying mediating effects of endogenous estrogen and insulin in the relation between obesity, alcohol consumption, and breast cancer. Cancer Epidemiol. Biomarkers Prev. 21, 1203–1212.

Hwang, K.T., Han, W., Kim, J., Moon, H.G., Oh, S., Song, Y.S., Kim, Y.A., Chang, M.S., Noh, D.Y., 2017. Prognostic influence of BCl2 on molecular subtypes of breast cancer. J. Breast Cancer 20, 54–64.

Hynes, N.E., MacDonald, G., 2009. ErbB receptors and signaling pathways in cancer. Curr. Opin. Cell Biol. 21, 177–184.

Ibrahim, E.M., Al-Foheidi, M.E., Al-Mansour, M.M., Kazkaz, G.A., 2014. The prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancer: a meta-analysis. Breast Cancer Res. Treat. 148, 467–476.

Ibrahim, S.A., Hassan, H., Vilardo, L., Kumar, S.K., Kumar, A.V., Kelsch, R., Schneider, C., Kiesel, L., Eich, H.T., Zucchi, I., Reinbold, R., Greve, B., Götte, M., 2013. Syndecan-1 (CD138) modulates triple-negative breast cancer stem cell properties via regulation of LRP-6 and IL-6-mediated STAT3 signaling. PLoS One 8, 1–14.

Ibrahim, Y.H., García-García, C., Serra, V., He, L., Torres-Lockhart, K., Prat, A., Anton, P., Cozar, P., Guzmán, M., Grueso, J., Rodríguez, O., Calvo, M.T., Aura, C., Díez, O., Rubio, I.T., Pérez, J., Rodón, J., Cortés, J., Ellisen, L.W., Scaltriti, M., Baselga, J., 2012. PI3K inhibition impairs BRCA1/2 expression and sensitizes BRCA-proficient triple-negative breast cancer to PARP inhibition. Cancer Discov. 2, 1036–1047.

Infante, J.R., Papadopoulos, K.P., Bendell, J.C., Patnaik, A., Burris, H.A., Rasco, D., Jones, S.F., Smith, L., Cox, D.S., Durante, M., Bellew, K.M., Park, J., Le, N.T., Tolcher, A.W., 2013. A phase 1b study of trametinib, an oral Mitogen-activated protein kinase kinase (MEK) inhibitor, in combination with gemcitabine in advanced solid tumours. Eur. J. Cancer 49, 2077–2085.

Iqbal, J., Ginsburg, O., Rochon, P.A., Sun, P., Narod, S.A., 2015. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. JAMA - J. Am. Med. Assoc. 313, 165–173.

Isakoff, S.J., Mayer, E.L., He, L., Traina, T.A., Carey, L.A., Krag, K.J., Rugo, H.S., Liu, M.C., Stearns, V., Come, S.E., Timms, K.M., Hartman, A.R., Borger, D.R., Finkelstein, D.M., Garber, J.E., Ryan, P.D., Winer, E.P., Goss, P.E., Ellisen, L.W., 2015. TBCRC009: A multicenter phase II clinical trial of platinum monotherapy with biomarker assessment in metastatic triple-negative breast cancer. J. Clin. Oncol. 33, 1902–1909.

Isakoff, S.J., Puhalla, S., Domchek, S.M., Friedlander, M., Kaufman, B., Robson, M., Telli, M.L., Diéras, V., Han, H.S., Garber, J.E., Johnson, E.F., Maag, D., Qin, Q., Giranda, V.L., Shepherd, S.P., 2017. A randomized Phase II study of veliparib with temozolomide or carboplatin/paclitaxel versus placebo with carboplatin/paclitaxel in BRCA1/2 metastatic breast cancer: Design and rationale. Futur. Oncol. 13, 307–320.

Ishibashi, H., Suzuki, T., Suzuki, S., Moriya, T., Kaneko, C., Takizawa, T., Sunamori, M., Handa, M., Kondo, T., Sasano, H., 2003. Sex steroid hormone receptors in human thymoma. J. Clin. Endocrinol. Metab. 88, 2309–2317.

Issa-Nummer, Y., Darb-Esfahani, S., Loibl, S., Kunz, G., Nekljudova, V., Schrader, I., Sinn, B.V., Ulmer, H.-U., Kronenwett, R., Just, M., Kühn, T., Diebold, K., Untch, M., Holms, F., Blohmer, J.-U., Habeck, J.-O., Dietel, M., Overkamp, F., Krabisch, P., von Minckwitz, G., Denkert, C., 2013. Prospective Validation of Immunological Infiltrate for Prediction of Response to Neoadjuvant Chemotherapy in HER2-Negative Breast Cancer – A Substudy of the Neoadjuvant GeparQuinto Trial. PLoS One 8, e79775.

Ivshina, A. V., George, J., Senko, O., Mow, B., Putti, T.C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H., Wong, J.E.L., Liu, E.T., Bergh, J., Kuznetsov, V.A., Miller, L.D., 2006. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res. 66, 10292–10301.

Iyengar, P., Combs, T.P., Shah, S.J., Gouon-Evans, V., Pollard, J.W., Albanese, C., Flanagan, L., Tenniswood, M.P., Guha, C., Lisanti, M.P., Pestell, R.G., Scherer, P.E., 2003. Adipocyte-secreted factors synergistically promote mammary tumorigenesis through induction of anti-apoptotic transcriptional programs and proto-oncogene stabilization. Oncogene 22, 6408–6423.

Jenkins, B.D., Martini, R.N., Hire, R., Brown, A., Bennett, B., Brown, I., Howerth, E.W., Egan, M., Hodgson, J., Yates, C., Kittles, R., Chitale, D., Ali, H., Nathanson, D., Nikolinakos, P., Newman, L., Monteil, M., Davis, M.B., 2019. Atypical chemokine receptor 1 (DARC/ACKR1) in breast tumors is associated with survival, circulating chemokines, tumor-infiltrating immune cells, and African ancestry. Cancer Epidemiol. Biomarkers Prev. 28, 690–700.

Jeong, H.M., Kim, R.N., Kwon, M.J., Oh, E., Han, J., Lee, S.K., Choi, J.S., Park, S., Nam, S.J., Gong, G.Y., Nam, J.W., Choi, D.H., Lee, H., Nam, B.H., Choi, Y. La, Shin, Y.K., 2017. Targeted exome sequencing of Korean triple-negative breast cancer reveals homozygous deletions associated with poor prognosis of adjuvant chemotherapy-treated patients. Oncotarget 8, 61538–61550.

Jézéquel, P., Kerdraon, O., Hondermarck, H., Guérin-Charbonnel, C., Lasla, H., Gouraud, W., Canon, J.L., Gombos, A., Dalenc, F., Delaloge, S., Lemonnier, J., Loussouarn, D., Verrièle, V., Campone, M., 2019. Identification of three subtypes of triple-negative breast cancer with potential therapeutic implications. Breast Cancer Res. 21, 1–14.

Jiagge, E., Chitale, D., Newman, L.A., 2018. Triple-Negative Breast Cancer, Stem Cells, and African Ancestry. Am. J. Pathol. 188, 271–279.

Jiagge, E., Jibril, A.S., Chitale, D., Bensenhaver, J.M., Awuah, B., Hoenerhoff, M., Adjei, E., Bekele, M., Abebe, E., Nathanson, S.D., Gyan, K., Salem, B., Oppong, J., Aitpillah, F., Kyei, I., Bonsu, E.O., Proctor, E., Merajver, S.D., Wicha, M., Stark, A., Newman, L.A., 2016. Comparative Analysis of Breast Cancer Phenotypes in African American, White American, and West Versus East African patients: Correlation Between African Ancestry and Triple-Negative Breast Cancer. Ann. Surg. Oncol. 23, 3843–3849.

Jiang, J.X., Riquelme, M.A., Zhou, J.Z., 2015. ATP, a double-edged sword in cancer. Oncoscience 2, 673–674.

Jiang, W., Wang, X., Zhang, C., Xue, L., Yang, L., 2020. Expression and clinical significance of MAPK and EGFR in triple–negative breast cancer. Oncol. Lett. 19, 1842–1848.

Jiang, Z., Li, X., Hu, J., Zhou, W., Jiang, Y., Li, G., Lu, D., 2006. Promoter hypermethylation-mediated down-regulation of LATS1 and LATS2 in human astrocytoma. Neurosci. Res. 56, 450–458.

John, T., Liu, G., Tsao, M.S., 2009. Overview of molecular testing in non-small-cell lung cancer: Mutational analysis, gene copy number, protein expression and other biomarkers of EGFR for the

prediction of response to tyrosine kinase inhibitors. Oncogene 28, S14–S23.

Johnson, R., Halder, G., 2014. The two faces of Hippo: Targeting the Hippo pathway for regenerative medicine and cancer treatment. Nat. Rev. Drug Discov. 13, 63–79.

Johnstone, R.W., Ruefli, A.A., Lowe, S.W., 2002. Apoptosis: A link between cancer genetics and chemotherapy. Cell 108, 153–164.

Jones, P.A., Laird, P.W., 1999. Cancer epigenetics comes of age. Nat. Genet. 21, 163–167.

Jones, S.R., Carley, S., Harrison, M., 2003. An introduction to power and sample size estimation.

Jovanovic, B., Mayer, I.A., Mayer, E.L., Abramson, V.G., Bardia, A., Sanders, M.E., Kuba, M.G., Estrada, M. V., Beeler, J.S., Shaver, T.M., Johnson, K.C., Sanchez, V., Rosenbluth, J.M., Dillon, P.M., Forero-Torres, A., Chang, J.C., Meszoely, I.M., Grau, A.M., Lehmann, B.D., Shyr, Y., Sheng, Q., Chen, S.C., Arteaga, C.L., Pietenpol, J.A., 2017. A randomized phase II neoadjuvant study of cisplatin, paclitaxel with or without everolimus in patients with stage II/III triple-negative breast cancer (TNBC): Responses and long-term outcome correlated with increased frequency of DNA damage response gene. Clin. Cancer Res. 23, 4035–4045.

Kahn, S.E., Zinman, B., Haffner, S.M., O'Neill, M.C., Kravitz, B.G., Yu, D., Freed, M.I., Herman, W.H., Holman, R.R., Jones, N.P., Lachin, J.M., Viberti, G.C., 2006. Obesity is a major determinant of the association of C-reactive protein levels and the metabolic syndrome in type 2 diabetes. Diabetes 55, 2357–2364.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K., 2017. KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45, D353–D361.

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457–D462.

Kang, C., Syed, Y.Y., 2020. Atezolizumab (in Combination with Nab-Paclitaxel): A Review in Advanced Triple-Negative Breast Cancer. Drugs 80, 601–607.

Karaayvaz, M., Cristea, S., Gillespie, S.M., Patel, A.P., Mylvaganam, R., Luo, C.C., Specht, M.C., Bernstein, B.E., Michor, F., Ellisen, L.W., 2018. Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. Nat. Commun. 9.

Karayiannakis, A.J., Nakopoulou, L., Gakiopoulou, H., Keramopoulos, A., Davaris, P.S., Pignatelli, M., 2001. Expression patterns of β-catenin in in situ and invasive breast cancer. Eur. J. Surg. Oncol. 27, 31–36.

Karpowicz, P., Perez, J., Perrimon, N., 2010. The Hippo tumor suppressor pathway regulates intestinal stem cell regeneration. Development 137, 4135–4145.

Kassambara, A., Kosinski, M., Biecek, P., Fabian, S., 2019. survminer: Drawing Survival Curves using "ggplot2."

Kaufmann, S.H., Gores, G.J., 2000. Apoptosis in cancer: Cause and cure. BioEssays 22, 1007–1017.

Kazan, J.M., El-Saghir, J., Saliba, J., Shaito, A., Jalaleddine, N., El-Hajjar, L., Al-Ghadban, S., Yehia, L., Zibara, K., El-Sabban, M., 2019. Cx43 expression correlates with breast cancer metastasis in MDA-MB-231 cells in vitro, in a mouse xenograft model and in human breast cancer tissues. Cancers (Basel). 11.

Keenan, T., Moy, B., Mroz, E.A., Ross, K., Niemierko, A., Rocco, J.W., Isakoff, S., Ellisen, L.W., Bardia, A., 2015. Comparison of the genomic landscape between primary breast cancer in African American versus white women and the association of racial differences with tumor recurrence. J. Clin.

Oncol. 33, 3621–3627.

Kennedy, R.D., Quinn, J.E., Mullan, P.B., Johnston, P.G., Harkin, D.P., 2004. The role of BRCA1 in the cellular response to chemotherapy. J. Natl. Cancer Inst. 96, 1659–1668.

Khelil, A.H., Denden, S., Leban, N., Daimi, H., Lakhdhar, R., Lefranc, G., Ben Chibani, J., Perrin, P., 2010. Hemoglobinopathies in North Africa: A review. Hemoglobin 34, 1–23.

Khosravi-Shahi, P., Cabezón-Gutiérrez, L., Custodio-Cabello, S., 2018. Metastatic triple negative breast cancer: Optimizing treatment options, new and emerging targeted therapies. Asia. Pac. J. Clin. Oncol. 14, 32–39. 8

Khramtsov, A.I., Khramtsova, G.F., Tretiakova, M., Huo, D., Olopade, O.I., Goss, K.H., 2010. Wnt/β-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. Am. J. Pathol. 176, 2911–2920. https://doi.org/10.2353/ajpath.2010.091125

Kidd, K.K., Speed, W.C., Pakstis, A.J., Furtado, M.R., Fang, R., Madbouly, A., Maiers, M., Middha, M., 2014. Forensic Science International : Genetics Progress toward an efficient panel of SNPs for ancestry inference 10, 23–32.

Kim, H.S., Moon, H.G., Han, W., Yom, C.K., Kim, W.H., Kim, J.H., Noh, D.Y., 2012. COX2 overexpression is a prognostic marker for Stage III breast cancer. Breast Cancer Res. Treat. 132, 51–59.

Kim, S.B., Maslyar, D.J., Dent, R., Im, S.A., Espié, M., Blau, S., Tan, A.R., Isakoff, S.J., Oliveira, M., Saura, C., Wongchenko, M.J., Kapp, A. V., Chan, W.Y., Singel, S.M., Baselga, J., Kim, S.B., Lee, K.S., Im, S.A., Espié, M., Wang, H.C., Blau, S., Dent, R., Tan, A., Sohn, J.H., De Laurentiis, M., Estevez, L.G., Huang, C.S., Romieu, G., Velez, M., Villanueva, R., Conte, P.F., Dakhil, S., Debled, M., Martin, A.G., Hurvitz, S., Kim, J.H., Levy, C., Oliveira, M., Rovira, P.S., Seo, J.H., Valero, V., Vidal, G., Wong, A., Allison, M.A.K., Figlin, R., Chan, D., Chen, S.C., Chen, Y.H., Cobleigh, M., De Braud, F., Dirix, L., Hansen, V., Bessard, A.H., Iannotti, N., Isakoff, S., Lawler, W., Montaño, A., Salkini, M., Seigel, L., 2017. Ipatasertib plus paclitaxel versus placebo plus paclitaxel as first-line therapy for metastatic triple-negative breast cancer (LOTUS): a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. Lancet Oncol. 18, 1360–1372.

Kim, T., Yang, S.J., Hwang, D., Song, J., Kim, M., Kyum Kim, S., Kang, K., Ahn, J., Lee, D., Kim, M.Y., Kim, S., Seung Koo, J., Seok Koh, S., Kim, S.Y., Lim, D.S., 2015. A basal-like breast cancer-specific role for SRF-IL6 in YAP-induced cancer stemness. Nat. Commun. 6.

Kimbung, S., Biskup, E., Johansson, I., Aaltonen, K., Ottosson-Wadlund, A., Gruvberger-Saal, S., Cunliffe, H., Fadeel, B., Loman, N., Berglund, P., Hedenfalk, I., 2012. Co-targeting of the PI3K pathway improves the response of BRCA1 deficient breast cancer cells to PARP1 inhibition. Cancer Lett. 319, 232–241.

King, M.C., Marks, J.H., Mandell, J.B., 2003. Breast and Ovarian Cancer Risks Due to Inherited Mutations in BRCA1 and BRCA2. Science (80-.). 302, 643–646.

Knudson, A.G., 1971. Mutation and Cancer: Statistical Study of Retinoblastoma. Proc. Nat. Acad. Sci. USA 68, 820–823.

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., Wilson, R.K., Ally, A., Balasundaram, M., Butterfield, Y.S.N., Carlsen, R., Carter, C., Chu, Andy, Chuah, E., Chun, H.J.E., Coope, R.J.N., Dhalla, N., Guin, R., Hirst, C., Hirst, M., Holt, R.A., Lee, D., Li, H.I., Mayo, M., Moore, R.A., Mungall, A.J., Pleasance, E., Robertson, A.G., Schein, J.E., Shafiei, A., Sipahimalani, P., Slobodan, J.R., Stoll, D., Tam, A., Thiessen, N., Varhol, R.J., Wye, N., Zeng, T., Zhao, Y., Birol, I., Jones, S.J.M., Marra, M.A., Cherniack, A.D., Saksena, G., Onofrio, R.C., Pho, N.H., Carter, S.L., Schumacher, S.E., Tabak, B., Hernandez, B., Gentry, J., Nguyen, H., Crenshaw, A., Ardlie, K., Beroukhim, R., Winckler, W., Getz, G., Gabriel, S.B., Meyerson, M., Chin, L., Kucherlapati, R., Hoadley, K.A., Auman, J.T., Fan, C.,

Turman, Y.J., Shi, Y., Li, L., Topal, M.D., He, X., Chao, H.H., Prat, A., Silva, G.O., Iglesia, M.D., Zhao, W., Usary, J., Berg, J.S., Adams, M., Booker, J., Wu, J., Gulabani, A., Bodenheimer, T., Hoyle, A.P., Simons, J. V., Soloway, M.G., Mose, L.E., Jefferys, S.R., Balu, S., Parker, J.S., Hayes, D.N., Perou, C.M., Malik, S., Mahurkar, S., Shen, H., Weisenberger, D.J., Triche, T., Lai, P.H., Bootwalla, M.S., Maglinte, D.T., Berman, B.P., Van Den Berg, D.J., Baylin, S.B., Laird, P.W., Creighton, C.J., Donehower, L.A., Noble, M., Voet, D., Gehlenborg, N., Di Cara, D., Zhang, Juinhua, Zhang, H., Wu, C.J., Yingchun Liu, S., Lawrence, M.S., Zou, L., Sivachenko, A., Lin, P., Stojanov, P., Jing, R., Cho, J., Sinha, Raktim, Park, R.W., Nazaire, M.D., Robinson, J., Thorvaldsdottir, H., Mesirov, J., Park, P.J., Reynolds, S., Kreisberg, R.B., Bernard, B., Bressler, R., Erkkila, T., Lin, J., Thorsson, V., Zhang, W., Shmulevich, I., Ciriello, G., Weinhold, N., Schultz, N., Gao, J., Cerami, E., Gross, B., Jacobsen, A., Sinha, Rileen, Aksoy, B.A., Antipin, Y., Reva, B., Shen, R., Taylor, B.S., Ladanyi, M., Sander, C., Anur, P., Spellman, P.T., Lu, Y., Liu, W., Verhaak, R.R.G., Mills, G.B., Akbani, R., Zhang, N., Broom, B.M., Casasent, T.D., Wakefield, C., Unruh, A.K., Baggerly, K., Coombes, K., Weinstein, J.N., Haussler, D., Benz, C.C., Stuart, J.M., Benz, S.C., Zhu, J., Szeto, C.C., Scott, G.K., Yau, C., Paull, E.O., Carlin, D., Wong, C., Sokolov, A., Thusberg, J., Mooney, S., Ng, S., Goldstein, T.C., Ellrott, K., Grifford, M., Wilks, C., Ma, S., Craft, B., Yan, C., Hu, Y., Meerzaman, D., Gastier-Foster, J.M., Bowen, J., Ramirez, N.C., Black, A.D., Pyatt, R.E., White, P., Zmuda, E.J., Frick, J., Lichtenberg, T.M., Brookens, R., George, M.M., Gerken, M.A., Harper, H.A., Leraas, K.M., Wise, L.J., Tabler, T.R., McAllister, C., Barr, T., Hart-Kothari, M., Tarvin, K., Saller, C., Sandusky, G., Mitchell, C., Iacocca, M. V., Brown, J., Rabeno, B., Czerwinski, C., Petrelli, N., Dolzhansky, O., Abramov, M., Voronina, O., Potapova, O., Marks, J.R., Suchorska, W.M., Murawa, D., Kycler, W., Ibbs, M., Korski, K., Spychała, A., Murawa, P., Brzeziński, J.J., Perz, H., Łaźniak, R., Teresiak, M., Tatka, H., Leporowska, E., Bogusz-Czerniewicz, M., Malicki, J., Mackiewicz, A., Wiznerowicz, M., Van Le, X., Kohl, B., Viet Tien, N., Thorp, R., Van Bang, N., Sussman, H., Phu, B.D., Hajek, R., Hung, N.P., Phuong, T.V.T., Thang, H.Q., Khan, K.Z., Penny, R., Mallery, D., Curley, E., Shelton, C., Yena, P., Ingle, J.N., Couch, F.J., Lingle, W.L., King, T.A., Gonzalez-Angulo, A.M., Dyer, M.D., Liu, S., Meng, X., Patangan, M., Waldman, F., Stöppler, H., Rathmell, W.K., Thorne, L., Huang, M., Boice, L., Hill, A., Morrison, C., Gaudioso, C., Bshara, W., Daily, K., Egea, S.C., Pegram, M.D., Gomez-Fernandez, C., Dhir, R., Bhargava, R., Brufsky, A., Shriver, C.D., Hooke, J.A., Campbell, J.L., Mural, R.J., Hu, H., Somiari, S., Larson, C., Deyarmin, B., Kvecher, L., Kovatich, A.J., Ellis, M.J., Stricker, T., White, K., Olopade, O., Luo, C., Chen, Y., Bose, R., Chang, L.W., Beck, A.H., Pihl, T., Jensen, M., Sfeir, R., Kahn, A., Chu, Anna, Kothiyal, P., Wang, Z., Snyder, E., Pontius, J., Ayala, B., Backus, M., Walton, J., Baboud, J., Berton, D., Nicholls, M., Srinivasan, D., Raman, R., Girshik, S., Kigonya, P., Alonso, S., Sanbhadti, R., Barletta, S., Pot, D., Sheth, M., Demchok, J.A., Shaw, K.R.M., Yang, L., Eley, G., Ferguson, M.L., Tarnuzzer, R.W., Zhang, Jiashan, Dillon, L.A.L., Buetow, K., Fielding, P., Ozenberger, B.A., Guyer, M.S., Sofia, H.J., Palchik, J.D., 2012. Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70.

Korde, L.A., Zujewski, J.A., Kamin, L., Giordano, S., Domchek, S., Anderson, W.F., Bartlett, J.M.S., Gelmon, K., Nahleh, Z., Bergh, J., Cutuli, B., Pruneri, G., McCaskill-Stevens, W., Gralow, J., Hortobagyi, G., Cardoso, F., 2010. Multidisciplinary meeting on male breast cancer: Summary and research recommendations. J. Clin. Oncol. 28, 2114–2122.

Koundouros, N., Poulogiannis, G., 2020. Reprogramming of fatty acid metabolism in cancer. Br. J. Cancer 122, 4–22.

Kreike, B., Hart, G., Bartelink, H., Van De Vijver, M.J., 2010. Analysis of breast cancer related gene expression using natural splines and the Cox proportional hazard model to identify prognostic associations. Breast Cancer Res. Treat. 122, 711–720.

Krishnamurthy, N., Kurzrock, R., 2018. Targeting the Wnt/beta-catenin pathway in cancer: Update on effectors and inhibitors. Cancer Treat. Rev. 62, 50–60.

Kroenke, C.H., Sweeney, C., Kwan, M.L., Quesenberry, C.P., Weltzien, E.K., Habel, L.A., Castillo, A.,

Bernard, P.S., Factor, R.E., Kushi, L.H., Caan, B.J., 2014. Race and breast cancer survival by intrinsic subtype based on PAM50 gene expression. Breast Cancer Res. Treat. 144, 689–699.

Kumar, P., Aggarwal, R., 2016. An overview of triple-negative breast cancer. Arch. Gynecol. Obstet. 293, 247–269.

Kummar, S., Wade, J.L., Oza, A.M., Sullivan, D., Chen, A.P., Gandara, D.R., Ji, J., Kinders, R.J., Wang, L., Allen, D., Coyne, G.O., Steinberg, S.M., Doroshow, J.H., 2016. Randomized phase II trial of cyclophosphamide and the oral poly (ADP-ribose) polymerase inhibitor veliparib in patients with recurrent, advanced triple-negative breast cancer. Invest. New Drugs 34, 355–363.

Kwa, M.J., Adams, S., 2018. Checkpoint inhibitors in triple-negative breast cancer (TNBC): Where to go from here. Cancer 124, 2086–2103.

Kwan, M.L., Kushi, L.H., Weltzien, E., Maring, B., Kutner, S.E., Fulton, R.S., Lee, M.M., Ambrosone, C.B., Caan, B.J., 2009. Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. Breast Cancer Res. 11, 1–13.

Kwong, A., Wong, K., Wong, C., Leung, C., 2009. The Epidemiology, Clinico-Pathologic and Prognostic Characteristics of Triple-Negative Breast Cancer Compared with Non Triple-Negative Breast Cancer. Cancer Res 69, 3071.

Lan, Q., Fang, Y., Mei, S., Xie, T., Liu, Y., Jin, X., Yang, G., Zhu, B., 2020. Next generation sequencing of a set of ancestry - informative SNPs : ancestry assignment of three continental populations and estimating ancestry composition for Mongolians. Mol. Genet. Genomics.

Larsson, S.C., Mantzoros, C.S., Wolk, A., 2007. Diabetes mellitus and risk of breast cancer: A meta-analysis. Int. J. Cancer 121, 856–862.

Laurenti, E., Wilson, A., Trumpp, A., 2009. Myc's other life: stem cells and beyond. Curr. Opin. Cell Biol. 21, 844–854.

Lee, E., McKean-Cowdin, R., Ma, H., Spicer, D. V., Van Den Berg, D., Bernstein, L., Ursin, G., 2011. Characteristics of triple-negative breast cancer in patients with a BRCA1 mutation: Results from a population-based study of young women. J. Clin. Oncol. 29, 4373–4380.

Lee, H.H., Wang, Y.N., Hung, M.C., 2015. Non-canonical signaling mode of the epidermal growth factor receptor family. Am. J. Cancer Res. 5, 2944–2958.

Lee, J.S.H., Kibbe, W.A., Grossman, R.L., 2018. Data Harmonization for a Molecularly Driven Health System. Cell 174, 1045–1048.

Lee, K. min, Giltnane, J.M., Balko, J.M., Schwarz, L.J., Guerrero-Zotano, A.L., Hutchinson, K.E., Nixon, M.J., Estrada, M. V., Sánchez, V., Sanders, M.E., Lee, T., Gómez, H., Lluch, A., Pérez-Fidalgo, J.A., Wolf, M.M., Andrejeva, G., Rathmell, J.C., Fesik, S.W., Arteaga, C.L., 2017. MYC and MCL1 Cooperatively Promote Chemotherapy-Resistant Breast Cancer Stem Cells via Regulation of Mitochondrial Oxidative Phosphorylation. Cell Metab. 26, 633-647.e7.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., Pietenpol, J.A., 2011. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. 121, 2750–2767.

Lehmann, B.D., Jovanović, B., Chen, X., Estrada, M. V., Johnson, K.N., Shyr, Y., Moses, H.L., Sanders, M.E., Pietenpol, J.A., 2016. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. PLoS One 11, e0157368.

Lehmann, B.D., Pietenpol, J.A., 2014. Identification and use of biomarkers in treatment strategies for triple negative breast cancer subtypes. J Pathol. 232, 142–150.

Lei, Q.-Y., Zhang, H., Zhao, B., Zha, Z.-Y., Bai, F., Pei, X.-H., Zhao, S., Xiong, Y., Guan, K.-L., 2008. TAZ

Promotes Cell Proliferation and Epithelial-Mesenchymal Transition and Is Inhibited by the Hippo Pathway. Mol. Cell. Biol. 28, 2426–2436.

Leithe, E., Sirnes, S., Omori, Y., Rivedal, E., 2006. Downregulation of gap junctions in cancer cells. Crit. Rev. Oncog. 12, 225–256.

Leon-Ferre, R.A., Polley, M.Y., Liu, H., Gilbert, J.A., Cafourek, V., Hillman, D.W., Elkhanany, A., Akinhanmi, M., Lilyquist, J., Thomas, A., Negron, V., Boughey, J.C., Liu, M.C., Ingle, J.N., Kalari, K.R., Couch, F.J., Visscher, D.W., Goetz, M.P., 2018. Impact of histopathology, tumor-infiltrating lymphocytes, and adjuvant chemotherapy on prognosis of triple-negative breast cancer. Breast Cancer Res. Treat. 167, 89–99.

Leung, L.K., Wang, T.T.Y., 1999. Paradoxical regulation of Bcl-2 family proteins by 17β-oestradiol in human breast cancer cells MCF-7. Br. J. Cancer 81, 387–392.

Li, D., Bi, F.F., Cao, J.M., Cao, C., Li, C.Y., Yang, Q., 2013. Effect of BRCA1 on epidermal growth factor receptor in ovarian cancer. J. Exp. Clin. Cancer Res. 32, 1–6.

Li, X., Li, M., Lian, Z., Zhu, H., Kong, L., Wang, P., Yu, J., 2016. Prognostic Role of Programmed Death Ligand-1 Expression in Breast Cancer: A Systematic Review and Meta-Analysis. Target. Oncol. 11, 753–761.

Li, X., Wu, D., Cooper, N.G.F., Rai, S.N., 2019. Sample size calculations for the differential expression analysis of RNA-seq data using a negative binomial regression model. Stat. Appl. Genet. Mol. Biol. 20180021, 1–17.

Li, Y.W., Xu, J., Zhu, G.Y., Huang, Z.J., Lu, Y., Li, X.Q., Wang, N., Zhang, F.X., 2018. Apigenin suppresses the stem cell-like properties of triple-negative breast cancer cells by inhibiting YAP/TAZ activity. Cell Death Discov. 4.

Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M.A., Goldstein, L.S.B., Abujarour, R., Ding, S., Guan, K.L., 2010. The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. Genes Dev. 24, 1106–1118.

Liang, P., Pardee, A.B., 2003. Analysing differential gene expression in cancer. Nat. Rev. Cancer 3, 869–876.

Liedtke, C., Mazouni, C., Hess, K.R., André, F., Tordai, A., Mejia, J.A., Symmans, W.F., Gonzalez-Angulo, A.M., Hennessy, B., Green, M., Cristofanilli, M., Hortobagyi, G.N., Pusztai, L., 2008. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J. Clin. Oncol. 26, 1275–1281.

Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A., Feleppa, F., Huschtscha, L.I., Thorne, H.J., Fox, S.B., Yan, M., French, J.D., Brown, M.A., Smyth, G.K., Visvader, J.E., Lindeman, G.J., 2009. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat. Med. 15, 907–913.

Lima, R.T., Martins, L.M., Guimarães, J.E., Sambade, C., Vasconcelos, M.H., 2004. Specific downregulation of bcl-2 and xIAP by RNAi enhances the effects of chemotherapeutic agents in MCF-7 human breast cancer cells. Cancer Gene Ther. 11, 309–316.

Lin, C.Y., Barry-Holson, K.Q., Allison, K.H., 2016. Breast cancer stem cells: Are we ready to go from bench to bedside? Histopathology 68, 119–137.

Lin, S.Y., Xia, W., Wang, J.C., Kwong, K.Y., Spohn, B., Wen, Y., Pestell, R.G., Hung, M.C., 2000. β-catenin, a novel prognostic marker for breast cancer: Its roles in cyclin D1 expression and cancer progression. Proc. Natl. Acad. Sci. U. S. A. 97, 4262–4266.

Lindner, R., Sullivan, C., Offor, O., Lezon-Geyda, K., Halligan, K., Fischbach, N., Shah, M., Bossuyt, V.,

Schulz, V., Tuck, D.P., Harris, L.N., 2013. Molecular phenotypes in triple negative breast cancer from African American patients suggest targets for therapy. PLoS One 8.

Litton, J.K., Blum, J.L., Im, Y.-H., Martin, M., Mina, L.A., Roche, H.H., 2015. A phase 3, open-label, randomized, parallel, 2-arm international study of the oral PARP inhibitor talazoparib (BMN 673) in BRCA mutation subjects with locally advanced and/or metastatic breast cancer (EMBRACA). J. Clin. Oncol. 33, TPS1107.

Liu, F., Lang, R., Zhao, J., Zhang, X., Pringle, G.A., Fan, Y., Yin, D., Gu, F., Yao, Z., Fu, L., 2011. CD8+ cytotoxic T cell and FOXP3+ regulatory T cell infiltration in relation to breast cancer survival and molecular subtypes. Breast Cancer Res. Treat. 130, 645–655.

Liu, J., Ye, L., Li, Q., Wu, X., Wang, B., Ouyang, Y., Yuan, Z., Li, J., Lin, C., 2018. Synaptopodin-2 suppresses metastasis of triple-negative breast cancer via inhibition of YAP/TAZ activity. J. Pathol. 244, 71–83.

Liu, S., Chia, S.K., Mehl, E., Leung, S., Rajput, A., Cheang, M.C.U., Nielsen, T.O., 2010. Progesterone receptor is a significant factor associated with clinical outcomes and effect of adjuvant tamoxifen therapy in breast cancer patients. Breast Cancer Res Treat 119, 53–61.

Liu, S., Wicha, M.S., 2010. Targeting breast cancer stem cells. J. Clin. Oncol. 28, 4006–4012.

Liu, Y., Lv, D. lai, Duan, J. jie, Xu, S. lin, Zhang, J. fang, Yang, X. jun, Zhang, X., Cui, Y. hong, Bian, X. wu, Yu, S. cang, 2014. ALDH1A1 expression correlates with clinicopathologic features and poor prognosis of breast cancer patients: A systematic review and meta-analysis. BMC Cancer 14, 1–11. 4

Loi, S., Drubay, D., Adams, S., Pruneri, G., Francis, P.A., Lacroix-Triki, M., Joensuu, H., Dieci, M.V., Badve, S., Demaria, S., Gray, R., Munzone, E., Lemonnier, J., Sotiriou, C., Piccart, M.J., Kellokumpu-Lehtinen, P.L., Vingiani, A., Gray, K., Andre, F., Denkert, C., Salgado, R., Michiels, S., 2019. Tumor-infiltrating lymphocytes and prognosis: A pooled individual patient analysis of early-stage triple-negative breast cancers. J. Clin. Oncol. 37, 559–569.

Loi, S., Dushyanthen, S., Beavis, P.A., Salgado, R., Savas, P., Combs, S., Rimm, D.L., Giltnane, J.M., Monica, V., Sanders, M.E., Cook, R.S., Pilkinton, M.A., Mallal, S.A., Wang, K., Miller, V.A., Stephens, P.J., Yelensky, R., Doimi, F.D., Ryzhov, S. V, Darcy, P.K., Carlos, L., Balko, J.M., Centre, M.C., Translational, C., Haven, N., Program, C., 2016. RAS/MAPK activation is associated with reduced tumor- infiltrating lymphocytes in triple-negative breast cancer: therapeutic cooperation between MEK and PD-1/PD-L1 immune checkpoint inhibitors. Clin Cancer Res 22, 1499–1509.

Loi, S., Michiels, S., Salgado, R., Sirtaine, N., Jose, V., Fumagalli, D., Kellokumpu-Lehtinen, P.L., Bono, P., Kataja, V., Desmedt, C., Piccart, M.J., Loibl, S., Denkert, C., Smyth, M.J., Joensuu, H., Sotiriou, C., 2014. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: Results from the FinHER trial. Ann. Oncol. 25, 1544–1550.

Loi, S., Sirtaine, N., Piette, F., Salgado, R., Viale, G., Van Eenoo, F., Rouas, G., Francis, P., Crown, J.P.A., Hitre, E., De Azambuja, E., Quinaux, E., Di Leo, A., Michiels, S., Piccart, M.J., Sotiriou, C., 2013. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. J. Clin. Oncol. 31, 860–867.

Lopes, N., Carvalho, J., Durães, C., Sousa, B., Gomes, M., Costa, J.L., Oliveira, C., Paredes, J., Schmitt, F., 2012. 1Alpha,25-dihydroxyvitamin D 3 induces de novo E-cadherin expression in triple-negative breast cancer cells by CDH1-promoter demethylation. Anticancer Res. 32, 249–257.

López-Knowles, E., Zardawi, S.J., McNeil, C.M., Millar, E.K.A., Crea, P., Musgrove, E.A., Sutherland, R.L., O'Toole, S.A., 2010. Cytoplasmic localization of â-catenin is a marker of poor outcome in breast

cancer patients. Cancer Epidemiol. Biomarkers Prev. 19, 301–309.

Lumachi, F., Brunello, A., Maruzzo, M., Basso, U., Basso, S.M.M., 2013. Treatment of Estrogen Receptor-Positive Breast Cancer. Curr. Med. Chem. 20, 596–604.

Lund, M.J., Trivers, K.F., Porter, P.L., Coates, R.J., Leyland-Jones, B., Brawley, O.W., Flagg, E.W., O'Regan, R.M., Gabram, S.G.A., Eley, J.W., 2009. Race and triple negative threats to breast cancer survival: A population-based study in Atlanta, GA. Breast Cancer Res. Treat. 113, 357–370.

Luo, W., Friedman, M.S., Shedden, K., Hankenson, K.D., Woolf, P.J., 2009. GAGE: Generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 10, 1–17.

Ly, M., Antoine, M., Dembl, A.K., Levy, P., Rodenas, A., Tour, B.A., Badiaga, Y., Dembl, B.K., Bagayogo, D.C., Diallo, Y.L., Kon, A.A., Callard, P., Bernaudin, J.F., Diallo, D.A., 2012. High incidence of triple-negative tumors in sub-saharan africa: A prospective study of breast cancer characteristics and risk factors in malian women seen in a bamako university hospital. Oncol. 83, 257–263.

Lyons, T.R., O'Brien, J., Borges, V.F., Conklin, M.W., Keely, P.J., Eliceiri, K.W., Marusyk, A., Tan, A.C., Schedin, P., 2011. Postpartum mammary gland involution drives progression of ductal carcinoma in situ through collagen and COX-2. Nat. Med. 17, 1109–1116.

Marotti, J.D., de Abreu, F.B., Wells, W.A., Tsongalis, G.J., 2017. Triple-Negative Breast Cancer: Next-Generation Sequencing for Target Identification. Am. J. Pathol. 187, 2133–2138.

Marra, A., Viale, G., Curigliano, G., 2019. Recent advances in triple negative breast cancer: The immunotherapy era. BMC Med. 17, 1–9.

Marti, P., Stein, C., Blumer, T., Abraham, Y., Dill, M.T., Pikiolek, M., Orsini, V., Jurisic, G., Megel, P., Makowska, Z., Agarinis, C., Tornillo, L., Bouwmeester, T., Ruffner, H., Bauer, A., Parker, C.N., Schmelzle, T., Terracciano, L.M., Heim, M.H., Tchorz, J.S., 2015. YAP promotes proliferation, chemoresistance, and angiogenesis in human cholangiocarcinoma through TEAD transcription factors. Hepatology 62, 1497–1510.

Martin, D.N., Boersma, B.J., Yi, M., Reimers, M., Howe, T.M., Yfantis, H.G., Tsai, Y.C., Williams, E.H., Lee, D.H., Stephens, R.M., Weissman, A.M., Ambs, S., 2009. Differences in the tumor microenvironment between African-American and European-American breast cancer patients. PLoS One 4.

Martin, T.A., Jiang, W.G., 2009. Loss of tight junction barrier function and its role in cancer metastasis. Biochim. Biophys. Acta - Biomembr. 1788, 872–891.

Massihnia, D., Galvano, A., Fanale, D., Perez, A., Castiglia, M., Incorvaia, L., Listì, A., Rizzo, S., Cicero, G., Bazan, V., Castorina, S., Russo, A., 2016. Triple negative breast cancer: Shedding light onto the role of pi3k/akt/mtor pathway. Oncotarget 7, 60712–60722.

Masuda, H., Baggerly, K.A., Wang, Y., Zhang, Y., Gonzalez-Angulo, A.M., Meric-Bernstam, F., Valero, V., Lehmann, B.D., Pietenpol, J.A., Hortobagyi, G.N., Symmans, W.F., Ueno, N.T., 2013. Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. Clin. Cancer Res. 19, 5533–5540.

Masuda, H., Masuda, N., 2016. The challenge to overcome triple-negative breast cancer heterogeneity, in: Personalized Treatment of Breast Cancer.

Masuda, N., Lee, S.J., Ohtani, S., Im, Y.H., Lee, E.S., Yokota, I., Kuroi, K., Im, S.A., Park, B.W., Kim, S.B., Yanagita, Y., Ohno, S., Takao, S., Aogi, K., Iwata, H., Jeong, J., Kim, A., Park, K.H., Sasano, H., Ohashi, Y., Toi, M., 2017. Adjuvant capecitabine for breast cancer after preoperative chemotherapy. N. Engl. J. Med. 376, 2147–2159.

Matejcic, M., Li, D.P., Prescott, N.J., Lewis, C.M., Mathew, C.G., Parker, M.I., 2011. Association of a deletion of GSTT2B with an altered risk of oesophageal squamous cell carcinoma in a South

African population: A Case-Control study. PLoS One 6.

Matter, K., Balda, M.S., 2007. Epithelial tight junctions, gene expression and nucleo-junctional interplay. J. Cell Sci. 120, 1505–1511.

Mavaddat, N., Barrowdale, D., Andrulis, I.L., Domchek, S.M., Eccles, D., Nevanlinna, H., Ramus, S.J., Spurdle, A., Robson, M., Sherman, M., Marie, A., Couch, F.J., Engel, C., Mcguffog, L., Healey, S., Sinilnikova, O.M., Southey, M.C., Terry, M.B., Goldgar, D., Malley, F.O., John, E.M., Janavicius, R., Tihomirova, L., Hansen, T.V.O., Nielsen, F.C., Osorio, A., Walker, L., Rogers, M.T., Side, L.E., Houghton, C., Weaver, J., Stoppa-lyonnet, D., Steele, L., Neuhausen, S.L., Nussbaum, R., Beattie, M., Odunsi, K., Sucheston, L., Gayther, S.A., Nathanson, K., Gross, J., Walsh, C., Karlan, B., 2012. Pathology of Breast and Ovarian Cancers among BRCA1 and BRCA2 Mutation Carriers : Results from the Consortium of Investigators of Modi fi ers of BRCA1 / 2 (CIMBA). Cancer Epidemiol Biomarkers Prev 21, 134–148.

McCarthy, D.J., Chen, Y., Smyth, G.K., 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 40, 4288–4297.

Mccarthy, D.J., Smyth, G.K., 2009. Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics 25, 765–771.

McCormack, V.A., Dos Santos Silva, I., 2006. Breast density and parenchymal patterns as markers of breast cancer risk: A meta-analysis. Cancer Epidemiol. Biomarkers Prev. 15, 1159–1169.

McWhorter, W.P., Mayer, W.J., 1987. Black/white differences in type of initial breast cancer treatment and implications for survival. Am. J. Public Health 77, 1515–1517.

Meng, W., Takeichi, M., 2009. Adherens junction: molecular architecture and regulation. Cold Spring Harb. Perspect. Biol 1:a002899.

Miller, K., Tong, Y., Jones, D.R., Walsh, T., Danso, M.A., Ma, C.X., 2015. Cisplatin with or without rucaparib after preoperative chemotherapy in patients with triple negative breast cancer: Final efficacy results of Hoosier Oncology Group BRE09-146. J. Clin. Oncol. 33, 1082.

Millikan, R.C., Newman, B., Tse, C.K., Moorman, P.G., Conway, K., Smith, L. V., Labbok, M.H., Geradts, J., Bensen, J.T., Jackson, S., Nyante, S., Livasy, C., Carey, L., Earp, H.S., Perou, C.M., 2008. Epidemiology of basal-like breast cancer. Breast Cancer Res. Treat. 109, 123–139.

Mittendorf, E.A., Philips, A. V., Meric-Bernstam, F., Qiao, N., Wu, Y., Harrington, S., Su, X., Wang, Y., Gonzalez-Angulo, A.M., Akcakanat, A., Chawla, A., Curran, M., Hwu, P., Sharma, P., Litton, J.K., Molldrem, J.J., Alatrash, G., 2014. PD-L1 expression in triple-negative breast cancer. Cancer Immunol. Res. 2, 361–370.

Molinero, L., Li, Y., Chang, C.W., Maund, S., Berg, M., Harrison, J., Fassò, M., O'Hear, C., Hegde, P., Emens, L.A., 2019. Tumor immune microenvironment and genomic evolution in a patient with metastatic triple negative breast cancer and a complete response to atezolizumab. J. Immunother. Cancer 7, 1–9.

Moran, M.S., 2014. Should Triple-Negative Breast Cancer (TNBC) Subtype Affect Local-Regional Therapy Decision Making? Am. Soc. Clin. Oncol. Educ. B. e32–e36.

Mori, H., Kubo, M., Yamaguchi, R., Nishimura, R., Osako, T., Arima, N., Okumura, Y., Okido, M., Yamada, M., Kai, M., Kishimoto, J., Oda, Y., Nakamura, M., 2017. The combination of PD-L1 expression and decreased tumorinfiltrating lymphocytes is associated with a poor prognosis in triple-negative breast cancer. Oncotarget 8, 15584–15592.

Morris, G.J., Naidu, S., Topham, A.K., Guiles, F., Xu, Y., McCue, P., Schwartz, G.F., Park, P.K., Rosenberg, A.L., Brill, K., Mitchell, E.P., 2007. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: A single-institution compilation compared

with the national cancer institute's surveillance, epidemiology, and end results database. Cancer 110, 876–884.

Mounir, M., Lucchetta, M., Silva, T.C., Olsen, C., Bontempi, G., Chen, X., Noushmehr, H., Colaprico, A., Papaleo, E., 2019. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEX. PLoS Comput. Biol. 15.

Moynahan, M.E., Chiu, J.W., Koller, B.H., Jasint, M., 1999. Brca1 controls homology-directed DNA repair. Mol. Cell 4, 511–518.

Mremi, A., Broadwater, G., Jackson, K., Amsi, P., Mbulwa, C., Hyslop, T., Ong, C., Hall, A., 2019. Breast cancer in Tanzanian, black American, and white American women: An assessment of prognostic and predictive features, including tumor infiltrating lymphocytes. PLoS One 14, 1–12.

Nabholtz, J.M., Chalabi, N., Radosevic-Robin, N., Dauplat, M.M., Mouret-Reynier, M.A., Van Praagh, I., Servent, V., Jacquin, J., Benmammar, K.E., Kullab, S., Bahadoor, M.R.K., Kwiatkowski, F., Cayre, A., Abrial, C., Durando, X., Bignon, Y.J., Chollet, P., Penault-Llorca, F., 2016. Multicentric neoadjuvant pilot Phase II study of cetuximab combined with docetaxel in operable triple negative breast cancer. Int. J. Cancer 138, 2274–2280.

Nachimuthu Natarajan, Takuma Nemoto, Nettlin, C., Gerald P. Murphy, 1985. Race-Related Differences in Breast Cancer Patients. Results of the 1982 national survey of breast cancer by the American College of Surgeons. Cancer 56, 1704–1709.

Nagata, Y., Lan, K.H., Zhou, X., Tan, M., Esteva, F.J., Sahin, A.A., Klos, K.S., Li, P., Monia, B.P., Nguyen, N.T., Hortobagyi, G.N., Hung, M.C., Yu, D., 2004. PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. Cancer Cell 6, 117–127.

Nakai, K., Hung, M.-C., Yamaguchi, H., 2016. A perspective on anti-EGFR therapies targeting triple-negative breast cancer. Am. J. Cancer Res. 6, 1609–23.

Nakajima, H., Ishikawa, Y., Furuya, M., Sano, T., Ohno, Y., Horiguchi, J., Oyama, T., 2014. Protein expression, gene amplification, and mutational analysis of EGFR in triple-negative breast cancer. Breast Cancer 21, 66–74.

Nakopoulou, Gakiopoulou, Keramopoulos, Giannopoulou, Athanassiadou, Mavrommatis, Davaris, 2000. C-Met Tyrosine Kinase Receptor Expression Is Associated With Abnormal B-Catenin Expression and Favourable Prognostic Factors in Invasive Breast Carcinoma. Histopathology 36, 313–325.

Nakshatri, H., Anjanappa, M., Bhat-Nakshatri, P., 2015. Ethnicity-dependent and -independent heterogeneity in healthy normal breast hierarchy impacts tumor characterization. Sci. Rep. 5, 1–14.

Nalwoga, H., Arnes, J.B., Wabinga, H., Akslen, L.A., 2010. Expression of aldehyde dehydrogenase 1 (ALDH1) is associated with basal-like markers and features of aggressive tumours in African breast cancer. Br. J. Cancer 102, 369–375.

Nanda, R., Schumm, L.P., Cummings, S., Fackenthal, J.D., Sveen, L., Ademuyiwa, F., Cobleigh, M., Esserman, L., Lindor, N.M., Neuhausen, S.L., Olopade, O.I., 2005. Genetic Testing in an Ethnically Diverse Cohort of High-Risk Women - A Comparative Analysis of BRCA1 and BRCA2 Mutations in American Families of European and African Ancestry. Jama 294, 1925–1933.

Negrini, S., Gorgoulis, V.G., Halazonetis, T.D., 2010. Genomic instability — an evolving hallmark of cancer 11, 220–228.

Nelder, J.A., Wedderburn, R.W.., 1972. Generalized linear models. J. R. Stat. Soc. Ser. A General, 370–384.

Newman, L.A., 2017. Breast Cancer Disparities: Socioeconomic Factors versus Biology. Ann. Surg. Oncol. 24, 2869–2875.

Newman, L.A., Griffith, K.A., Jatoi, I., Simon, M.S., Crowe, J.P., Colditz, G.A., 2006. Meta-analysis of survival in African American and white American patients with breast cancer: Ethnicity compared with socioeconomic status. J. Clin. Oncol. 24, 1342–1349.

Newman, L.A., Jenkins, B., Chen, Y., Oppong, J.K., Adjei, E., Jibril, A.S., Hoda, S., Cheng, E., Chitale, D., Bensenhaver, J.M., Awuah, B., Bekele, M., Abebe, E., Kyei, I., Aitpillah, F., Adinku, M., Nathanson, S.D., Jackson, L., Jiagge, E., Merajver, S., Petersen, L.F., Proctor, E., Gyan, K.K., Martini, R., Kittles, R., Davis, M.B., 2019. Hereditary Susceptibility for Triple Negative Breast Cancer Associated with Western Sub-Saharan African Ancestry: Results from an International Surgical Breast Cancer Collaborative. Ann. Surg. 270, 484–492.

Newman, L.A., Kaljee, L.M., 2017. Health disparities and triple-negative breast cancer in african American women a review. JAMA Surg. 152, 485–493.

Nielsen, F.C., Van Overeem Hansen, T., Sørensen, C.S., 2016. Hereditary breast and ovarian cancer: New genes in confined pathways. Nat. Rev. Cancer 16, 599–612.

Nielsen, T.O., Hsu, F.D., Jensen, K., Cheang, M., Karaca, G., Hu, Z., Hernandez-Boussard, T., Livasy, C., Cowan, D., Dressler, L., Akslen, L.A., Ragaz, J., Gown, A.M., Gilks, C.B., Van De Rijn, M., Perou, C.M., 2004. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin. Cancer Res. 10, 5367–5374.

O'Meara, T., Safonov, A., Casadevall, D., Qing, T., Silber, A., Killelea, B., Hatzis, C., Pusztai, L., 2019. Immune microenvironment of triple-negative breast cancer in African-American and Caucasian women. Breast Cancer Res. Treat. 175, 247–259.

O'Toole, S.A., Beith, J.M., Millar, E.K.A., West, R., McLean, A., Cazet, A., Swarbrick, A., Oakes, S.R., 2013. Therapeutic targets in triple negative breast cancer. J. Clin. Pathol. 66, 530–542.

Oda, N., Shimazu, K., Naoi, Y., Morimoto, K., Shimomura, A., Shimoda, M., Kagara, N., Maruyama, N., Kim, S.J., Noguchi, S., 2012. Intratumoral regulatory T cells as an independent predictive factor for pathological complete response to neoadjuvant paclitaxel followed by 5-FU/epirubicin/cyclophosphamide in breast cancer patients. Breast Cancer Res. Treat. 136, 107–116.

Okabe, M., Toh, U., Iwakuma, N., Saku, S., Akashi, M., Kimitsuki, Y., Seki, N., Kawahara, A., Ogo, E., Itoh, K., Akagi, Y., 2017. Predictive factors of the tumor immunological microenvironment for long-term follow-up in early stage breast cancer. Cancer Sci. 108, 81–90.

Olopade, O.I., Fackenthal, J.D., Dunston, G., Tainsky, M.A., Collins, F., Whitfield-Broome, C., 2003. Breast cancer genetics in African Americans. Cancer 97, 236–245.

Oluwagbemiga, L.A., Atoyebi, O., Adesunkanmi, A.R.K., 2015. Seventeen years after BRCA1: what is the BRCA mutation status of the breast cancer patients in Africa? – a systematic review. Springerplus 4, 1.

Omilian, A.R., Wei, L., Chen, C., Elisa, H., Song, V.B., Thaer, L., Ambrosone, C.B., Yao, S., 2020. Somatic mutations of triple - negative breast cancer : a comparison between Black and White women. Breast Cancer Res. Treat.

Osborne, B.C.K., Zhao, H.H., Fuqua, S.A.W., 2000. Selective Estrogen Receptor Modulators : Structure , Function , and Clinical Use. Biol. Neoplasia 18, 3172–3186.

Owens, O.L., Jackson, D.D., Thomas, T.L., Friedman, D.B., Hébert, J.R., 2013. African american men's and women's perceptions of clinical trials research: Focusing on prostate cancer among a high-risk population in the South. J. Health Care Poor Underserved 24, 1784–1800.

Ozawa, M., Baribault, H., Kemler, R., 1989. The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species. EMBO J. 8, 1711–1717.

Ozretic, P., Alvir, I., Sarcevic, B., Vujaskovic, Z., Rendic-Miocevic, Z., Roguljic, A., Beketic-Oreskovic, L., 2018. Apoptosis regulator bcl-2 is an independent prognostic marker for worse overall survival in triple-negative breast cancer patients. Int. J. Biol. Markers 33, 109–115.

Pakstis, A.J., Speed, W.C., Soundararajan, U., Rajeevan, H., Judith, R.K., Li, H., Kenneth, K.K., 2019. Population relationships based on 170 ancestry SNPs from the combined Kidd and Seldin panels. Sci. Rep. 1–8.

Palmer, J.R., Boggs, D.A., Wise, L.A., Ambrosone, C.B., Adams-Campbell, L.L., Rosenberg, L., 2011. Parity and lactation in relation to estrogen receptor negative breast cancer in African American women. Cancer Epidemiol Biomarkers Prev 20, 1883–1891.

Palmer, J.R., Castro-Webb, N., Bertrand, K., Bethea, T.N., Denis, G. V., 2017. Type II diabetes and incidence of estrogen receptor negative breast cancer in African American women. Cancer Res. 77, 6462–6469.

Pang, J., Toy, K.A., Griffith, K.A., Awuah, B., Quayson, S., Newman, L.A., Kleer, C.G., 2012. Invasive breast carcinomas in Ghana: High frequency of high grade, basal-like histology and high EZH2 expression. Breast Cancer Res. Treat. 135, 59–66.

Papadopoulos, K., 2006. Targeting the Bcl-2 Family in Cancer Therapy. Semin. Oncol. 33, 449–456.

Park, H.S., Jang, M.H., Kim, E.J., Kim, H.J., Lee, H.J., Kim, Y.J., Kim, J.H., Kang, E., Kim, S.W., Kim, I.A., Park, S.Y., 2014. High EGFR gene copy number predicts poor outcome in triple-negative breast cancer. Mod. Pathol. 27, 1212–1222.

Park, J.H., Ahn, J.H., Kim, S.B., 2018. How shall we treat early triple-negative breast cancer (TNBC): From the current standard to upcoming immuno-molecular strategies. ESMO Open 3, 1–16.

Patel, N., Weekes, D., Drosopoulos, K., Gazinska, P., Noel, E., Rashid, M., Mirza, H., Quist, J., Brasó-maristany, F., Mathew, S., Ferro, R., Pereira, A.M., Prince, C., Noor, F., Francesch-domenech, E., Marlow, R., Rinaldis, E. De, Grigoriadis, A., Linardopoulos, S., Marra, P., Tutt, A.N.J., 2018. Integrated genomics and functional validation identifies malignant cell specific dependencies in triple negative breast cancer. Nat. Commun.

Pelicano, H., Zhang, W., Liu, J., Hammoudi, N., Dai, J., Xu, R.H., Pusztai, L., Huang, P., 2014. Mitochondrial dysfunction in some triple-negative breast cancer cell lines: Role of mTOR pathway and therapeutic potential. Breast Cancer Res. 16, 1–16.

Peng, R., Wang, S., Shi, Y., Liu, D., Teng, X., Qin, T., Zeng, Y., Yuan, Z., 2011. Patients 35 years old or younger with operable breast cancer are more at risk for relapse and survival: A retrospective matched case-control study. Breast 20, 568–573.

Pereira, B., Chin, S.-F., Rueda, O.M., Vollan, H.-K.M., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.-J., Tsui, D.W.Y., Liu, B., Dawson, S.-J., Abraham, J., Northen, H., Peden, J.F., Mukherjee, A., Turashvili, G., Green, A.R., McKinney, S., Oloumi, A., Shah, S., Rosenfeld, N., Murphy, L., Bentley, D.R., Ellis, I.O., Purushotham, A., Pinder, S.E., Børresen-Dale, A.-L., Earl, H.M., Pharoah, P.D., Ross, M.T., Aparicio, S., Caldas, C., 2016. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat. Commun. 7, 11479.

Perez, E.A., 2007. Safety profiles of tamoxifen and the aromatase inhibitors in adjuvant therapy of hormone-responsive early breast cancer. Ann. Oncol. 18, 26–35.

Perou, C.M., Sùrlie, T., Eisen, M.B., Rijn, M. Van De, Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, I., Pergamenschikov, A., Williams, C., Zhu, S.X., Lùnning, P.E.,

Brown, P.O., Botstein, D., Grant, S., 2000. Molecular portraits of human breast tumours. Nature 406, 747–752.

Petrini, I., Meltzer, P.S., Zucali, P.A., Luo, J., Lee, C., Santoro, A., Lee, H.S., Killian, K.J., Wang, Y., Tsokos, M., Roncalli, M., Steinberg, S.M., Wang, Y., Giaccone, G., 2012. Copy number aberrations of BCL2 and CDKN2A/B identified by array-CGH in thymic epithelial tumors. Cell Death Dis. 3, 1–11.

Piccolo, S., Dupont, S., Cordenonsi, M., 2014. The biology of YAP/TAZ: Hippo signaling and beyond. Physiol. Rev. 94, 1287–1312.

Pierobon, M., Frankenfeld, C.L., 2013. Obesity as a risk factor for triple-negative breast cancers: A systematic review and meta-analysis. Breast Cancer Res. Treat. 137, 307–314.

Plotnikov, A., Zehorai, E., Procaccia, S., Seger, R., 2011. Biochimica et Biophysica Acta The MAPK cascades : Signaling components , nuclear roles and mechanisms of nuclear translocation ☆. BBA - Mol. Cell Res. 1813, 1619–1633.

Polak, P., Kim, J., Braunstein, L.Z., Karlic, R., Haradhavala, N.J., Tiao, G., Rosebrock, D., Livitz, D., Kübler, K., Mouw, K.W., Kamburov, A., Maruvka, Y.E., Leshchiner, I., Lander, E.S., Golub, T.R., Zick, A., Orthwein, A., Lawrence, M.S., Batra, R.N., Caldas, C., Haber, D.A., Laird, P.W., Shen, H., Ellisen, L.W., D'Andrea, A.D., Chanock, S.J., Foulkes, W.D., Getz, G., 2017. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat. Genet. 49, 1476–1486.

Popgeorgiev, N., Jabbour, L., Gillet, G., 2018. Subcellular localization and dynamics of the Bcl-2 family of proteins. Front. Cell Dev. Biol. 6, 1–11.

Porta, C., Paglino, C., Mosca, A., 2014. Targeting PI3K/Akt/mTOR signaling in cancer. Front. Oncol. 4 APR, 1–11.

Prasad, C.P., Gupta, S.D., Rath, G., Ralhan, R., 2008. Wnt signaling pathway in invasive ductal carcinoma of the breast: Relationship between β-catenin, disheveled and cyclin D1 expression. Oncology 73, 112–117.

Prat, A., Adamo, B., Cheang, M.C.U., Anders, C.K., Carey, L.A., Perou, C.M., 2013. Molecular Characterization of Basal-LikeandNon-Basal-Like Triple-Negative Breast Cancer. Oncologist 18, 123–133.

Prat, A., Guo, H., Ganesan, P., Parker, J.S., Gao, M., Moulder, S., Karginova, O., Lu, Y., Lee, J.J., Fan, C., Liang, J., Janku, F., Livasy, C., Lorenzi, P.L., Valero, V., Herschkowitz, J.I., Bai, S., Zinner, R.G., He, X., Hawke, D.H., Naing, A., Perou, C.M., Li, J., Fu, S., Dogruluk, T., Tsimberidou, A.M., Scott, K.L., Hong, D., Jonasch, E., Stephen, B., Mills, G.B., Stephens, P., Ding, Z., Yelensky, R., Meric-Bernstam, F., Kurzrock, R., Wheler, J.J., 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 12, R68.

Prat, A., Lluch, A., Albanell, J., Barry, W.T., Fan, C., Chacón, J.I., Parker, J.S., Calvo, L., Plazaola, A., Arcusa, A., Seguí-Palmer, M.A., Burgues, O., Ribelles, N., Rodriguez-Lescure, A., Guerrero, A., Ruiz-Borrego, M., Munarriz, B., López, J.A., Adamo, B., Cheang, M.C.U., Li, Y., Hu, Z., Gulley, M.L., Vidal, M.J., Pitcher, B.N., Liu, M.C., Citron, M.L., Ellis, M.J., Mardis, E., Vickery, T., Hudis, C.A., Winer, E.P., Carey, L.A., Caballero, R., Carrasco, E., Martín, M., Perou, C.M., Alba, E., 2014. Predicting response and survival in chemotherapy-treated triple-negative breast cancer. Br. J. Cancer 111, 1532–1541.

Prat, A., Parker, J.S., Fan, C., Perou, C.M., 2012. PAM50 assay and the three-gene model for identifying the major and clinically relevant molecular subtypes of breast cancer. Breast Cancer Res. Treat. 135, 301–306.

Prat, A., Perou, C.M., 2011. Deconstructing the molecular portraits of breast cancer. Mol. Oncol. 5, 5–23. https://doi.org/10.1016/j.molonc.2010.11.003

Pruneri, G., Vingiani, A., Bagnardi, V., Rotmensz, N., De Rose, A., Palazzo, A., Colleoni, A.M., Goldhirsch, A., Viale, G., 2016. Clinical validity of tumor-infiltrating lymphocytes analysis in patients with triple-negative breast cancer. Ann. Oncol. 27, 249–256.

Quail, D.F., Joyce, J.A., 2013. Microenvironmental regulation of tumor progression and metastasis. Nat. Med. 19, 1423–1437.

Rakha, E.A., El-Rehim, D.A., Paish, C., Green, A.R., Lee, A.H.S., Robertson, J.F., Blamey, R.W., Macmillan, D., Ellis, I.O., 2006. Basal phenotype identifies a poor prognostic subgroup of breast cancer of clinical importance. Eur. J. Cancer 42, 3149–3156.

Rakha, E.A., El-Sayed, M.E., Green, A.R., Lee, A.H.S., Robertson, J.F., Ellis, I.O., 2007. Prognostic markers in triple-negative breast cancer. Cancer 109, 25–32.

Razzaghi, H., Quesnel-Crooks, S., Sherman, R., Joseph, R., Kohler, B., Andall-Brereton, G., Ivey, M.A., Edwards, B.K., Mery, L., Gawryszewski, V., Saraiya, M., 2016. Leading causes of cancer mortality — Caribbean Region, 2003–2013. Morb. Mortal. Wkly. Rep. 65, 1395–1400.

Reed, J.C., 1998. Bcl-2 family proteins 3225–3236.

Ring, B.Z., Hout, D.R., Morris, S.W., Lawrence, K., Schweitzer, B.L., Bailey, D.B., Lehmann, B.D., Pietenpol, J.A., Seitz, R.S., 2016. Generation of an algorithm based on minimal gene sets to clinically subtype triple negative breast cancer patients. BMC Cancer 16, 1–8.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009a. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009b. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11.

Robinson, M.D., Smyth, G.K., 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics 9, 321–332.

Robinson, M.D., Smyth, G.K., 2007. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics 23, 2881–2887.

Robson, M., Im, S.A., Senkus, E., Xu, B., Domchek, S.M., Masuda, N., Delaloge, S., Li, W., Tung, N., Armstrong, A., Wu, W., Goessl, C., Runswick, S., Conte, P., 2017. Olaparib for metastatic breast cancer in patients with a germline BRCA mutation. N. Engl. J. Med. 377, 523–533.

Rodler, E.T., Kurland, B.F., Griffin, M., Gralow, J.R., Porter, P., Yeh, R.F., Gadi, V.K., Guenthoer, J., Beumer, J.H., Korde, L., Strychor, S., Kiesel, B.F., Linden, H.M., Thompson, J.A., Swisher, E., Chai, X., Shepherd, S., Giranda, V., Specht, J.M., 2016. Phase I study of veliparib (ABT-888) combined with cisplatin and vinorelbine in advanced triple-negative breast cancer and/or BRCA mutation-associated breast cancer. Clin. Cancer Res. 22, 2855–2864.

Ronen, A., Glickman, B.W., 2001. Human DNA repair genes. Environ. Mol. Mutagen. 37, 241–283.

Rong, X., Liang, Y., Han, Q., Zhao, Y., Jiang, G., Zhang, X., Lin, X., Liu, Y., Zhang, Y., Han, X., Zhang, M., Luo, Y., Li, P., Wei, L., Yan, T., Wang, E., 2019. Molecular Mechanisms of Tyrosine Kinase Inhibitor Resistance Induced by Membranous/Cytoplasmic/Nuclear Translocation of Epidermal Growth Factor Receptor. J. Thorac. Oncol. 14, 1766–1783.

Rosenberg, P., 2015. Abstract 1850. Present. Am. Assoc. Cancer Res. Annu. Meet.

Rouzier, R., Perou, C.M., Symmans, W.F., Ibrahim, N., Cristofanilli, M., Anderson, K., Hess, K.R., Stec, J., Ayers, M., Wagner, P., Morandi, P., Fan, C., Rabiul, I., Ross, J.S., Hortobagyi, G.N., Pusztai, L., 2005. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin.

Cancer Res. 11, 5678–5685.

Ruiz-Narváez, E.A., Rosenberg, L., Cozier, Y.C., Cupples, L.A., Adams-Campbell, L.L., Palmer, J.R., 2010. Polymorphisms in the TOX3/LOC643714 locus and risk of breast cancer in African-American women. Cancer Epidemiol. Biomarkers Prev. 19, 1320–1327. https://doi.org/10.1158/1055-9965.EPI-09-1250

Saal, L.H., Holm, K., Maurer, M., Memeo, L., Su, T., Wang, X., Mansukhani, M., Enoksson, J., Hibshoosh, H., Yu, J.S., Malmstro, P., 2005. PIK3CA Mutations Correlate with Hormone Receptors , Node Metastasis , and ERBB2 , and Are Mutually Exclusive with PTEN Loss in Human Breast Carcinoma 2554–2560.

Sachdev, J.C., Ahmed, S., Mirza, M.M., Farooq, A., Kronish, L., Jahanzeb, M., 2010. Does race affect outcomes in triple negative breast cancer? Breast Cancer Basic Clin. Res. 4, 23–33.

Sager, R., 1997. Expression genetics in cancer: Shifting the focus from DNA to RNA. Proc. Natl. Acad. Sci. U. S. A. 94, 952–955.

Saini, K.S., Loi, S., de Azambuja, E., Metzger-Filho, O., Saini, M.L., Ignatiadis, M., Dancey, J.E., Piccart-Gebhart, M.J., 2013. Targeting the PI3K/AKT/mTOR and Raf/MEK/ERK pathways in the treatment of breast cancer. Cancer Treat. Rev. 39, 935–946.

Sánchez-Tilló, E., De Barrios, O., Siles, L., Cuatrecasas, M., Castells, A., Postigo, A., 2011. β-catenin/TCF4 complex induces the epithelial-to-mesenchymal transition (EMT)-activator ZEB1 to regulate tumor invasiveness. Proc. Natl. Acad. Sci. U. S. A. 108, 19204–19209.

Sangai, Takafumi, Akcakanat, A., Chen, Huiqin, Tarco, E., Wu, Yun, Do, K.-A., Miller, T.W., Arteaga, C.L., Mills, G.B., Maria Gonzalez-Angulo, A., Meric-Bernstam, F., Sangai, T, Chen, H, Wu, Y, 2012. Biomarkers of Response to Akt Inhibitor MK-2206 in Breast Cancer Experimental Design—MK-2206 effects on cell signaling were assessed in vitro and in vivo. Its antitumor efficacy was assessed in vitro in a panel of cancer cell lines with differing PIK3CA a. Clin Cancer Res. Clin Cancer Res. Oct. 15, 5816–5828.

Saraiva, D.P., Guadalupe Cabral, M., Jacinto, A., Braga, S., 2017. How many diseases is triple negative breast cancer: The protagonism of the immune microenvironment. ESMO Open 2.

Saraiva, D.P., Jacinto, A., Borralho, P., Braga, S., 2018. HLA-DR in Cytotoxic T Lymphocytes Predicts Breast Cancer Patients' Response to Neoadjuvant Chemotherapy 9.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: An open-source platform for biological-image analysis. Nat. Methods 9, 676–682.

Schlessinger, J., 2014. Receptor tyrosine kinases: Legacy of the first two decades. Cold Spring Harb. Perspect. Biol. 6, 1–14.

Schmadeka, R., Harmon, B.E., Singh, M., 2014. Triple-negative breast carcinoma: Current and emerging concepts. Am. J. Clin. Pathol. 141, 462–477.

Schmid, P., Adams, S., Rugo, H.S., Schneeweiss, A., Barrios, C.H., Iwata, H., Dieras, V., Hegg, R., Im, S.A., Shaw Wright, G., Henschel, V., Molinero, L., Chui, S.Y., Funke, R., Husain, A., Winer, E.P., Loi, S., Emens, L.A., 2018. Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. N. Engl. J. Med. 379, 2108–2121.

Schmid, P., Rugo, H.S., Adams, S., Schneeweiss, A., Barrios, C.H., Iwata, H., Diéras, V., Henschel, V., Molinero, L., Chui, S.Y., Maiya, V., Husain, A., Winer, E.P., Loi, S., Emens, L.A., 2020. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-

blind, placebo-controlled, phase 3 trial. Lancet Oncol. 21, 44–59.

Schneider, V.A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.C., Kitts, P.A., Murphy, T.D., Pruitt, K.D., Thibaud-Nissen, F., Albracht, D., Fulton, R.S., Kremitzki, M., Magrini, V., Markovic, C., McGrath, S., Steinberg, K.M., Auger, K., Chow, W., Collins, J., Harden, G., Hubbard, T., Pelan, S., Simpson, J.T., Threadgold, G., Torrance, J., Wood, J.M., Clarke, L., Koren, S., Boitano, M., Peluso, P., Li, H., Chin, C.S., Phillippy, A.M., Durbin, R., Wilson, R.K., Flicek, P., Eichler, E.E., Church, D.M., 2017. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res. 27, 849–864.

Schwartz, T., Stark, A., Pang, J., Awuah, B., Kleer, C.G., Quayson, S., Kingman, S., Aitpillah, F., Abantanga, F., Jiagge, E., Oppong, J.K., Osei-Bonsu, E., Martin, I., Yan, X., Toy, K., Adjei, E., Wicha, M., Newman, L.A., 2013. Expression of aldehyde dehydrogenase 1 as a marker of mammary stem cells in benign and malignant breast lesions of Ghanaian women. Cancer 119, 488–494.

Seewaldt, V.L., Scott, V., 2007. Rapid progression of basal-type breast cancer. N. Engl. J. Med. 356, 2015.

Selvendiran, K., Ahmed, S., Dayton, A., Ravi, Y., Kuppusamy, M.L., Bratasz, A., Rivera, B.K., Kálai, T., Hideg, K., Kuppusamy, P., 2010. HO-3867, a synthetic compound, inhibits the migration and invasion of ovarian carcinoma cells through downregulation of fatty acid synthase and focal adhesion kinase. Mol. Cancer Res. 8, 1188–1197.

Serre, D., Pääbo, S., 2004. Evidence for gradients of human genetic diversity within and among continents. Genome Res. 14, 1679–1685.

Servick, K., 2014. Breast cancer: A world of differences. Science. 343, 1452–1453.

Shah, S.P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., Bashashati, A., Prentice, L.M., Khattra, J., Burleigh, A., Yap, D., Bernard, V., Mcpherson, A., Shumansky, K., Crisan, A., Giuliany, R., Heravi-moussavi, A., Rosner, J., Lai, D., Birol, I., Varhol, R., Tam, A., Dhalla, N., Zeng, T., Ma, K., Chan, S.K., Griffith, M., Moradian, A., Cheng, S.G., Morin, G.B., Watson, P., Gelmon, K., Chia, S., Chin, S., Curtis, C., Rueda, O.M., Pharoah, P.D., 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 486, 395–399.

Shall, S., de Murcia, G., 2000. Poly(ADP-ribose) polymerase-1: What have we learned from the deficient mouse model? Mutat. Res. - DNA Repair 460, 1–15. https://doi.org/10.1016/S0921-8777(00)00016-1

Shang, S., Hua, F., Hu, Z.W., 2017. The regulation of β-catenin activity and function in cancer: Therapeutic opportunities. Oncotarget 8, 33972–33989.

Sharma, D., Davidson, N.E., 2013. Obesity and breast cancer: A multipartite connection. J. Mammary Gland Biol. Neoplasia 18, 253–255.

Shavers, V.L., Brown, M.L., 2002. Racial and ethnic disparities in the receipt of cancer treatment. J. Natl. Cancer Inst. 94, 334–357.

Shen, Y., Dong, W., Esteva, F.J., Kau, S.W., Theriault, R.L., Bevers, T.B., 2007. Are there racial differences in breast cancer treatments and clinical outcomes for women treated at M.D. Anderson Cancer Center? Breast Cancer Res. Treat. 102, 347–356.

Shimelis, H., LaDuca, H., Hu, C., Hart, S.N., Na, J., Thomas, A., Akinhanmi, M., Moore, R.M., Brauch, H., Cox, A., Eccles, D.M., Ewart-Toland, A., Fasching, P.A., Fostira, F., Garber, J., Godwin, A.K., Konstantopoulou, I., Nevanlinna, H., Sharma, P., Yannoukakos, D., Yao, S., Feng, B.J., Davis, B.T., Lilyquist, J., Pesaran, T., Goldgar, D.E., Polley, E.C., Dolinsky, J.S., Couch, F.J., 2018. Triple-negative breast cancer risk genes identified by multigene hereditary cancer panel testing. J. Natl. Cancer Inst. 110, 855–862.

Shinde, S.S., Forman, M.R., Kuerer, H.M., Yan, K., Peintinger, F., Hunt, K.K., Hortobagyi, G.N., Pusztai, L., Symmans, W.F., 2010. Higher parity and shorter breastfeeding duration. Cancer 116, 4933–4943.

Shoemaker, M.L., White, M.C., Wu, M., Weir, H.K., Romieu, I., 2018. Differences in breast cancer incidence among young women aged 20–49 years by stage and tumor characteristics, age, race, and ethnicity, 2004–2013. Breast Cancer Res. Treat. 169, 595–606.

Siddharth, S., Sharma, D., 2018. Racial disparity and triple-negative breast cancer in African-American women: A multifaceted affair between obesity, biology, and socioeconomic determinants. Cancers (Basel). 10, 1–19.

Sigismund, S., Avanzato, D., Lanzetti, L., 2018. Emerging functions of the EGFR in cancer. Mol. Oncol. 12, 3–20.

Sikov, W.M., Berry, D.A., Perou, C.M., Singh, B., Cirrincione, C.T., Tolaney, S.M., Kuzma, C.S., Pluard, T.J., Somlo, G., Port, E.R., Golshan, M., Bellon, J.R., Collyar, D., Hahn, O.M., Carey, L.A., Hudis, C.A., Winer, E.P., 2015. Impact of the addition of carboplatin and/or bevacizumab to neoadjuvant once-per-week paclitaxel followed by dose-dense doxorubicin and cyclophosphamide on pathologic complete response rates in stage II to III triple-negative breast cancer: CALGB 40603 (A. J. Clin. Oncol. 33, 13–21.

Silber, J.H., Rosenbaum, P.R., Clark, A.S., Giantonio, B.J., Ross, R.N., Teng, Y., Wang, M., Niknam, B.A., Ludwig, J.M., Wang, W., Even-Shoshan, O., Fox, K.R., 2013. Characteristics associated with differences in survival among black and white women with breast cancer. JAMA - J. Am. Med. Assoc. 310, 389–397.

Silva, T.C., Colaprico, A., Olsen, C., D'Angelo, F., Bontempi, G., Ceccarelli, M., Noushmehr, H., 2016. TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages [version 1; referees: 1 approved, 1 approved with reservations]. F1000Research 5.

Silver, D.P., Richardson, A.L., Eklund, A.C., Wang, Z.C., Szallasi, Z., Li, Q., Juul, N., Leong, C.O., Calogrias, D., Buraimoh, A., Fatima, A., Gelman, R.S., Ryan, P.D., Tung, N.M., De Nicolo, A., Ganesan, S., Miron, A., Colin, C., Sgroi, D.C., Ellisen, L.W., Winer, E.P., Garber, J.E., 2010. Efficacy of neoadjuvant cisplatin in triple-negative breast cancer. J. Clin. Oncol. 28, 1145–1153.

Sims, A.H., Howell, A., Howell, S.J., Clarke, R.B., 2007. Origins of breast cancer subtypes and therapeutic implications. Nat. Clin. Pract. Oncol. 4, 516–525.

Singh, A.M., Dalton, S., 2009. The Cell Cycle and Myc Intersect with Mechanisms that Regulate Pluripotency and Reprogramming. Cell Stem Cell 5, 141–149.

Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A., Mcguire, W.L., 1987. Human Breast Cancer: Correlation of Relapse and Survival with Amplification of the HER-2/neu Oncogene. Science (80-.). 235, 177–182.

Smid, M., Wang, Y., Zhang, Y., Sieuwerts, A.M., Yu, J., Klijn, J.G.M., Foekens, J.A., Martens, J.W.M., 2008. Subtypes of breast cancer show preferential site of relapse. Cancer Res. 68, 3108–3114.

Solovieff, N., Hartley, S.W., Baldwin, C.T., Klings, E.S., Gladwin, M.T., Taylor, J.G., Kato, G.J., Farrer, L.A., Steinberg, M.H., Sebastiani, P., 2011. Ancestry of African Americans with sickle cell disease. Blood Cells, Mol. Dis. 47, 41–45.

Somlo, G., Frankel, P.H., Arun, B.K., Ma, C.X., Garcia, A.A., Cigler, T., Cream, L. V., Harvey, H.A., Sparano, J.A., Nanda, R., Chew, H.K., Moynihan, T.J., Vahdat, L.T., Goetz, M.P., Beumer, J.H., Hurria, A., Mortimer, J., Piekarz, R., Sand, S., Herzog, J., Van Tongeren, L.R., Ferry-Galow, K. V., Chen, A.P., Ruel, C., Newman, E.M., Gandara, D.R., Weitzel, J.N., 2017. Efficacy of the PARP inhibitor veliparib with carboplatin or as a single agent in patients with germline BRCA1- or BRCA2-associated metastatic breast cancer: California Cancer Consortium trial NCT01149083. Clin. Cancer Res. 23,

4066–4076.

Sonugür, F.G., Akbulut, H., 2019. The Role of Tumor Microenvironment in Genomic Instability of Malignant Tumors 10, 1–7.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., Rijn, M. Van De, Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Eystein, P., Børresen-dale, A., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications 98.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C.M., Lønning, P.E., Brown, P.O., Botstein, D., 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets 100, 8418–8423.

Soucek, L., Jucker, R., Panacchia, L., Ricordy, R., Tatò, F., Nasi, S., 2002. Omomyc, a potential Myc dominant negative, enhances Myc-induced apoptosis. Cancer Res. 62, 3507–3510.

St John, M.A.R., Tao, W., Fei, X., Fukumoto, R., Carcangiu, M.L., Brownstein, D.G., Parlow, A.F., McGrath, J., Xu, T., 1999. Mice deficient of Lats1 develop soft-tissue sarcomas, ovarian tumours and pituitary dysfunction. Nat. Genet. 21, 182–186.

Staal, F.J.T., Sen, J.M., 2008. The canonical Wnt signaling pathway plays an important role in lymphopoiesis and hematopoiesis. Eur. J. Immunol. 38, 1788–1794.

Stark, A., Kleer, C.G., Martin, I., Awuah, B., Nsiah-Asare, A., Takyi, V., Braman, M., Quayson, S.E., FWACP, Zarbo, R., Wicha, M., 2010. African ancestry and higher prevalence of triple-negative breast cancer. Cancer. Cancer 116, 4926–4932.

Stead, L.A., Lash, T.L., Sobieraj, J.E., Chi, D.D., Westrup, J.L., Charlot, M., Blanchard, R.A., Lee, J.C., King, T.C., Rosenberg, C.L., 2009. Triple-negative breast cancers are increased in black women regardless of age or body mass index. Breast Cancer Res. 11, 1–10.

Steinmann, K., Sandner, A., Schagdarsurengin, U., Damman, R.H., 2009. Frequent promoter hypermethylation of tumor-related genes in head and neck squamous cell carcinoma. Oncol. Rep. 22, 1519–1526.

Stephens, J.M., Lee, J., Pilch, P.F., 1997. Tumor necrosis factor-α-induced insulin resistance in 3T3-L1 adipocytes is accompanied by a loss of insulin receptor substrate-1 and GLUT4 expression without a loss of insulin receptor-mediated signal transduction. J. Biol. Chem. 272, 971–976.

Stewart, C., Pepper, M.S., 2016. Cystic fibrosis on the African continent. Genet. Med. 18, 653–662.

Stewart, P.A., Luks, J., Roycik, M.D., Sang, Q.X.A., Zhang, J., 2013. Differentially expressed transcripts and dysregulated signaling pathways and networks in African American breast cancer. PLoS One 8, 1–13.

Stine, Z.E., Walton, Z.E., Altman, B.J., Hsieh, A.L., Dang, C. V., 2015. MYC, metabolism, and cancer. Cancer Discov. 5, 1024–1039.

Sturtz, L.A., Melley, J., Mamula, K., Shriver, C.D., Ellsworth, R.E., 2014. Outcome disparities in African American women with triple negative breast cancer: A comparison of epidemiological and molecular factors between African American and Caucasian women with triple negative breast cancer. BMC Cancer 14.

Sugita, B., Gill, M., Mahajan, A., Duttargi, A., Kirolikar, S., Almeida, R., Regis, K., Oluwasanmi, O.L., Marchi, F., Marian, C., Makambi, K., Kallakury, B., Sheahan, L., J.Cavalli, I., Ribeiro, E.M., Madhavan, S., Boca, S., Gusev, Y., Cavalli, L.R., 2016. Differentially expressed miRNAs in triple negative breast cancer between African-American and non-Hispanic white women. Oncotarget 7, 79274–79291.

Sullivan, H.C., Oprea-Ilies, G., Adams, A.L., Page, A.J., Kim, S., Wang, J., Cohen, C., 2014. Triple-negative breast carcinoma in African American and caucasian women: Clinicopathology, immunomarkers, and outcome. Appl. Immunohistochem. Mol. Morphol. 22, 17–23.

Sullivan, N.J., Sasser, A.K., Axel, A.E., Vesuna, F., Raman, V., Ramirez, N., Oberyszyn, T.M., Hall, B.M., 2009. Interleukin-6 induces an epithelial-mesenchymal transition phenotype in human breast cancer cells. Oncogene 28, 2940–2947.

Sun, G., Irvine, K.D., 2011. Regulation of Hippo signaling by Jun kinase signaling during compensatory cell proliferation and regeneration, and in neoplastic tumors. Dev. Biol. 350, 139–151.

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. 0, 1–41.

Swanson, G., Lin, C., 1994. Survival patterns among younger women with breast cancer: the effects of age, race, stage, and treatment. Natl Cancer Inst Monogr 69–77.

Szabo, C.I., King, M.C., 1997. Population genetics of BRCA1 and BRCA2. Am. J. Hum. Genet. 60, 1013–1020.

Szychot, E., Brodkiewicz, A., Peregud-Pogorzelski, J., 2013. Will Therapies that Target Tumour Suppressor Genes Czy terapia celowana prowadząca do utraty funkcji. Adv Clin Exp Med 22, 861–864.

Takahashi, K., Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126, 663–676.

Takahashi, Y., Miyoshi, Y., Takahata, C., Irahara, N., Taguchi, T., Tamaki, Y., Noguchi, S., 2005. Down-regulation of LATS1 and LATS2 mRNA expression by promoter hypermethylation and its association with biologically aggressive phenotype in human breast cancers. Clin. Cancer Res. 11, 1380–1385.

Taliaferro-Smith, L.T., Oberlick, E., Liu, T., McGlothen, T., Alcaide, T., Tobin, R., Donnelly, S., Commander, R., Kline, E., Nagaraju, G.P., Havel, L., Marcus, A., Nahta, R., O'Regan, R., 2015. FAK activation is required for IGF1R-mediated regulation of EMT, migration, and invasion in mesenchymal triple negative breast cancer cells. Oncotarget 6, 4757–4772.

Tawfik, K., Kimler, B.F., Davis, M.K., Fan, F., Tawfik, O., 2012. Prognostic significance of Bcl-2 in invasive mammary carcinomas: A comparative clinicopathologic study between "triple-negative" and non-"triple-negative" tumors. Hum. Pathol. 43, 23–30.

Tayyari, F., Gowda, G.A.N., Olopade, O.F., Berg, R., Yang, H.H., Lee, M.P., Ngwa, W.F., Mittal, S.K., Raftery, D., Mohammed, S.I., 2018. Metabolic profiles of triple-negative and luminal A breast cancer subtypes in African-American identify key metabolic differences. Oncotarget 9, 11677–11690.

Teleki, I., Szasz, A.M., Maros, M.E., Gyorffy, B., Kulka, J., Meggyeshazi, N., Kiszner, G., Balla, P., Samu, A., Krenacs, T., 2014. Correlations of differentially expressed gap junction connexins cx26, cx30, cx32, cx43 and cx46 with breast cancer progression and prognosis. PLoS One 9, 1–12.

Ter Brugge, P., Kristel, P., Van Der Burg, E., Boon, U., De Maaker, M., Lips, E., Mulder, L., De Ruiter, J., Moutinho, C., Gevensleben, H., Marangoni, E., Majewski, I., Jozwiak, K., Kloosterman, W., Van Roosmalen, M., Duran, K., Hogervorst, F., Turner, N., Esteller, M., Cuppen, E., Wesseling, J., Jonkers, J., 2016. Mechanisms of therapy resistance in patient-derived xenograft models of brca1-deficient breast cancer. J. Natl. Cancer Inst. 108, 1–12.

Therneau, T., 1999. A package for survival analysis in S. Citeseer 1–83.

Thodberg, S., Neilson, E.H.J., 2020. Application of Plant Flavoproteins. Catalysts 10.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J., Doumbo, O., Ibrahim, M., Juma, A.T., Maritha, J., Lema, G., Moore, J.H., Mortensen, H., Nyambo, T.B., Omar, S.A., Powell, K., Pretorius, G.S., Smith, M.W., Thera, A., Wambebe, C., Weber, J.L., Williams, S.M., 2009. The Genetic Structure and History of Africans and African Americans. Science. 324, 1035–1044.

Tolcher, A W, Bendell, J.C., Papadopoulos, K.P., Iii, H.A.B., Patnaik, A., Jones, S.F., Rasco, D., Cox, D.S., Durante, M., Bellew, K.M., Park, J., Le, N.T., Infante, J.R., 2015. A phase IB trial of the oral MEK inhibitor trametinib (GSK1120212) in combination with everolimus in patients with advanced solid tumors 58–64.

Tolcher, Anthony W, Patnaik, A., Papadopoulos, K.P., Rasco, D.W., Becerra, C.R., Allred, A.J., Orford, K., Aktan, G., Ferron, G., Ibrahim, N., Gauvin, J., Motwani, M., Cornfeld, M., 2015. Phase I study of the MEK inhibitor trametinib in combination with the AKT inhibitor afuresertib in patients with solid tumors and multiple myeloma 183–189.

Toma, T.T., Dawson, J.M., Adjeroh, D.A., 2018. Human ancestry indentification under resource constraints -- what can one chromosome tell us about human biogeographical ancestry? 11.

Tomczak, K., Czerwińska, P., Wiznerowicz, M., 2015. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. Wspolczesna Onkol. 1A, A68–A77.

Tomioka, N., Azuma, M., Ikarashi, M., Yamamoto, M., Sato, M., Watanabe, K. ichi, Yamashiro, K., Takahashi, M., 2018. The therapeutic candidate for immune checkpoint inhibitors elucidated by the status of tumor-infiltrating lymphocytes (TILs) and programmed death ligand 1 (PD-L1) expression in triple negative breast cancer (TNBC). Breast Cancer 25, 34–42.

Tovey, H., Bliss, J., Tutt, A., Morden, J., Jarman, K., Martin, S., Kernaghan, S., Toms, C., Kilburn, L., 2015. Managing non-proportionality of hazards (PH) within TNT: a randomised phase III trial of carboplatin compared to docetaxel for patients with metastatic or recurrent locally advanced triple negative (TN) or brca1/2 breast cancer (BC). Trials 16, P150.

Tse, G.M., Tan, P.H., Lau, K.M., De Andrade, V.P., Lui, P.C.W., Vong, J.S., Chaiwun, B., Lam, C.C., Yu, A.M., Moriya, T., 2009. Breast cancer in the elderly: A histological assessment. Histopathology 55, 441–451.

Tsutsui, S., Yasuda, K., Suzuki, K., Takeuchi, H., Nishizaki, T., Higashi, H., Era, S., 2006. Bcl-2 protein expression is associated with p27 and p53 protein expressions and MIB-1 counts in breast cancer. BMC Cancer 6, 1–7.

Turner, N., Tutt, A., Ashworth, A., 2004. Hallmarks of “BRCAness” in sporadic cancers. Nat. Rev. Cancer 4, 814–819.

Turner, N.C., Lord, C.J., Iorns, E., Brough, R., Swift, S., Elliott, R., Rayter, S., Tutt, A.N., Ashworth, A., 2008. A synthetic lethal siRNA screen identifying genes mediating sensitivity to a PARP inhibitor. EMBO J. 27, 1368–1377.

Tutt, A., Bertwistle, D., Valentine, J., Gabriel, A., Swift, S., Ross, G., Griffin, C., Thacker, J., Ashworth, A., 2001. Mutation in Brca2 stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences. EMBO J. 20, 4704–4716.

Tutt, A., Robson, M., Garber, J.E., Domchek, S.M., Audeh, M.W., Weitzel, J.N., Friedlander, M., Arun, B., Loman, N., Schmutzler, R.K., Wardley, A., Mitchell, G., Earl, H., Wickens, M., Carmichael, J., 2010. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: A proof-of-concept trial. Lancet 376, 235–244.

van Schie, E.H., van Amerongen, R., 2020. Aberrant WNT/CTNNB1 Signaling as a Therapeutic Target in Human Breast Cancer: Weighing the Evidence. Front. Cell Dev. Biol. 8, 1–14.

Varghese, F., Bukhari, A.B., Malhotra, R., De, A., 2014. IHC profiler: An open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples. PLoS One 9.

Vaux, D.L., Cory, S., Adams, J.M., 1988. Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. Nature.

Viale, G., Rotmensz, N., Maisonneuve, P., Bottiglieri, L., Montagna, E., Luini, A., Veronesi, P., Intra, M., Torrisi, R., Cardillo, A., Campagnoli, E., Goldhirsch, A., Colleoni, M., 2009. Invasive ductal carcinoma of the breast with the "triple- negative" phenotype: Prognostic implications of EGFR immunoreactivity. Breast Cancer Res. Treat. 116, 317–328.

Vici, P., Ercolani, C., Di Benedetto, A., Pizzuti, L., Di Lauro, L., Sperati, F., Terrenato, I., Gamucci, T., Natoli, C., Di Filippo, F., Botti, C., Barba, M., Mottolese, M., De Maria, R., Maugeri-Saccà, M., 2016. Topographic expression of the Hippo transducers TAZ and YAP in triple-negative breast cancer treated with neoadjuvant chemotherapy. J. Exp. Clin. Cancer Res. 35, 1–8.

Vivanco, I., Rohle, D., Versele, M., Iwanami, A., Kuga, D., Oldrini, B., Tanaka, K., Dang, J., Kubek, S., Palaskas, N., Hsueh, T., Evans, M., Mulholland, D., Wolle, D., Rajasekaran, S., Rajasekaran, A., Liau, L.M., Cloughesy, T.F., Dikic, I., Brennan, C., Wu, H., Mischel, P.S., Perera, T., Mellinghoff, I.K., 2010. The phosphatase and tensin homolog regulates epidermal growth factor receptor (EGFR) inhibitor response by targeting EGFR for degradation. Proc. Natl. Acad. Sci. U. S. A. 107, 6459–6464.

Vogelstein, Bert; Lane, David; J. Levine, A., 2000. Surfing p53 Network. Nature 408, 307–310.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Kinzler, K.W., 2013. Cancer genome landscapes. Science (80-.). 340, 1546–1558.

Von Minckwitz, G., Schneeweiss, A., Loibl, S., Salat, C., Denkert, C., Rezai, M., Blohmer, J.U., Jackisch, C., Paepke, S., Gerber, B., Zahm, D.M., Kümmel, S., Eidtmann, H., Klare, P., Huober, J., Costa, S., Tesch, H., Hanusch, C., Hilfrich, J., Khandan, F., Fasching, P.A., Sinn, B. V., Engels, K., Mehta, K., Nekljudova, V., Untch, M., 2014. Neoadjuvant carboplatin in patients with triple-negative and HER2-positive early breast cancer (GeparSixto; GBG 66): A randomised phase 2 trial. Lancet Oncol. 15, 747–756.

Von Minckwitz, G., Untch, M., Blohmer, J.U., Costa, S.D., Eidtmann, H., Fasching, P.A., Gerber, B., Eiermann, W., Hilfrich, J., Huober, J., Jackisch, C., Kaufmann, M., Konecny, G.E., Denkert, C., Nekljudova, V., Mehta, K., Loibl, S., 2012. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J. Clin. Oncol. 30, 1796–1804.

Vona-Davis, L., Rose, D.P., 2009. The influence of socioeconomic disparities on breast cancer tumor biology and prognosis: A review. J. Women's Heal. 18, 883–893.

Vona-davis, L., Rose, D.P., Hazard, H., Howard-mcnatt, M., Adkins, F., Partin, J., Hobbs, G., 2008. Triple-Negative Breast Cancer and Obesity in a Rural Appalachian Population 17, 3319–3325.

Walter, M., Liang, S., Ghosh, S., Hornsby, P.J., Li, R., 2009. Interleukin 6 secreted from adipose stromal cells promotes migration and invasion of breast cancer cells. Oncogene 28, 2745–2755.

Wang, Jianhua, Shi, M., Ling, R., Xia, Y., Luo, S., Fu, X., Xiao, F., Li, J., Long, X., Wang, Jianguo, Hou, Z., Chen, Y., Zhou, B., Xu, M., 2011. Adjuvant chemotherapy and radiotherapy in triple-negative breast carcinoma: A prospective randomized controlled multi-center trial. Radiother. Oncol. 100, 200–204.

Wang, Y., Beydoun, M.A., 2007. The obesity epidemic in the United States - Gender, age, socioeconomic, racial/ethnic, and geographic characteristics: A systematic review and meta-regression analysis. Epidemiol. Rev. 29, 6–28.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57–63.

Warnes, A.G.R., Bolker, B., Bonebakker, L., Huber, W., Liaw, A., Lumley, T., Magnusson, A., Moeller, S., Schwartz, M., 2020. Package 'gplots'.

Wei, B., Wang, J., Bourne, P., Yang, Q., Hicks, D., Bu, H., Tang, P., 2008. Bone metastasis is strongly associated with estrogen receptor-positive/progesterone receptor-negative breast carcinomas. Hum. Pathol. 39, 1809–1815.

Wei, M., Grushko, T.A., Dignam, J., Hagos, F., Nanda, R., Sveen, L., Xu, J., Fackenthal, J., Tretiakova, M., Das, S., Olopade, O.I., 2005. BRCA1 promoter methylation in sporadic breast cancer is associated with reduced BRCA1 copy number and chromosome 17 aneusomy. Cancer Res. 65, 10692–10699.

Wein, L., Savas, P., Luen, S.J., Virassamy, B., Salgado, R., Loi, S., 2017. Clinical validity and Utility of Tumor-infiltrating lymphocytes in routine clinical practice for breast cancer patients: Current and future directions. Front. Oncol. 7, 156.

Weinberg, R.A. (Robert A., 1996. How Cancer Arises.

Weisberg, S.P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R.L., Ferrante, A.W., 2003. Obesity is associated with macrophage accumulation in adipose tissue. J. Clin. Invest. 112, 1796–1808.

Weisman, P.S., Ng, C.K.Y., Brogi, E., Eisenberg, R.E., Won, H.H., Piscuoglio, S., De Filippo, M.R., Ioris, R., Akram, M., Norton, L., Weigelt, B., Berger, M.F., Reis-Filho, J.S., Wen, H.Y., 2016. Genetic alterations of triple negative breast cancer by targeted next-generation sequencing and correlation with tumor morphology. Mod. Pathol. 29, 476–488.

Wendt, M.K., Williams, W.K., Pascuzzi, P.E., Balanis, N.G., Schiemann, B.J., Carlin, C.R., Schiemann, W.P., 2015. The Antitumorigenic Function of EGFR in Metastatic Breast Cancer is Regulated by Expression of Mig6. Neoplasia (United States) 17, 124–133.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

WOLFF, A.C., HAMMOND, M.E.H., SCHWARTZ, J.N., HAYES, D.F., 2007. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Arch. Pathol. Lab. Med. 131, 1330–1331.

Woodward, W.A., Huang, E.H., McNeese, M.D., Perkins, G.H., Tucker, S.L., Strom, E.A., Middleton, L., Hahn, K., Hortobagyi, G.N., Buchholz, T.A., 2006. African-American race is associated with a poorer overall survival rate for breast cancer patients treated with mastectomy and doxorubicin-based chemotherapy. Cancer 107, 2662–2668.

World Health Organization, 2008. Waist Circumference and Waist-Hip Ratio: Report of a WHO Expert Consultation. WHO Libr. Cat. Data Geneva.

Wozniak, M.A., Modzelewska, K., Kwong, L., Keely, P.J., 2004. Focal adhesion regulation of cell behavior. Biochim. Biophys. Acta - Mol. Cell Res. 1692, 103–119.

Wright, H.J., Hou, J., Xu, B., Cortez, M., Potma, E.O., Tromberg, B.J., Razorenova, O. V., 2017. CDCP1 drives triple-negative breast cancer metastasis through reduction of lipid-droplet abundance and stimulation of fatty acid oxidation. Proc. Natl. Acad. Sci. U. S. A. 114, E6556–E6565.

Wu, C.H., Van Riggelen, J., Yetil, A., Fan, A.C., Bachireddy, P., Felsher, D.W., 2007. Cellular senescence is an important mechanism of tumor regression upon c-Myc inactivation. Proc. Natl. Acad. Sci. U. S. A. 104, 13028–13033.

Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M.L., Visvader, J.E., Smyth, G.K., 2010. ROAST: Rotation gene set tests for complex microarray experiments. Bioinformatics 26, 2176–2182.

Xin, M., Kim, Y., Sutherland, L.B., Murakami, M., Qi, X., McAnally, J., Porrello, E.R., Mahmoud, A.I., Tan,

W., Shelton, J.M., Richardson, J.A., Sadek, H.A., Bassel-Duby, R., Olson, E.N., 2013. Hippo pathway effector Yap promotes cardiac regeneration. Proc. Natl. Acad. Sci. U. S. A. 110, 13839–13844.

Xu, J., Chen, Y., Huo, D., Khramtsov, A., Khramtsova, G., Zhang, C., Goss, K.H., Olopade, O.I., 2016. β-catenin regulates c-Myc and CDKN1A expression in breast cancer cells. Mol. Carcinog. 55, 431–439.

Xu, J., Chen, Y., Olopade, O.I., 2010. MYC and Breast Cancer. Genes Cancer 1, 629–640.

Xu, J., Prosperi, J.R., Choudhury, N., Olopade, O.I., Goss, K.H., 2015. B-Catenin Is Required for the Tumorigenic Behavior of Triple-Negative Breast Cancer Cells. PLoS One 10, 1–11.

Xue, J.M., Liu, Y., Wan, L.H., Zhu, Y.X., 2020. Comprehensive Analysis of Differential Gene Expression to Identify Common Gene Signatures in Multiple Cancers. Med. Sci. Monit. 26, e919953.

Yarden, Y., 2001. The EGFR family and its ligands in human cancer: Signalling mechanisms and therapeutic opportunities. Eur. J. Cancer 37, 3–8.

Yarden, Y., Sliwkowski, M.X., 2001. Untangling the ErbB signalling network. Nat. Rev. Mol. Cell Biol. 2, 127–137.

Yi, Y.W., Park, J.S., Kwak, S.J., Seong, Y.S., 2015. Co-treatment with BEZ235 enhances sensitivity of BRCA1-negative breast cancer cells to olaparib. Anticancer Res. 35, 3829–3838.

Yip, K.W., Reed, J.C., 2008. Bcl-2 family proteins and cancer. Oncogene 27, 6398–6406.

Yoshida, K., Miki, Y., 2004. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer Sci. 95, 866–871.

You, J.S., Jones, P.A., 2012. Cancer Genetics and Epigenetics: Two Sides of the Same Coin? Cancer Cell 22, 9–20.

Yu, F.X., Guan, K.L., 2013. The Hippo pathway: Regulators and regulations. Genes Dev. 27, 355–371.

Yudell, M., Roberts, D., DeSalle, R., Tishkoff, S., 2016. Science and society: Taking race out of human genetics. Science (80-.). 351, 564–565.

Zaha, D.C., 2014. Significance of immunohistochemistry in breast cancer. World J. Clin. Oncol. 5, 382–392.

Zaky, S., Lund, M., May, K., Hair, B., Godette, K., Butler, E., Holmes, L., O'Regan, R., Landry, J., 2009. The Triple Threat of Recurrence after Breast Conserving Therapy: Race, Receptor Status and Age. Cancer Res 69, 6045.

Zanconato, F., Cordenonsi, M., Piccolo, S., 2016. YAP/TAZ at the Roots of Cancer. Cancer Cell 29, 783–803.

Zhang, B., Fackenthal, J.D., Niu, Q., Huo, D., Sveen, W.E., DeMarco, T., Adebamowo, C.A., Ogundiran, T., Olopade, O.I., 2009. Evidence for an ancient BRCA1 mutation in breast cancer patients of yoruban ancestry. Fam. Cancer 8, 15–22.

Zhang, J., Antonyak, M.A., Singh, G., Cerione, R.A., 2013. A Mechanism for the Upregulation of EGF Receptor Levels in Glioblastomas. Cell Rep. 3, 2008–2020.

Zhang, J., Fackenthal, J.D., Huo, D., Zheng, Y., Olopade, O.I., 2010. Searching for large genomic rearrangements of the BRCA1 gene in a Nigerian population. Breast Cancer Res. Treat. 124, 573–577.

Zhang, L.L., Liu, J., Lei, S., Zhang, J., Zhou, W., Yu, H.G., 2014. PTEN inhibits the invasion and metastasis of gastric cancer via downregulation of FAK expression. Cell. Signal. 26, 1011–1020.

Zhang, P., Yin, Y., Mo, H., Zhang, B., Wang, X., Li, Q., Yuan, P., Wang, J., Zheng, S., Cai, R., Ma, F., Fan, Y., Xu, B., 2016. Better pathologic complete response and relapse-free survival after carboplatin

plus paclitaxel compared with epirubicin plus paclitaxel as neoadjuvant chemotherapy for locally advanced triple-negative breast cancer: A randomized phase 2 trial. Oncotarget 7, 60647–60656.

Zhao, B., Lei, Q.Y., Guan, K.L., 2008. The Hippo-YAP pathway: new connections between regulation of organ size and cancer. Curr. Opin. Cell Biol. 20, 638–646.

Zhao, B., Li, L., Lei, Q., Guan, K.-L., 2010. The Hippo–YAP pathway in organ size control and tumorigenesis: an updated version. Gen Devel 24, 862–874.

Zhao, J., Guan, J.L., 2009. Signal transduction by focal adhesion kinase in cancer. Cancer Metastasis Rev. 28, 35–49.

Zheng, J., 2012. Energy metabolism of cancer: Glycolysis versus oxidative phosphorylation (review). Oncol. Lett. 4, 1151–1157.

Zinkel, S., Gross, A., Yang, E., 2006. BCL2 family in DNA damage and cell cycle control. Cell Death Differ. 13, 1351–1359.

Zörnig, M., Hueber, A.O., Baum, W., Evan, G., 2001. Apoptosis regulators and their role in tumorigenesis. Biochim. Biophys. Acta - Rev. Cancer 1551.

Zujewski, J.A., Rubinstein, L., 2017. CREATE-X a role for capecitabine in early-stage breast cancer: an analysis of available data. npj Breast Cancer 3, 1–4.

APPENDIX A

Manuscript submitted to *Breast Cancer Research and Treatment*

(under review)

Differential expression between African-ancestry and White patients diagnosed with Triple-Negative Breast Cancer: EGFR, Myc, Bcl2 and β-Catenin as ancestry-associated markers

Short title: Differential expression between African-ancestry and White patients diagnosed with TNBC

Ana T. Matias[1], Ana Jacinta-Fernandes[2,3], Ana-Teresa Maia[2,3,4], Sofia Braga[1,5,6], António Jacinto[1], M. Guadalupe Cabral*[1], Patrícia H. Brito*[7,8]

1 - iNOVA4Health, CEDOC, NOVA Medical School, NMS, Universidade Nova de Lisboa, Lisbon, Portugal

2 - Centre for Biomedical Research, CBMR, Universidade do Algarve, Faro, Portugal

3 - Algarve Biomedical Center, ABC, Universidade do Algarve, Faro, Portugal

4 - Department of Biomedical Sciences and Medicine, DCBM, Universidade do Algarve, Faro, Portugal

5 - Hospital Professor Doutor Fernando Fonseca, HFF, Amadora, Portugal

6 - Instituto CUF de Oncologia, Lisbon, Portugal

7 - UCIBIO, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, Caparica, Portugal

8- Nova Medical School, NMS, Universidade Nova de Lisboa, Lisboa, Portugal

*Corresponding Authors: Patrícia H. Brito (phbrito@fct.unl.pt) and M. Guadalupe Cabral (guadalupe.cabral@nms.unl.pt)

Abstract

Purpose Triple-negative breast cancer (TNBC) has a higher incidence, a younger age of onset, and a more aggressive behavior in African-ancestry women. This severity has been assumed to have biological causes. Here we analyzed ancestry-associated differential gene and protein expression between African-ancestry and White TNBC patients, controlling for patients' menopause status and pathological staging.

Methods Differential gene expression analyses (DGEA) were performed using The Cancer Genome Atlas (TCGA)'s RNA-sequencing data. Gene set enrichment analysis (GSEA) and Ingenuity Pathway Analysis (IPA), focusing on network design, highlighted candidate genes, further validated through immunohistochemistry using TNBC's from patients followed in Portugal.

Results TCGA data of 52 African-American and 90 White TNBC patients indicate a lower survival probability of African-Americans, particularly in premenopausal and stage II disease (log-rank p=0.019 and 0.0038, respectively). African-Americans's transcriptomic profile is characterized by downregulated Hippo pathway and cellular community gene sets, and breast cancer gene set upregulation, comparing with White patients. IPA's network design highlighted central roles for EGFR, Myc and Bcl2 genes in discriminating differences in transcription profiles. These genes and β-Catenin, consensually reported to be required in TNBC tumorigenesis, were further validated through immunohistochemistry. African-ancestry patients from TCGA and validation experiments are distinct, yet, both groups had matched differential gene and protein expression profiles, respectively.

Conclusions We found ancestry-associated transcriptomic patterns between African-ancestry and White TNBCs, particularly when controlling for menopause status or staging. EGFR, Myc, Bcl2 and β-catenin gene and protein differential expression matching results in distinct populations suggest these markers as being important indicators of TNBC's ancestry-associated development.

Keyword Triple-negative breast cancer. Ancestry-associated disease risk . African ancestry . Differential gene expression . RNA-sequencing data

Abbreviations

AA - African-American

BC - Breast cancer

DEG - Differential expressed genes

DGEA - Differential gene expression analysis

FC - Fold-change

FDR - False discovery rate

FFPE - Formalin-fixed paraffin-embedded

GSEA - Gene set enrichment analysis

HER2 - Human Epidermal growth factor Receptor 2

HFF - Hospital Professor Doutor Fernando Fonseca

HR - Hormone receptor

IHC - Immunohistochemistry

IPA - Ingenuity Pathway Analysis

KEGG - Kyoto Encyclopedia of Genes and Genomes

MDS - Multi-dimensional scaling

TCGA - The Cancer Genome Atlas

TNBC – Triple-negative breast cancer

Introduction

Triple-negative (TNBC) is the breast cancer (BC) subtype that neither expresses estrogen receptor, nor progesterone receptor, nor has Human Epidermal growth factor Receptor 2 (HER2) amplification. TNBC presents an aggressive behavior, with large locally advanced breast lesion or metastatic disease developing shortly after adjuvant chemotherapy (Collett et al., 2005; Seewaldt and Scott, 2007). TNBC patients' survival rate is dismal (Garrido-Castro et al., 2019; Saraiva et al., 2017) and, only recently, a targeted therapy was approved (Kang and Syed, 2020). Such targeted therapy consists in the combination of the checkpoint inhibitor, atezolizumab, with nab-paclitaxel, and has been approved as target therapy specifically for programmed death-ligand 1 (PD-L1+) advanced TNBCs (Kang and Syed, 2020). Globally, TNBC accounts for 10-20% of invasive BC diagnoses (Carey et al., 2006; Rakha et al., 2007, 2006; Vona-davis et al., 2008). However, TNBC is disproportionately prevalent in African-ancestry women, comparing with women of White/Caucasian descent (P. Boyle, 2012; DeSantis et al., 2019; Huo et al., 2009; Keenan et al., 2015; Newman et al., 2019; Stark et al., 2010). TNBC in African-ancestry women also presents higher mortality rates (Bauer et al., 2007; DeSantis et al., 2019; Honório et al., 2016), a younger age of onset (Carey et al., 2006; Copson et al., 2014; Huo et al., 2009; Lund et

al., 2009; Zaky et al., 2009), and a faster and more aggressive tumor development and relapse after chemotherapy (Copson et al., 2014; Frasci et al., 2009), independently of other risk factors for BC (Stead et al., 2009). While factors for such disparity may include advanced disease stage at diagnosis, socioeconomic status, and lack of access to treatment (Newman, 2017), biological disparities have been suggested as one of the main factors involved in the ancestry-associated BC discrepancy (Grunda et al., 2012; Newman, 2017; Siddharth and Sharma, 2018). Although studies have investigated ancestry-associated differences in TNBC patients concerning gene and protein expression (Huo et al., 2017; Lindner et al., 2013), somatic alterations (Huo et al., 2017; Newman et al., 2019; Omilian et al., 2020; Shimelis et al., 2018), and metabolic processes (Pelicano et al., 2014; Tayyari et al., 2018), implications in TNBC development and clinical setting are not yet clarified.

To identify ancestry-associated molecular differences between African-ancestry and White TNBC patients, taking into consideration patients' menopause status and stage of the disease, with potential impact in patients' prognosis and treatment, we performed differential gene expression analysis (DGEA) using The Cancer Genome Atlas (TCGA) data. To ascertain if the differences observed between patients from TCGA are transposed to clinical samples, we selected three differentially expressed candidates, EGFR, Myc, Bcl2, as well as β-Catenin, known to be important in TNBC tumorigenesis (Dey et al., 2013b, 2013a; Xu et al., 2015), to validate through immunohistochemistry in samples from a Portuguese hospital. Remarkably, despite the geographic and cultural differences of the validation and TCGA populations, the immunohistochemistry results match the bioinformatic analysis.

Methods

Patient populations

From TCGA's BC project data, "TCGA-BRCA", we identified 52 African-ancestry and 89 White patients diagnosed with primary TNBC, who had undergone RNA-sequencing. Patients were diagnosed between 1998 and 2013, and clinical information was submitted to TCGA between August 2010 and March 2015. RNA-sequencing and clinical data were retrieved from TCGA using *TCGAbiolinks* package (Colaprico et al., 2016; Mounir et al., 2019) in R environment (version 3.6.0), as of March 6, 2019. Formalin-fixed paraffin-embedded (FFPE) TNBC samples collected from 12 African-ancestry and 11 White TNBC patients diagnosed between January 2012 and August 2018 at the Hospital Professor Doutor Fernando Fonseca (HFF) were used to independently validate the results obtained with the TCGA cohort.

Differential Gene Expression Analysis

Differential gene expression analysis (DGEA) was performed with *edgeR* (Robinson et al., 2009b), version 3.26.8, using generalized linear model (gml)-*edgeR* pipeline (McCarthy et al., 2012) and a false discovery rate (FDR) cutoff of ≤ 0.05. We applied a filtering threshold of at least 10 counts-per-million in at least two or more patients/libraries in each DGEA. The list of differential expressed genes (DEG) was further narrowed down using *glmTreat()* function by testing whether the differential expression was significantly above a $\log_2$-fold-change (FC) of 1.2, allowing the identification of more biologically meaningful genes. Comparisons between different DEG lists was performed with *UpSetR* package (Conway et al., 2017). Gene set enrichment analysis (GSEA) and Ingenuity Pathway Analysis (IPA, QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis) software were used to identify key regulators and pathways.

For GSEA, *GAGE* "Generally Applicable Gene-set Enrichment" package (Luo et al., 2009) was used to bring Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets into the R environment. *edgeR*'s *fry()* function, based on the ROAST method (Wu et al., 2010), was used to perform GSEA. Gene sets were considered enriched when FDR≤0.05, taking into consideration both up- and downregulated genes in the set.

IPAs' core analysis, using the complete lists of genes and respective FC and FDR obtained from DGEA, allowed the identification of relevant altered networks, as well as the central network regulators, i.e., DEG with more direct and indirect connections in its network, being potential drivers for ancestry-associated biological differences. Ingenuity Knowledge Base (Genes only) was used as reference. Only networks with a score above 30 were considered. A threshold of FDR≤0.05 was applied to highlight DEG in a network.

Immunohistochemistry

Immunohistochemistry (IHC) was performed to validate candidate genes potentially involved in TNBC ancestry-associated discrepancy, using FFPE samples. After deparaffinization, endogenous peroxidase blocking and antigen retrieval, samples were incubated 1h with primary antibody (Primary antibodies and respective dilutions are listed in the Supplementary Table S1), followed by 1h incubation with the secondary antibody (Peroxidase IgG Fraction Monoclonal Mouse Anti-Rabbit IgG, light chain specific, code 211-032-171, Jackson ImmunoResearch). Bright-DAB Kit (VWR) was used for protein detection. Images were acquired using Axio Imager Z2 microscope (Zeiss). Protein expression was evaluated using Fiji ImageJ (Schindelin et al., 2012) software and a semiquantitative approach to assign an H-score (Ishibashi et al., 2003; John et al., 2009). The intensity of protein expression was measured using the scale 0-3, 0 being negative and 3 being very high expression. The H-Score was calculated by multiplying the intensity value with its area, regardless of cellular location. Wilcoxon-Mann-Whitney test was used to compare mean H-score of each protein between African-ancestry and White patients using the R function *wilcox.test()* and GraphPad Prism 6.01 software.

Statistical analysis

Wilcoxon-Mann-Whitney test was used to compare age of diagnosis and survival between groups of patients. Pearson's Chi-squared test with Yates' continuity correction was used to test for statistically significant differences between the percentage of patients of each ancestry regarding staging and menopause status, using *chisq.test()* R function. *survival* (Therneau, 1999) and *survminer* (Kassambara et al., 2019) packages were used to design and test for differences regarding the time of survival between African-ancestry and White TNBC patients according to staging and menopause status. P-values<0.05 were considered statistically significant.

Results

TNBC incidence and patients' survival

Distribution of clinicopathologic features from 940 African-American (AA) and White BC patients from "TCGA-BRCA" project are depicted in Table 1. AAs are diagnosed with BC at earlier age (56.29 vs 58.61 years, p=0.037) and are more likely to develop TNBC than their White counterparts (28.42% vs 11.89%, p<0.0001). Differences in TNBC patients' age, pathological stage and menopause status at

TNBC diagnosis were not statistically significant, as well as months of follow-up, vital status and time to death (Supplementary – Patients).

TNBC patients were divided according to their ancestry and controlled for menopause status and pathological stage of TNBC at diagnosis, since these variables may contribute to molecular disparities in TNBC biology. The survival probability for each group was analyzed up to 12 years follow-up after initial diagnosis (Fig. 1). AA patients tended to have a faster and higher mortality than White patients in line with what have been reported (Fig 1). Interestingly, this tendency is most significant within premenopausal and stage II TNBC patients (Fig. 1b and e, log-rank p=0.019 and 0.0038, respectively).

Differential gene expression analysis

DGEA between patients' groups was firstly explored with multi-dimensional scaling (MDS) plots (Fig. S1). In accordance with previous studies (Huo et al., 2017; Stewart et al., 2013), and despite some outlier results, differences in gene expression profiles are higher between TNBC and other subtypes of BC (Hormone receptor [HR+] and HER2) patients, than between patients diagnosed with HR+ and HER2 BCs (Fig. S1a), regardless of patients' ancestry (Fig. S1b). The MDS plot does not show a clear differential expression profile between AA and White TNBC patients at the first two dimensions (Fig. S1c). However, ancestry-associated discrepancies in disease severity and previous reported differences (Chang et al., 2018; Huo et al., 2017; Keenan et al., 2015; Lindner et al., 2013; Stewart et al., 2013; Sugita et al., 2016), motivated further analyses taking into consideration the data substructure.

The following contrasts were subjected to DGEA between AA and White TNBC patients: 1) All AA vs all White patients; 2) AA vs White patients with matched menopause status; 3) AA vs White patients with matched TNBC stage; 4) AA vs White patients with matched menopause status and TNBC stage.

Figure 2 shows the number of significant DEG (FDR≤0.05) in AA patients compared to matching White patients obtained from the different analyses, as well as the number of matching DEG between different contrasts, with the exception of the contrasts that resulted in 2 or less DEG (Supplementary - Matching DEG). Differential expression among stage II patients resulted in the higher number of DEG (1776). This result highlights the importance of disease's staging in explaining the differences between the two patients' groups. DGEA including all TNBC patients recovered 1122 DEG and DGEA for post-menopause showed 718 DEGs. Furthermore, DGEA for stage III had threefold the DEG than for stage I (94 vs 30, respectively). Despite studies consistently highlighting the earlier age of TNBC onset in African-ancestry patients (Carey et al., 2006; Copson et al., 2014; Huo et al., 2009; Zaky et al., 2009), our DGEA results suggest that differences in the transcription profile between patients groups tend to be greater when controlling for disease's pathological stage than when controlling for menopause status (see Supplementary - DGEA for full DGEA results).

To determine if these differences could be observed in normal mammary tissue, we performed identical DGEA with RNA-sequencing data from normal-adjacent BC tissue from AA and White patients from TCGA (76 DEG, 6 AA vs 105 White). This analysis showed that the large majority of the DEG previously identified are associated with TNBC progression, as most DEG were not observed in normal tissue. In fact, only two genes - CXCL10 and LTB - were found to be simultaneously differentially expressed between AA and White patients in both TNBC and normal tissue.

DGEA results showed that there are indeed gene expression differences to justify TNBC's epidemiological and clinical observations. Thus, we proceeded to refine DGEA results, in order to find potential leading molecules involved in the ancestry-associated TNBC discrepancy.

Gene set enrichment analysis and Ingenuity Pathway Analysis

GSEA were performed to contextualize ancestry-associated biological differences between AA and White TNBC patients taking into account menopausal status and disease's pathological stage. GSEA included gene sets from the following collections: signal transduction, cellular community - eukaryotes and cell motility, cell growth and death, cancer: overview, where the BC gene set was included, as well as the drug resistance - antineoplastic collection, women-specific endocrine system collection gene sets, and gene sets related to immuno oncology, from the immune system collection. Only the groups, which showed significantly enriched gene sets (FDR≤0.05, Fig. 3), were analyzed: all TNBC, post-menopause (independently of the disease stage), stage II (independently of menopausal status) and post-menopause and stage II (see Supplementary – GSEA for complete results). Considering the gene sets' FC net direction among these four groups of patients, as well as their statistical significance, we highlight Hippo signaling pathway (Fig. 3 and Fig. S2a) as the most consistently downregulated enriched gene set, and cellular community and cell motility (Fig. 3 and Fig. S2b-g) as the most downregulated gene set collections in all groups of patients. We also highlight BC gene set (Fig. 3 and Fig. S2h) as being consistently upregulated in AA patients in all comparisons, and MAPK signaling pathway gene set (Fig. 3 and Fig. S2i) as being upregulated specifically in the group with stage II disease patients.

To identify potential driving molecules involved in the ancestry-associated TNBC discrepancy, we also performed IPA's core analysis with focus on network design, using the complete lists of genes obtained through DGEA for each group, except for pre- and postmenopausal stage I and III disease, due to their low number of DEG. IPA'networks consist in maps of known interactions between molecules, where highly interconnected networks are likely to represent significantly altered biological functions. Complete lists of networks and respective central regulators, scores, molecules involved, and top diseases and functions, according to patient's menopause status and/or staging, are listed in Supplementary - Networks file. From these analyses, we highlight EGFR as central network regulator (Fig. S3) due to its high score in the comparisons with all TNBC patients and with patients presenting post-menopause, stage II disease, or both. Specifically, *EGFR* is downregulated in AA patients, comparing with White patients, in these four groups (log_2FC=-2.238, FDR=1.08E-04; log_2FC=-2.911, FDR=9.41E-05; log_2FC=-2.775, FDR=9.85E-05; and log_2FC=-3.739, FDR=4.46E-05, respectively). Furthermore, EGFR is involved in enriched gene sets from the Signal transduction collection, including MAPK, Ras, Phospholipase D, PI3K-Akt, Rap1, FoxO, calcium, ErbB, Jak- STAT and HIF-1 signaling pathways; in gene sets from the Cancer and Drug resistance: antineoplastic collections, including BC, pathways, proteoglycans and microRNAs in cancer and EGFR tyrosine kinase inhibitor resistance; in Cellular community and Cell motility collections, including regulation of actin cytoskeleton, focal adhesion and gap junction gene sets; and in oxytocin and estrogen signaling pathways gene sets from the Endocrine system collection.

We also highlight Myc as a central network regulator (Fig. S4) due to its score in the comparisons with all TNBC patients, postmenopausal and postmenopausal stage II disease patients. Indeed, *MYC* is upregulated in AA patients in these groups (log_2FC=0.754, FDR=2.70E-02; log_2FC=1.051, FDR=1.13E-02; log_2FC=1.279, and FDR=5.16E-03, respectively). Also, Myc is involved in enriched gene sets from the Signal transduction collection, including Hippo, MAPK, Wnt, PI3K-Akt, ErbB, Jak-STAT and TGF-beta signaling pathways; Cancer collection's BC, pathways, proteoglycan, microRNAs and transcriptional misregulation in cancer gene sets; in Cell growth and death collection's cellular senescence and cell cycle gene sets; and signaling pathways regulating pluripotency of stem cells gene set from Cellular community collection.

Finally, we highlight Bcl2 as a highly scored central network regulator (Fig. S5) specifically downregulated in premenopausal AA patients (log_2FC=-2,043, FDR=3,49E-02). No enriched gene sets

were observed with FDR.mixed≤0.05, possible due to the small set of significant DEG observed in this group (83 DEG, Fig. 2). Interestingly, when comparing with normal-adjacent BC tissue from AA patients, *BCL2* is downregulated in TNBC patients, considering all menopause status (log_2FC=-2.19, FDR=1.10E-04, Supplementary - DGEA), suggesting that *BCL2* is downregulated throughout TNBC development, particularly in premenopausal AA patients.

Thus, EGFR, Myc and Bcl2 were selected to be further investigated through IHC in a collection of TNBC samples obtained from a Portuguese hospital. This procedure was performed to ascertain if the differential protein and mRNA expression would match, even though these populations present geographic and cultural differences from the populations sampled in TCGA. Additionally, we investigated β-catenin, encoded by *CTNNB1*, due to the well-described role of Wnt/β-catenin signaling in BC and TNBC tumorigenesis (Dey et al., 2013b, 2013a; Xu et al., 2015) and worse prognosis (Dey et al., 2013b; Geyer et al., 2011; Khramtsov et al., 2010; López-Knowles et al., 2010). Also, *CTNNB1* is downregulated in postmenopausal stage II disease AA patients (log_2FC=-0.935, FDR=1.63E-02), and is involved in the following enriched gene sets: Hippo and Wnt signaling pathways, from the Signal transduction collection; BC gene set; and Cellular community collection's signaling pathways regulating pluripotency of stem cells gene set.

Validation of EGFR, Myc, Bcl2 and β-catenin protein expression

African-ancestry patients from the validation cohort (Table 2) were diagnosed with TNBC at a significantly younger age, compared to White patients (47.75 vs 64.82 years, p=0.007). Hence, premenopausal TNBC cases among African-ancestry patients are in a significantly higher proportion, compared to White patients (58.33% vs 0, p=0.010). None of the other clinical variables had statistically significant differences between the two populations in the validation cohort.

Validation of gene expression through isolation of RNA from FFPE samples followed by quantitative RT-PCR was not possible due to the poor quality of the RNA isolated. Thus, IHC was performed to validate the selected candidates, showing that EGFR, Myc, Bcl2 and β-Catenin proteins are differentially expressed between African-ancestry and White patients (p≤0.05, Fig. 4) and the net direction of protein expression intensity matches with DGEA results. Thus, these observations corroborate that while Myc is upregulated in African-ancestry patients, EGFR, Bcl2 and *CTNNB1*/β-Catenin are downregulated, in comparison with White patients.

Examples of positive and low/negative staining between African-ancestry and White patients for each of the tested proteins are shown in Figure 5.

Discussion

Incidence and clinical differences between TNBC from African-ancestry and White women are expected to be heavily influenced by differences in the transcriptomic profile and by the regulation of pathways associated with tumorigenesis. Identification of ancestry-associated markers should contribute to understand more clearly the features of this disease in both populations. Additionally, these studies may also contribute for patients' prognosis and establishment of more effective targeted therapies, potentially closing the widening mortality gap between these populations. Our study is one of the few to systematically examine ancestry-associated differences at transcriptomic level in TNBC patients, with the particularity of considering patients' menopause status and disease's pathological stage. Furthermore, the subsequent validation of candidate markers through IHC, in sociocultural and geographically distinct patients, make our findings more robust.

"TCGA-BRCA" epidemiological study confirmed a higher TNBC incidence in AA patients, corroborating previous TCGA-based studies (Ademuyiwa et al., 2017; Huo et al., 2017; Keenan et al., 2015; O'Meara et al., 2019; Stewart et al., 2013), but we also show a decreased survival probability in AAs diagnosed specifically at pre-menopause and at stage II disease. Also, although the younger age of onset often described among African-ancestry TNBC patients [6, 14, 18], and observed in the validation cohort (Table 2), DGEA results suggest that differences in ancestry-associated transcriptomic profiles in TNBC patients are more pronounced when taking into consideration the disease staging than when considering patients' age or hormonal differences caused by menopause. The most distinct transcriptional profile between AA and White patients is observed among stage II disease. Also, stage III resulted in a threefold DEG than stage I disease, suggesting that differences in transcriptional profile increase with TNBC progression. GSEA was performed to identify altered pathways and biological processes between AA and White TNBC patients. Among the enriched gene sets observed, we highlighted Hippo signaling pathway as the most downregulated gene set. According to the literature, downregulation of Hippo pathway components (Fig. S2a), including LATS1, may promote the transcription of genes involved in cell proliferation and competition, cell death inhibition, epithelial-to-mesenchymal transition, tumor metastasis, and tumorigenesis (Harvey et al., 2013; Lei et al., 2008; Marti et al., 2015; Piccolo et al., 2014; Zanconato et al., 2016; Zhao et al., 2010). Furthermore, gene sets from the cellular community and cell motility gene collections, are also consistently downregulated in AA patients. The downregulation of gene sets from these KEGG collections suggests that the integrity of cell-cell contacts and actin cytoskeleton organization, fundamental in epithelial tissue homeostasis, are impaired (Bhat et al., 2019; Cavallaro and Christofori, 2004), promoting tumor cells dissociation and subsequent metastasis (Bogenrieder and Herlyn, 2003). Interestingly, BC gene set (Fig. S2h), that includes, among others, EGFR, MYC or CTNNB1 genes, is upregulated in AA patients compared to White patients, suggesting that these women have a higher activation of components and processes involved in BC development, which may be translated into TNBC's faster progression and aggressive behavior. Finally, we also highlight MAPK signaling pathway gene set (Fig. S2i), which is exclusively upregulated in stage II disease AA patients. MAPK cascades regulate a variety of cellular processes, including proliferation, differentiation, apoptosis and stress responses, playing a crucial role in the survival and development of cancer cells (Guo et al., 2020). Thus, DEG involved in MAPK pathway exclusively identified in stage II AA patients, including NTRK1, GADD45C, MAPK12, TRADD or FGFR2, may contribute to the observed worst prognosis of AA TNBC patients when comparing with matching White patients. Following GSEA, IPA's network design, through central regulator identification and network scoring, successfully recognized EGFR, Myc and Bcl2 as important candidate ancestry-associated markers, a result that, together with β-Catenin expression, we corroborated through IHC in an independent cohort.

EGFR is a transmembrane receptor for members of the epidermal growth factor family, reportedly promoting cell proliferation, motility, and survival via activation of various downstream signaling pathways, including MAPK and PI3K-AKT-mTOR (Yarden, 2001). EGFR is frequently reported as overexpressed in several cancers, including TNBC, being associated with poor prognosis, shorter disease-free survival and aggressiveness (Ali and Wendt, 2017; Foidart et al., 2019; Nakai et al., 2016; Nielsen et al., 2004; Park et al., 2014; Viale et al., 2009). Yet, some studies reported EGFR under-expression as being associated with BC progression, metastasis formation (Choong et al., 2007; Wendt et al., 2015), and bad prognosis (Kreike et al., 2010). Kreike's study (Kreike et al., 2010) suggested that EGFR expression in BC patients has a nonlinear relation with disease outcome, with both lower and higher expression associated with worse outcome, compared to intermediate expression levels. Thus, the precise role of EGFR remains puzzlingly elusive, particularly in TNBC development and prognosis in African-ancestry patients. Here, we show that EGFR is downregulated in AA comparing to White patients, therefore contributing to further clarify this issue.

Myc is a transcription factor that regulates up to 15% of human genes. It is involved in cell growth, proliferation, metabolism, differentiation, stress pathways, drug resistance and apoptosis, often associated with sustaining cell growth in many different cancers (Camarda et al., 2016; Dang et al., 2006; Fallah et al., 2017). Oncogenic Myc expression is increased in TNBC, as compared to other BC subtypes (Carey et al., 2018; Horiuchi et al., 2012; Koboldt et al., 2012; Lee et al., 2017), contributing to chemoresistance and poor prognosis (Carey et al., 2018; Lee et al., 2017). Interestingly, Myc has been stated as being overexpressed in TNBC in studies mostly composed by White patients or cell lines (Carey et al., 2018; Horiuchi et al., 2012; Koboldt et al., 2012; Lee et al., 2017), and here we observed that African-ancestry patients have an even higher expression of Myc.

Apoptosis regulator Bcl2 reportedly acts by promoting cell survival instead of driving cell proliferation. Impaired apoptosis may preserve preneoplastic cells of their need for trophic factors, potentially facilitating metastasis to distant sites (Adams and Cory, 2007). Previously, two studies suggested Bcl2 expression as a favorable prognostic marker for HR-positive BCs (Choi et al., 2014; Eom et al., 2016), and other study as an independent indicator of favorable prognosis for all types of early-stage BC (Dawson et al., 2010). Nonetheless, the effects of *BCL2*/Bcl2 low expression in cancer, and its implications in patients' prognosis and treatment, are poorly explored. We observed downregulation of *BCL2* expression specifically in premenopausal AAs comparinh with White patients. IHC also showed that all African-ancestry patients, mostly premenopausal, had low/negative Bcl2 expression. Interestingly, *BCL2* is significantly downregulated in TNBC compared to normal-adjacent BC tissue in AAs (Supplementary - DGEA), suggesting that this transcription impairment may contribute to TNBC development in premenopausal African-ancestry patients. These results are in line with the Abdel-Fatah study (Abdel-Fatah et al., 2013), which showed that Bcl2's negative expression in early primary-TNBCs was significantly associated with high proliferation and increased risk of death and recurrence. Furthermore, they also observed that although Bcl2-negative patients improved survival upon adjuvant-anthracycline combination-chemotherapy treatment, when these patients do not achieve a pathological complete response, their prognosis is worse (Abdel-Fatah et al., 2013). Thus, Bcl2 expression may be a promising prognosis marker for African-ancestry TNBC patients, particularly at younger ages.

Finally, β-Catenin plays a central role in canonical Wnt signaling pathway and as an intermediate in other signal transduction pathways, including PI3K-AKT pathway. The role and impact of *CTNNB1*/β-catenin expression and misregulation in BC have been massively studied (Bànkfalvi et al., 1999; Chung et al., 2004; Dillon et al., 1998; Dolled-Filhart et al., 2006; Fanelli et al., 2008; Karayiannakis et al., 2001; Lin et al., 2000; López-Knowles et al., 2010; Nakopoulou et al., 2000; Prasad et al., 2008), following the crescent number of studies about Wnt/β-catenin signaling targeting as a therapeutic target in many tumors, especially colorectal carcinomas (Cheng et al., 2019; Krishnamurthy and Kurzrock, 2018; O'Toole et al., 2013; Shang et al., 2017). In TNBC, enrichment of canonical Wnt signal pathway and β-catenin expression is reportedly required for TNBC development by controlling numerous tumor-associated properties, including migration, stemness, anchorage-independent growth and chemosensitivity (Xu et al., 2015), being associated with poor clinical outcome (Dey et al., 2013b, 2013a; Geyer et al., 2011; Khramtsov et al., 2010; López-Knowles et al., 2010). In this study, we observed that African-ancestry TNBC patients express less *CTNNB1*/β-catenin compared to White patients, despite the usual worst prognosis in this population, suggesting that stratifying patients by ancestry is relevant to better understand the role of *CTNNB1*/β-catenin in TNBC.

Remarkably, even though DGEA were performed using USA patients' RNA-sequencing data and IHC in patients' samples from Portuguese hospitals, gene and protein differential expression show equivalent results. AA population is predominantly of Niger-Kordofanian/western Africa ancestry (Bryc et al., 2010; Tishkoff et al., 2009), while 8/12 African-ancestry patients followed in Portugal are from Western Africa countries, namely Cape Verde, São Tomé and Guinea-Bissau (Supplementary - Patients file). These observations suggest that at least patients with Western Africa ancestry, besides having an higher

incidence of TNBC (Huo et al., 2009; Newman et al., 2019), also share some TNBC-associated features, distinct from White patients.

Our results reinforce that TNBC is a heterogeneous disease with ancestry-associated biological features that should also be considered in addition to the molecular disparities between normal vs BC or TNBC vs other BC. African-ancestry patients are less represented in clinical trials (Haynes-Maslow et al., 2014; Owens et al., 2013; Shavers and Brown, 2002), but our study shows how important it is to include a diverse group of participants in tumor marker, cell-based and clinical trial studies. Importantly, because human populations are genetically heterogeneous and lack clear-cut genetic boundaries (Serre and Pääbo, 2004; Tishkoff et al., 2009), it is wise to considerer that although we suggest EGFR, Myc, Bcl2 and β-catenin expression as ancestry-associated markers in African-ancestry TNBC patients, other African-ancestry patients might have the chromosomic regions where those genes are located, or the chromosomic regions of genes regulating these markers, of other ancestries (Bryc et al., 2010). Nevertheless, although self-described ancestry should not substitute the specific disease's etiology for each individual patient, predictions based on patients' self-described ancestry in clinical settings may be helpful in predicting patients' prognosis and proper treatments, being an easy and inexpensive guiding method.

Acknowledgements

We are grateful to all HFF's BC patients that agreed to participate in this study, as well as the Pathology department for the assistance regarding sample collection. We also thank Ramiro Magno and Joana Xavier, at Universidade do Algarve, and Bruno Alexandre at Instituto de Biologia Experimental e Tecnológica for assistance in this research. We acknowledge Ana Farinho and Telmo Pereira for assistance with the IHC and microscopy, respectively.

Funding

This work was supported by national Portuguese funding through Fundação para a Ciência e Tecnologia - PD/BD/114053/2015, iNOVA4Health – UIDB/04462/2020 and ALGARVE2020 POCI-01-0145-FEDER-022184-"GenomePT" grant.

Conflict of interest

Authors declare no conflicts of interest.

Ethical approval

The Review Boards at NMS|FCM and HFF approved the use of patient samples in this study. All participants gave informed consent and this study was performed in accordance with the ethical standards outlined in the 1964 Declaration of Helsinki. Informed consent was obtained from all individual participants included in TCGA. Information about TCGA can be found at https://www.cancer.gov/about nci/organization/ccg/research/structural-genomics/tcga

Code availability

The authors declare that all the code used in the analyses will be made available from the corresponding authors upon request.

References

All References can be consulted in Chapter VI.

Manuscript Tables

Table 25 Description of TCGA's African-American and White breast cancer patients.[a]

	Black or African-American (n=183)	White (n=757)	P-value
Age at **BC** diagnosis, mean (IQR)	56.29 (47 - 67)	58.61 (49 - 67)	0.037[b]
Tumor type, n (%)			
TNBC	52 (28.42)	90 (11.89)	<0.0001
HER2	7 (3.83)	22 (2.91)	0.684
Hormone receptor	112 (61.20)	589 (77.81)	<0.0001
Indeterminate or not evaluated	12 (6.56)	56 (7.40)	
Age at diagnosis of **TNBC**, mean (IQR)	55.87 (48 - 62.75)	54.23 (46.25 - 62)	0.780[b]
Stage at **TNBC** diagnosis, n (%)			
I	9 (17.31)	19 (21.11)	0.741
II	32 (61.54)	56 (62.22)	>0.999
III	10 (19.23)	11 (12.22)	0.375
IV	1 (1.92)	1 (1.11)	NA
Indeterminate or not evaluated	-	3 (3.33)	NA
Menopause status at **TNBC** diagnosis, n (%)			
Pre (<6 months since LMP, no prior bilateral ovariectomy and not on estrogen replacement)	9 (17.31)	29 (32.22)	0.082
Peri (6-12 months since last menstrual period)	3 (5.77)	1 (1.11)	NA
Post (prior bilateral ovariectomy or >12 months since LMP with no prior hysterectomy)	32 (61.54)	55 (61.11)	0.999
Indeterminate or not evaluated	8 (15.48)	5 (5.56)	NA
Follow-up **TNBC** patients, months (IQR)[c],	36.83 (14.70 - 50.77)	28.79 (5.78 – 44.92)	0.932[b]
Vital Status of **TNBC** patients, n			0.883
Alive	45 (86.54)	80 (88.89)	
Dead	7 (13.46)	10 (11.11)	NA
Time to death of **TNBC** patients, months (IQR)	30.75 (25.13 - 36.80)	42.73 (12.43 - 58.94)	0.743[b]

[a] Some values do not sum to heading totals because of missing data.
[b] Wilcoxon-Mann-Whitney test. Otherwise, the comparisons are by Pearson's Chi-squared test.
[c] Months of follow-up after initial diagnosis.
NA – Not available

Table 2 Description of the validation cohort.

	African-ancestry (n=12)	White (n=11)	P-value
Age at diagnosis of **TNBC**, mean (IQR)	47.75 (37 - 54.25)	64.82 (58 - 70)	0.007[a]
Stage at **TNBC** diagnosis, n (%)			
I	1 (8.33)	4 (36.36)	0.262
II	4 (33.33)	2 (18.18)	0.725
III	2 (16.67)	2 (18.18)	>0.999
IV	4 (33.33)	2 (18.18)	0.725
Indeterminate or not evaluated	2 (16.67)	1 (9.1)	NA
Menopause status at **TNBC** diagnosis, n (%)			
Pre (<6 months since LMP, no prior bilateral ovariectomy and not on estrogen replacement)	7 (58.33)	0	0.010
Post (prior bilateral ovariectomy or >12 months since LMP with no prior hysterectomy)	3 (25)	10 (90.9)	0.006
Indeterminate or not evaluated	2 (16.67)	1 (9.1)	NA
Follow-up **TNBC** patients, mean (IQR)[b]	43 (21 - 58)	57 (45 - 73)	0.131
Vital Status of **TNBC** patients, n (%)			>0.999
Alive	10 (83.33)	10 (90.9)	NA
Dead	2 (16.67)	1 (9.1)	NA
Time to death of **TNBC** patients, months	22 - 33	53	0.667

[a] Wilcoxon-Mann-Whitney test. Otherwise, the comparisons are by Fisher exact test.
[b] Months of follow-up after initial diagnosis.
NA – Not available
IQR – Interquartile range
LMP – Last menstrual period

Manuscript figures

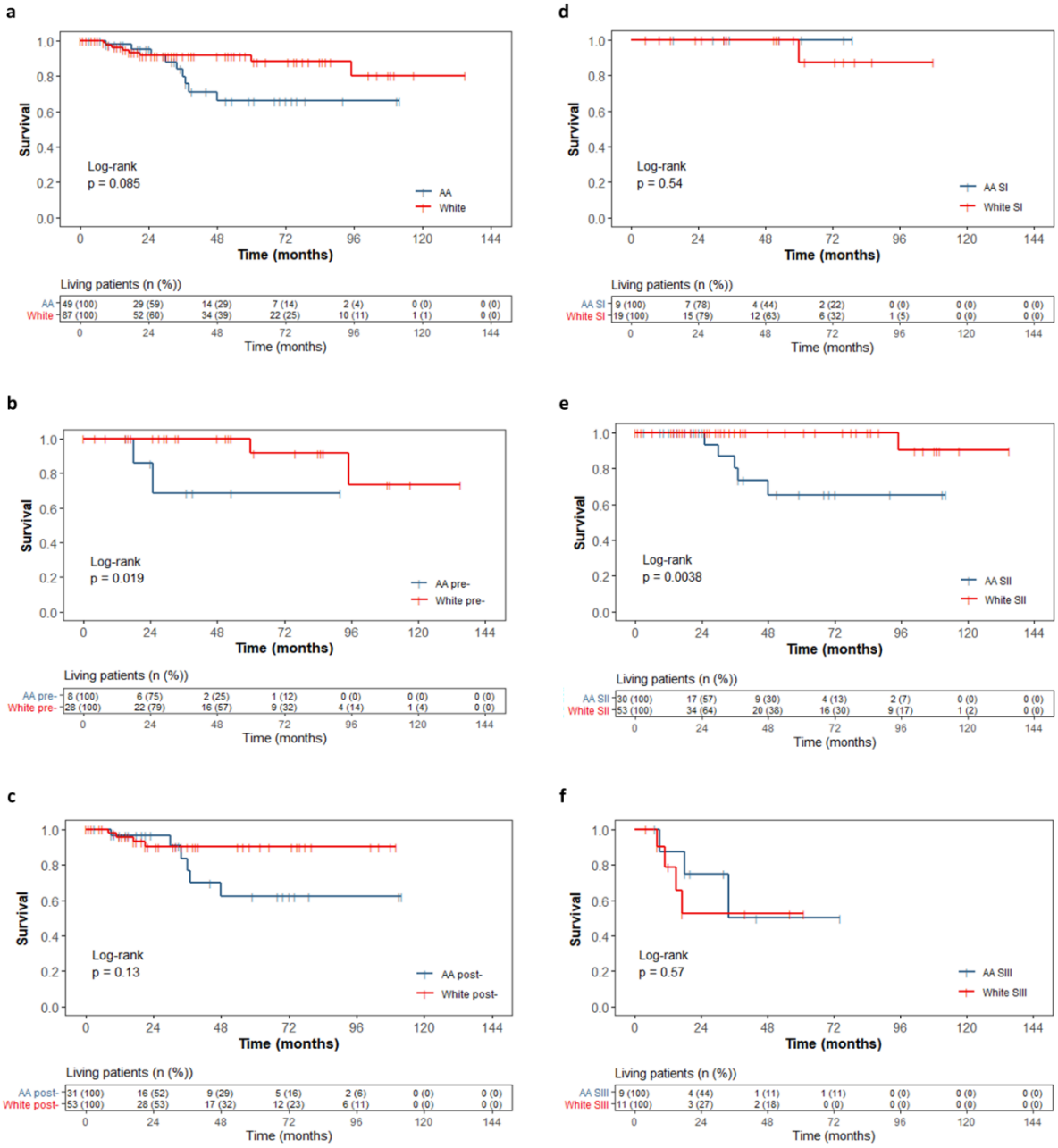


Fig. 1 Kaplan-Meier survival curves of TNBC cases stratified by ancestry up to 12 years follow-up after initial diagnosis. **a** All TNBC patients. **b** Premenopausal (pre-) TNBC patients. **c** Postmenopausal (post-) TNBC patients. **d** Stage I (SI) TNBC patients. **e** Stage II (SII) TNBC patients. **f** Stage III (SIII) TNBC patients. Blue survival curves correspond to African-American patients (AA), red curves show White patients. Values do not sum to heading totals because of missing follow-up information.

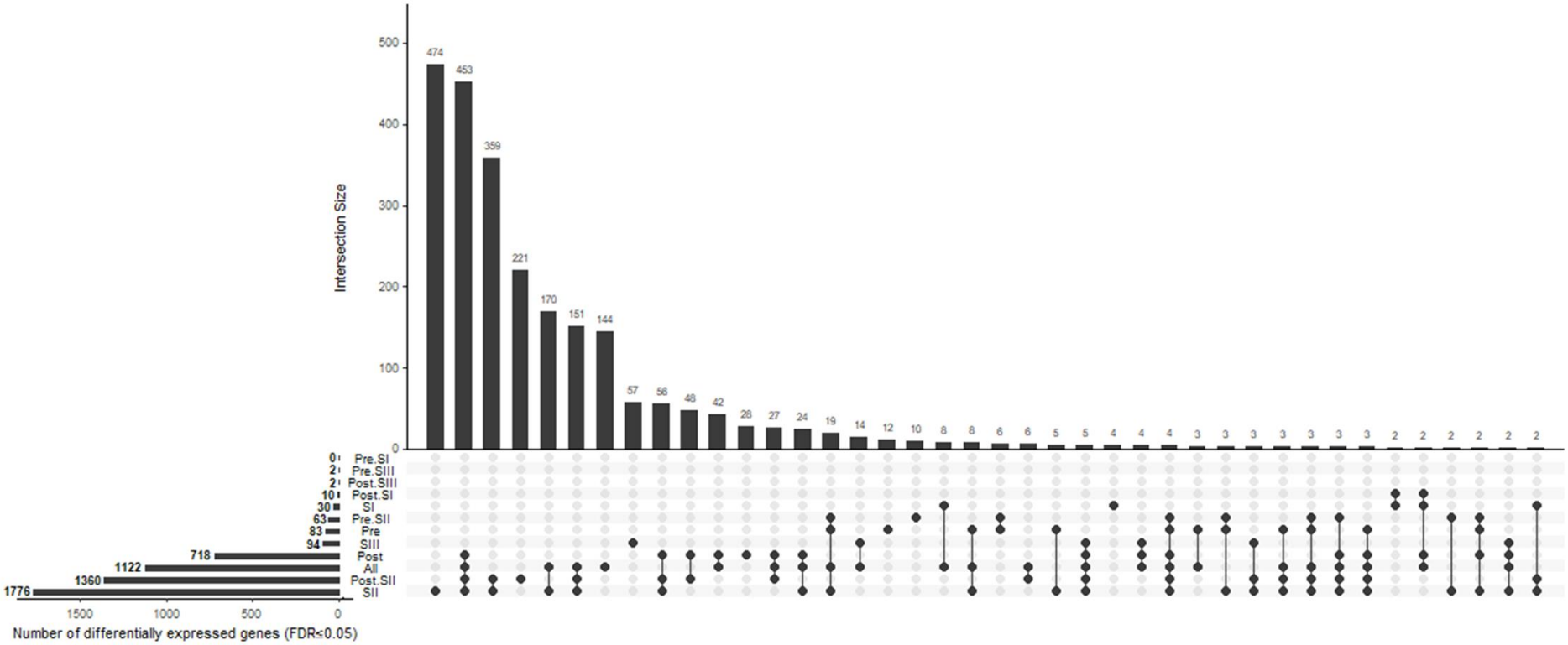


Fig. 2 UpSet plot of the differentially expressed genes (FDR≤0.05) idenfied in TNBC AA patients comparing with TNBC White patients, obtained in the different differential gene expression analyses (DGEA). The intersection size represents the number of matching genes. All - DGEA considering all TNBC patients (52 AA vs 89 White). Pre - DGEA considering premenopausal patients (9 AA vs 29 White). Post - DGEA considering postmenopausal patients (32 AA vs 57 White). SI - DGEA considering stage I disease patients (9 AA vs 19 White). SII - DGEA considering stage II disease patients (32 AA vs 59 White). SIII - DGEA considering stage III disease patients (10 AA vs 12 White). PreSI - DGEA considering premenopausal stage I disease patients (2 AA vs 7 White). PreSII - DGEA considering premenopausal stage II disease patients (5 AA vs 17 White). PreSIII - DGEA considering premenopausal stage III disease patients (2 AA vs 4 White). PostSI - DGEA considering postmenopausal stage I disease patients (3 AA vs 11 White). PostSII - DGEA considering postmenopausal stage II disease patients (22 AA vs 36 White). PostSIII - DGEA considering postmenopausal stage III disease patients (7 AA vs 6 White). AA - African-American.

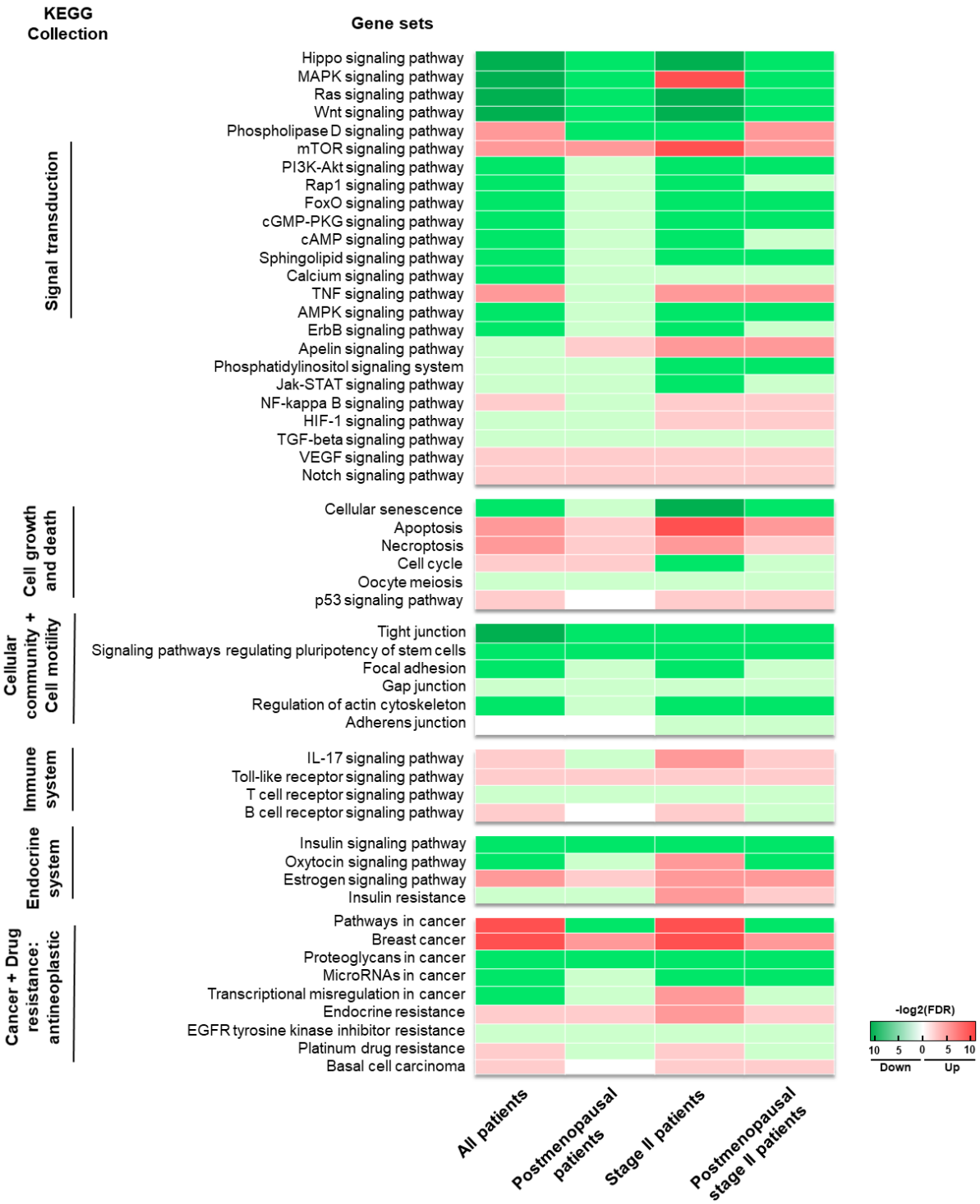


Fig. 3 Enriched gene sets in AA patients. Each column corresponds to a group of patients, in the following order: all TNBC patients, postmenopausal patients, stage II disease patients and postmenopausal stage II disease patients. Green color - negative net direction (Down). Red color - positive net direction (Up). White color - not statistically significant (FDR.Mixed≥0.05). Color intensity is proportional to - $\log_2$(FDR.Mixed) value.

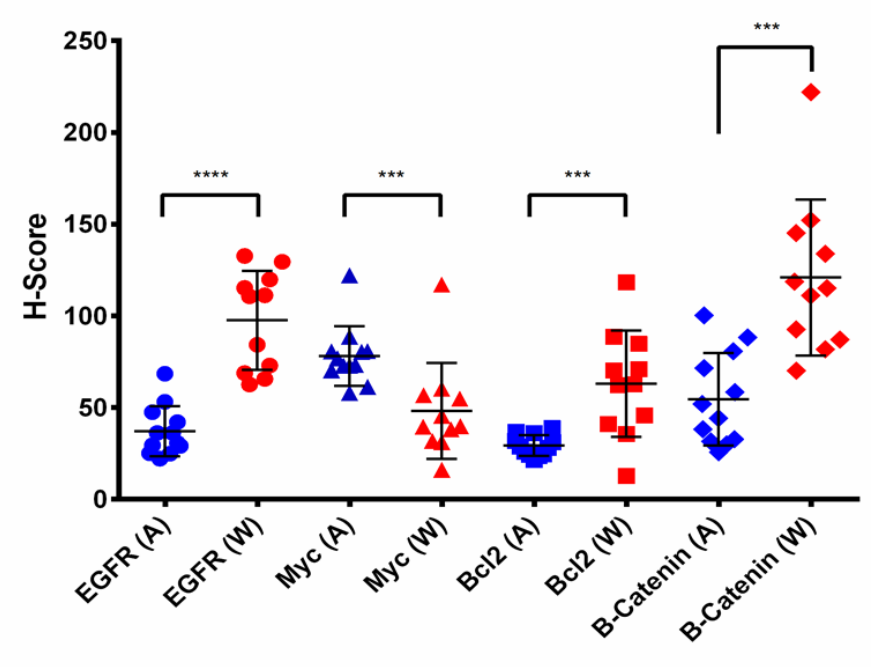


Fig. 4 Differences in H-score between African-ancestry (n=12) and White (n=11) patients are significantly different for EGFR (p=5.92E-06), Myc (p=4.01E-04), Bcl2 (p=9.81E-04) and β-Catenin (p=1.44E-04) expression and have the same net direction as the results obtained in DGEA. A/Blue - African-ancestry patients. W/Red - White patients. Circles - EGFR expression. Triangles - Myc expression. Squares - Bcl2 staining. Diamonds - β-catenin expression.

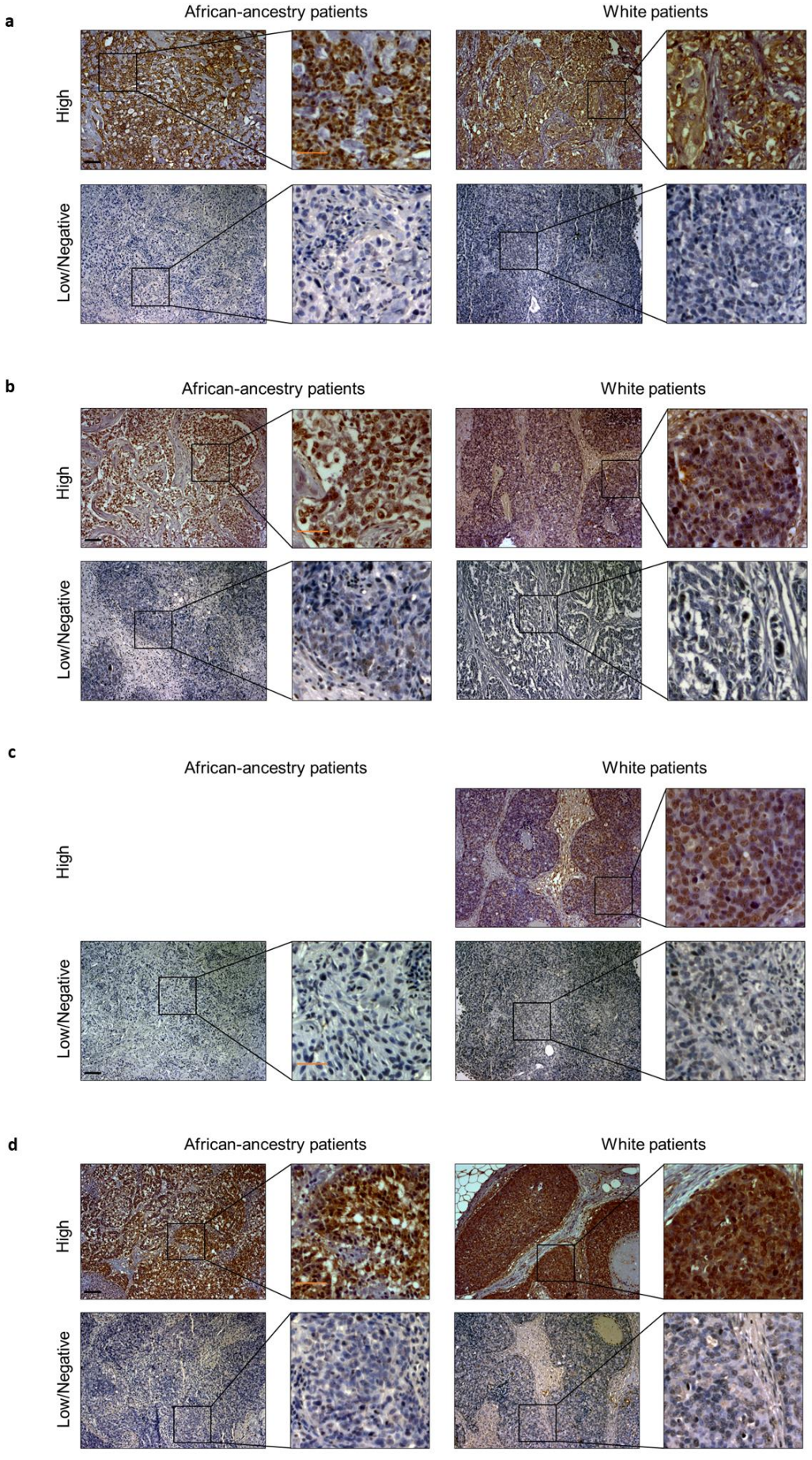

a
African-ancestry patients
White patients
High
Low/Negative
b
African-ancestry patients
White patients
High
Low/Negative
c
African-ancestry patients
White patients
High
Low/Negative
d
African-ancestry patients
White patients
High
Low/Negative

Fig. 5 Immunohistochemistry staining of formalin-fixed paraffin embed TNBC tissue from African-ancestry patients (left) and White patients (right) with **a.** anti-EGFR IgG, **b.** anti-Myc IgG, **c.** anti-Bcl2 IgG and **d.** anti-β-Catenin IgG. Brown color indicates positive reactivity (top), showing protein expression. Black scale bar - 100 µm. Orange scale bar - 50 µm

Manuscript supplementary information

Please consult the Supplementary Information – Manuscript file.