UNIVERSIDADE NOVA DE LISBOA

MScCBBi

MASTER IN
COMPUTATIONAL BIOLOGY
& BIOINFORMATICS

SPECIALIZATION Multi-Omics for Life and Health Sciences

Pedro Filipe Albino Duarte
BSc in Biology

Disclosing trans-omics signatures of gliomas based on network inference and unsupervised learning methods

September, 2024

ITQB NOVA
NOVA SCHOOL OF SCIENCE & TECHNOLOGY
NOVA MEDICAL SCHOOL
NOVA IMS Information Management School
DESDE 1902 INSTITUTO DE HIGIENE E MEDICINA TROPICAL UNIVERSIDADE NOVA DE LISBOA

NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

DEPARTMENT OF
LIFE SCIENCES

CAUSAL INSIGHTS FROM MULTI-OMICS DATA: UNCOVERING GENE NETWORKS AND POTENTIAL THERAPEUTIC TARGETS IN GLIOMA

PEDRO FILIPE ALBINO DUARTE

Bsc in Biology

Adviser: Dr. Marta Belchior Lopes
Assistant Researcher, NOVA School of Science and Technology

Co-adviser: Dr. Roberta Coletti
Researcher, NOVA School of Science and Technology

Examination Committee

Chair: Prof. Dr. Paula Maria Theriaga Mendes Bernardo Gonçalves
Associate Professor, NOVA School of Science and Technology

Members: Prof. Dr. Alexandra Sofia Martins Carvalho
Assistant Professor, Instituto Superior Técnico da Universidade de Lisboa

Dr. Marta Belchior Lopes
Assistant Researcher, NOVA School of Science and Technology

MASTER IN COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

NOVA University Lisbon
September, 2024

Causal insights from multi-omics data: uncovering gene networks and potential therapeutic targets in glioma

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v7.1.18) [1].

Acknowledgements

Despite being a quite solitary work, this master's thesis could not be made without the tremendous support of my supervisors, colleagues, friends and family.

Firstly I would like to address my deepest gratitude to my supervisors, who provided me one of the most fulfilling academic years I could have imagined, and for always believing in me. To Dr. Marta Lopes, thank you for all your remarkable guidance, insightful feedback, unwavering support, and restless availability. Words cannot express how thankful I am for accepting me in this journey, and for all the encouragement and trust you placed me, always propelling me to do better. To Dr. Roberta Coletti, I deeply appreciate your relentless support and availability, constructive feedback and infectious joy through this journey.

To my master's collegues, especially to Diogo, Marta and Lena for still caring for me and enjoying my company despite my prolonged absences and delays. I also want to say thanks to my friends Sara, André, Maria, Carolina, Catarina, Timóteo, Filipe, Mauro, Rodrigo and Ivan for all your companionship and always driving me to be a better person in every aspects. To Adriana, and David that besides the distance are always present and make me feel cared.

To my siblings, Fábio and Sara, thank you for being my anchors, my sources of motivation and the greateast examples, I could ask for. Words cannot convey the love and gratitude I feel toward you both. To my siblings in bond, Gonçalo and Lage for all your incredible support, encouragement and motivation have meant the world to me.

To my dad, thank you for always believing in me, for always being there and offering words of encouragement, even in the toughest times. To my mum, the brightest star in my path, who despite the distance is always there with me.

To my "earthly mothers", Paula and Camélia, your boundless love, encouragement and belief in me have been an endless source of strength. I cannot express how much you both mean to me and how thankful I am for everything you do.

Lastly to Rita, my safe haven and source of happiness, I cannot put to words how much I am grateful for your motivation unconditional support and fondness towards me. Your patience with my forgetfulness means the world to me.

”

“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 Curie, Physicist and chemist

Abstract

Gliomas are the most prevalent primary tumors in the Central Nervous System (CNS), characterized by high heterogeneity and complexity, making subtype classification and biomarker selection challenging yet crucial for improving patient outcomes and therapeutic strategies. Advances in high-throughput technologies have reshaped glioma classification by enabling detailed molecular profiling. However, attending the high complexity and diversity of this disease, it is essential to understand the underlying regulatory systems. To address this complexity, network-based approaches were employed to investigate Gene Regulatory Networks (GRN) across glioma subtypes. Using Integrative Differential Network Analysis in Genomics (iDINGO) for differential analysis across different omics and causal discovery algorithms, particularly the Fast Causal Inference (FCI) algorithm, key regulatory interactions were uncovered. Our analysis identified distinct biomarker networks across glioma subtypes. In astrocytoma, the *HLA* genes (*HLA-DOA*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*) and *CD74*, along with the *CCL* genes (*CCL3*, *CCL4* and *CCL4L2*) subnetwork, emerged as potential biomarkers. In glioblastoma (GBM), the interaction between *S100A8* and *S100A9* appeared to be a significant prognostic marker. The interactions between *ANXA2* and *ANXA2P2*, as well as *COL4A1* and *COL4A2*, were relevant across both astrocytoma and glioblastoma. Moreover, subnetworks involving *MOBP*, *MBP* and *ERMN*; *C1Q* genes (*C1QA*, *C1QB* and *C1QC*), and *GABR* genes (*GABRA1*, *GABRG2*, *GABRB2*) were identified as important across all glioma subtypes. *CYBA* was implicated in both astrocytoma and GBM, while *STX1A* was highlighted as a key biomarker for oligodendroglioma. These findings reveal potential mechanisms driving glioma heterogeneity and highlight promising therapeutic targets specific to each subtype. By integrating network discovery methods, this study provides a deeper understanding of glioma biology.

Keywords: Causal discovery, FCI, iDINGO, Glioma, Multi-omics

Resumo

Os gliomas são os tumores primários mais prevalecentes no Sistema Nervoso Central, sendo caracterizados por uma elevada heterogenidade e complexidade. Assim sendo, tanto a classificação dos diferentes subtipos da doença, como a seleção de biomarcadores, são tarefas desafiantes, embora fundamentais para otimizar estratégias terapêuticas e melhorar o prognóstico dos pacientes. Avanços recentes nas tecnologias de sequenciação de alto rendimento permitiram uma restruturação da classificação dos gliomas através da delineação detalhada dos seus perfis moleculares. Contudo, atendendo à complexidade desta doença, é crucial compreender os sistemas subjacentes que regulam os diferentes subtipos. Para abordar esta questão, foram aplicados métodos baseados em redes para investigar as redes regulatórias de genes nos subtipos de glioma. O iDINGO foi utilizado para efetuar uma análise diferencial de expressão em diferentes ómicas, enquanto algoritmos de descoberta causal, particularmente o FCI, foram empregues para identificar interações regulatórias chave. A nossa análise identificou potenciais redes biomarcadoras específicas para os subtipos de glioma. Relativamente ao astrocitoma, os genes *HLA* (*HLA-DOA*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*) e *CD74*, juntamente com a subrede composta pelos genes *CCL* (*CCL3*, *CCL4* e *CCL4L2*), revelaram-se como potenciais biomarcadores. No glioblastoma, a interação entre *S100A8* and *S100A9* destacou-se como um marcador de diagnóstico promissor. As interações entre *ANXA2* e *ANXA2P2*, tal como *COL4A1* e *COL4A2*, mostraram-se relevantes tanto para astrocitoma como para glioblastoma. Paralelamente, as subredes envolvendo os genes *MOBP*, *MBP* e *ERMN*; os genes *C1Q* (*C1QA*, *C1QB* and *C1QC*), bem como os genes *GABR* foram identificadas como possíveis biomarcadores em todos os subtipos. Por sua vez, o gene *CYBA* foi implicado tanto no astrocitoma como no glioblastoma, enquanto o gene *STX1A* foi destacado como biomarcador chave no oligodendroglioma. Tais descobertas permitem vislumbrar mecanismos responsáveis pela heterogeneidade em glioma, além de apontarem para alvos terapêuticos para cada subtipo. Ao integrar métodos de descoberta de redes, este estudo providencia uma visão aprofundada sobre a biologia dos gliomas.

Palavras chave: Redes de causalidade, FCI, iDINGO, Glioma, Multi-omicas

Contents

List of Figures

List of Tables

Acronyms

α-KG	α-Ketoglutarate *(p. 6)*
2HG	2-hydroxygluterate *(pp. 6, 7)*
2OG	2-oxoglutarate *(p. 6)*
AIC	Alkaike Informatioon Criterion *(p. 33)*
AKT	Serine/Threonine Kinase *(pp. 9, 65, 68, 69)*
ANM	Additive Noise Model *(pp. 16, 25, 26, 34, 40, 42, 62)*
ATRX	alpha-thalassemia/mental retardation and X-linked *(pp. 7, 8)*
BIC	Bayesian Information Criterion *(pp. 14, 15, 32, 33, 40, 61)*
BN	Bayesian Networks *(pp. 12, 13, 15)*
CDK	Cyclin Dependant Kinase *(pp. 7, 8)*
CDKN2A/B	Cyclin Dependent Kinase 2 A/B *(p. 7)*
CIC	Capicua Transcriptional Repressor *(p. 8)*
CNS	Central Nervous System *(pp. iv, 4, 5, 7)*
CPC	Conservative PC algorithm *(p. 18)*
CPDAG	Completed Partially Directed Acyclic Graphs *(pp. 17, 18, 20, 21, 31)*
DAG	Directed Acyclic Graph *(pp. 12–15, 17–21, 25, 31, 32)*
DEG	Differentially Expressed Genes *(p. 9)*
DINGO	Differential Network Analysis in Genomics *(pp. 9, 10, 16, 17, 29, 30, 38)*
EBIC	Extended Bayesian information criteria *(p. 30)*
ECM	Extracellular matrix *(pp. 65, 68)*
EGF	Epidermal Growth Factor *(p. 8)*
EGFR	Epidermal growth factor receptor *(pp. 5, 8, 9, 65)*
FCI	Fast Causal Inference *(pp. iv, v, 13, 16, 19, 20, 32, 40, 42, 44, 61, 70, 71)*

FCM Functional Causal Models *(pp. 15, 16, 21, 62)*
FDR False Discovery Rate *(pp. 35, 37, 40, 42, 62, 210)*
FGES Fast Greedy Equivalent Search *(p. 21)*
FGFR3 Fibroblast Growth Factor Receptor 3 *(pp. 8, 9)*
FN False Negatives *(p. 35)*
FP False Positives *(p. 35)*
FUBP1 Far Upstream Element Binding Protein 1 *(p. 8)*

GAP GTPase Activating Proteins *(p. 9)*
GBM glioblastoma *(pp. iv, 1, 2, 5–10, 29, 30, 38–40, 42, 44–47, 60–71)*
GES Greedy Equivalent Search *(pp. 15, 16, 20, 21, 32, 40, 61)*
GGM Gaussian Graphical Model *(p. 10)*
GO Gene Ontology *(pp. 2, 36, 37, 67)*
GOEA Gene Ontology Enrichment Analysis *(pp. 36, 37, 46)*
GRN Gene Regulatory Networks *(pp. iv, 11, 12, 15, 19–21, 61, 71)*
GSC Glioma Stem Cells *(p. 4)*

HSIC Hilbert Schmidt Independence Criterion *(p. 25)*

ICA Independent Component Analysis *(pp. 21–23, 34)*
IDH Isocitrate Dehydrogenase *(pp. 5–8)*
iDINGO Integrative Differential Network Analysis in Genomics *(pp. iv, v, 16, 17, 30, 37–39, 44–46, 61–67, 69–71)*
IL-6 Interleukin-6 *(pp. 64, 69)*

KCI Kernel-based conditional independence *(pp. 31, 32, 34, 40)*
KEGG Kyoto Encyclopedia of Genes and Genomes *(pp. 2, 36, 46, 64, 66, 68)*

LARS Least-Angle Regression *(p. 33)*
LGG Lower Grade Glioma *(pp. 5, 6, 29, 30, 38, 39, 61, 68)*
LiNGAM Linear Non-Gaussian and Acyclic model *(pp. 16, 21–24, 33, 34, 40, 42, 62)*

MAG Maximal Ancestral Graph *(p. 19)*
MEC Markov Equivalent Class *(pp. 14, 15, 17–20, 26, 31, 32, 34)*
MGMT O^6 -Methylguanine-DNA-Methyltransferase *(pp. 8, 9)*
MHC Major Histocompability Complex *(pp. 62, 63)*
MPC Majority Rule PC algorithm *(pp. 18, 19)*
mTOR Rapamycin *(p. 9)*

NF1 Neurofibromin 1 *(p. 9)*

PAG Partial Ancestral Graphs *(pp. 19, 32)*
PC Peter-Clark *(pp. 13, 16–20, 26, 31, 32, 34, 40, 42, 61, 62)*
PD-1 Programmed Cell Death protein-1 *(p. 63)*
PDGFRA Platelet-Derived Growth Factor Receptor α *(pp. 8, 9)*
PI3K Phosphatidylinositol-3-Kinase *(pp. 9, 65, 68, 69, 71)*
PTEN Phosphatase and Tensin Homolog *(p. 9)*

RFCI Really Fast Causal Inference *(p. 20)*
RKHS Reproducing Kernel Hilbert Space *(pp. 21, 33, 40)*
RTK Tyrosine Kinase *(pp. 8, 9)*

SHD Structural Hamming Distance *(pp. 35, 42, 62)*

TCGA The Cancer Genome Atlas *(pp. 2, 6, 10, 11, 16, 27, 29, 38, 61, 68)*
TERT Telomerase reverse transcriptase promotor mutation *(pp. 5, 7, 8)*
TGF-α Transforming Growth Factor α *(p. 8)*
TLR4 Toll-like receptor 4 *(p. 67)*
TMZ Temozolomide *(p. 9)*
TN True Negatives *(p. 35)*
TP True Positives *(p. 35)*
TP53 tumour protein p53 *(p. 7)*
TPM Transcripts Per Million *(p. 27)*

WHO World Health Organization *(pp. 1, 2, 4, 5, 8, 27, 70)*

1

Introduction

This introductory chapter outlines this research's motivation, primary goals, and objectives and underscores this study's significant role in advancing our understanding of glioma research. It is detailed and framed within a funded project's context, structure, and organization, emphasizing its potential impact on the field.

1.1 Motivation and Goals

Gliomas are the most common Central Nervous System (CNS) primary tumours [2, 3]. Despite aggressive treatments, the overall survival of the patients has not improved. Given that, there is an enormous urge to find new therapeutical targets. The high heterogeneity, diversity, and complexity associated with this disease make it challenging to distinctly classify the different subtypes. Many efforts have been made to classify each patient according to a specific glioma subtype, aiming to improve early diagnosis. Recent advances in high throughput sequencing technology provided different perspectives on the molecular profiles of glioma, allowing, accordingly, changes through time in the classification of the different subtypes of glioma [4]. In 2021, the World Health Organization (WHO) Classification of Tumors of the Central Nervous System revised the classification of glioma. According to this version, glioma is divided into three subtypes: astrocytoma, GBM and oligodendroglioma. These subtypes were classified based mainly on the sample's molecular profiles [5].

To tackle the complexity and heterogeneity of this disease, it is crucial to understand the underlying systems responsible for such attributes. One strategy that can be applied to perform such analysis is to employ network science to model the critical aspects of gliomas. The networks can represent, for instance, gene regulatory networks or gene co-expression networks. Such an approach can provide insights into the mechanistic differences between the different subtypes and elucidate the relevant features and biomarkers of the disease [6].

Undirected network methods have already been applied to uncover he network structure in cancer-related expression data, and glioma is no exception [7]. There is, however,

a lack of studies that focus on retrieving the directed relations from cancer omics data [8].

This research takes a unique approach by utilizing causal discovery algorithms to unravel the transcriptomic and methylomic mechanisms in an individual and integrated approach for each glioma subtype (astrocytoma, GBM, oligodendroglioma). The biological properties of the most relevant subnetworks will be assessed with enrichment analysis, comprising the enrichment of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways terms, providing a comprehensive understanding of the disease.

There is a solid collection of publicly available multi-omics datasets containing cancer samples, such as The Cancer Genome Atlas (TCGA). This accessibility allows one to find patterns and new routes to discover treatments, otherwise impossible, in an in-silico manner. The easy access to glioma patients' data from this database can be crucial in discovering new mechanisms regarding this disease. The data available in TCGA regarding glioma, namely the TCGA-LGG and TCGA-GBM cohorts, is grouped according to the outdated 2007 WHO guidelines [9]. The patients classification was already updated according with the 2021 WHO guidelines, in Mendonça et al. [10].

Thus, considering the updated data from TCGA this investigation aims to find novel biomarkers for glioma, provide insightful information for a better molecular characterization of each subtype, and propose potential therapeutical targets and possible drug applications. This will be achieved by applying causal discovery methods, an almost unexplored method to study the dynamics of molecular biological data features.

1.2 Scope

This work was done under the research project "MONET: Multi-omic Networks in Gliomas" (PTDC/CCI-BIO/4180/2020), funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P. This project aims to identify glioma molecular biomarkers through a multi-omics approach. This research is a result of a collaboration between the Center for Mathematics and Applications (NOVA Math), NOVA Laboratory for Computer Science and Technology (NOVA LINCS), Life and HealthSciences Research Institute of Universidade do Minho (ICVS), Instituto de Telecomunicações (IT) and Instituto de Engenharia de Sistemas e Computadores (INESC-ID).

1.3 Structure

After this section, this dissertation is organized in the following way. Chapter 2 presents an overview of glioma classification and the up-to-date biomarkers that characterize each subtype, the resources to integrate and explore multi-omics data as well as uncovering their underlying causal structure. In Chapter 3, the workflow of this work is described: data collection and preprocessing, the different network discovery methods, and evaluation and validation of the results. Chapter 4 focuses on the results and discussion of this

project. Lastly, Chapter 5 covers the conclusions of this research and future perspectives that can be highlighted.

2

State of the art

This chapter focuses on the evolution of the classification of the different subtypes of glioma, with emphasis on the necessity for new biomarkers to characterize this disease. It also revises the main pathways and differential genetic characteristics that specify each subtype and surveys different network approaches to be applied in the context of this disease, with a more considerable prominence on causal discovery methods. The algorithms, extensions, and biological applications associated with each network method will be discussed.

2.1 Glioma context and evolution of the classification

Brain tumours are neoplasms that occur in the CNS. These tumors are characterized by high mortality and morbidity rates associated with their location and highly invasive growth. Malignant tumours can be classified as primary or secondary [11]. Gliomas are the most common type of primary malignant brain tumours, comprising around 25% of all primary brain tumours and around 80% of all malignant brain tumours in adults [6, 12]. Gliomas arise from glial cells, a type of cells whose primary functions are supporting and protecting the neurons, maintaining homeostasis, and forming myelin sheaths around them. Specifically, those cells can be astrocytes, oligodendrocytes, and ependymal [11]. The high heterogeneity and the capacity of invasive proliferation associated with gliomas substantially increase the difficulty of making therapeutic approaches. Those properties derive from the capacity of Glioma Stem Cells (GSC) to develop tumours in the glial context of varying heterogeneity, accommodating a high potential for self-renewing [13].

The classification of gliomas has been a topic of debate for many years. In 2007, the WHO laid the foundation for diagnosing CNS tumours. That classification was founded on two primary principles: the first was based on tumour typing, where the gliomas were classified based on their similarity to their assumed healthy neuroglial cells of origin, namely astrocytes, oligodendrocytes, or ependymal cells; in contrast, the second one was sustained on a presumed ascendent histologic measure of malignancy that would grade the tumours on a four-tier system (I-IV grade) [2, 11, 14]. This classification was solely

sustained on a histopathological criterion, where all astrocytic phenotype tumours were grouped separately from all oligodendroglial tumours despite clinical disparities inside each group [15]. Nevertheless, recent advances in high throughput sequencing technologies provided a different insight into the molecular profiling of brain tumours, and gliomas were no exception [11, 16]. In that context, a new classification of CNS tumours by the WHO emerged in 2016. This classification aimed at integrating phenotypic and genotypic characteristics, privileging the latter, thus achieving a more objective classification and diagnosis of CNS tumours whilst reducing interobserver subjectivity [11, 15]. According to the 2016 WHO classification, the diffuse gliomas (grades II-IV) were classified based on the mutational status of the Isocitrate Dehydrogenase (IDH) genes (IDH1 and IDH2). These types of gliomas comprised five different entities (Diffuse astrocytoma WHO grade II, IDH-mutant; Anaplastic astrocytoma WHO grade III, IDH-mutant; GBM WHO grade IV, IDH-mutant; Oligodendroglioma WHO grade II, IDH-mutant and 1p/19q-codeleted; Anaplastic oligodendroglioma WHO grade III, IDH-mutant and 1p/19q-codeleted) [17, 18]. Despite the advances from the last classification, the interconnection between histological and genetic criteria still caused some discordant classifications, questioning the integrated application of both approaches [15]. Subsequently, in 2021, the WHO published a revised classification of gliomas, emphasizing the molecular features in tumour diagnosis even more to the detriment of histopathological analysis [19, 20]. The main motive for such revision was the even broader availability of molecular profiling techniques developed since the last classification. One of the techniques fundamental for new molecular considerations in the previous revision was the profiling of methylomic features, given that it provided an insight into the regulation of the genetic traits [5]. Additionally, the new classification shifted from grading with Roman numerals towards using Arabic Numerals, besides crisping the difference between the diffuse gliomas affecting adults and those impacting children. The adult-type diffuse gliomas are now classified based on the status of the mutation status on IDH alongside the 1p/19q-codeletion [21].

In the last published classification, only three entities were considered, namely Astrocytoma (grades 2,3 and 4), which is IDH-mutant; Oligodendroglioma (grades 2,3), being IDH-mutant and 1p/19q codeleted; and GBM (grade 4), IDH-wildtype [5, 21]. GBM, precisely, now differentiates from the IDH-mutant entities, not only from the IDH-mutation status but also from a few histological or genetic features. These include necrosis, microvascular proliferation, *Telomerase reverse transcriptase promotor mutation (TERT)* gene, the amplification of *Epidermal growth factor receptor (EGFR)*, a gain of chromosome 7 and loss of chromosome 10 (+7/-10) [5, 22].

The expected survival is quite variable among the different types of glioma. GBM, the more aggressive type, has a median overall survival of 16 months and real-world 5-year survival of around 7%. Otherwise, Lower Grade Glioma (LGG), both astrocytoma and oligodendroglioma, depending widely on various factors, have an overall survival expectancy between 3 and 15 years with a 5-year overall survival prospect between 58% and 72%. It is relevant to point out that tumours of oligodendroglial origin have a more

significant likelihood of survival and good prognosis than those of astrocytic nature [22–24].

The availability of public datasets such as TCGA was fundamental in advancing the molecular characterization of various cancers, glioma included, providing valuable information through various omics technologies such as genomics, transcriptomics, epigenomics and proteomics [25]. Even though specialized entities continuously revise TCGA data, the data collection has not been updated since 2013. That is particularly relevant in the glioma context given that some glioma cases, in the TCGA data collection, were classified before the 2016 or 2021 WHO guidelines [26]. Despite that, recent studies have updated the clinical classification of glioma samples in TCGA, according to the 2021 WHO guidelines [7, 26].

Overall, it is remarkable to acknowledge the contribution of the evolution of sequencing techniques in the molecular and genetic understanding of gliomas. Such changes are pivotal for classifying and diagnosing different disease subtypes, allowing for a better prognosis and improving the resources to develop personalized treatments.

2.2 Molecular characterization and biomarker profiling of glioma subtypes

Gliomas have already been the subject of study for numerous attempts at therapeutic strategies, but multiple reasons can be attributed to justify the lack of success in such a quest. Multiple intricate reasons, such as tumour heterogeneity, a dynamic tumour microenvironment comprised of neurons, microglia cells, astrocytes, oligodendrocytes and immune cells, as well as a higher level of plasticity compared to most tumour types, all together contribute to therapeutic resistance [27]. With that in mind, there is a need to explore molecular biomarkers individually for each glioma subtype to specify each subtype better and discover candidates for targeted treatments. The remainder of this chapter focuses on already-known biomarkers and pathways of interest for both LGG and GBM.

2.2.1 Astrocytoma

Astrocytomas are preponderantly characterized by their mutation in IDH1 and IDH2 [28]. IDH1 and IDH2 catalyze the metabolic reaction of reversible conversion of Isocitrate to α-Ketoglutarate (α-KG) or 2-oxoglutarate (2OG) and converting NADP+ to NADPH. IDH1/2 reactions are reversible, with the direction of the reaction being determined by the substrate/product concentrations, with both contributing to the maintenance of local and cellular NADPH levels and cellular redox balance [29]. Mutations in IDH1 are mainly present in the hotspot arginine at codon R132, while mutations in IDH2 are found in codon 172. Moreover, it has been reported that these IDH mutants have a neomorphic catalytic action, converting α-KG to 2-hydroxygluterate (2HG), suppressing the histone

lysine methylases [30]. It is noteworthy to mention that some studies report the association between 2HG and tumorigenesis, being this compound, classified as an oncometabolite that affects the methylation pattern of both DNA and histone, the activity of DNA repair proteins, immune response, as well as collagen maturation [31, 32].

Apart from IDH mutation, astrocytoma is also characterized by modifications in tumour protein p53 (TP53) along with *alpha-thalassemia/mental retardation and X-linked (ATRX)* mutations [5]. When active, *TP53* is an important tumour suppressor gene. It acts in response to several internal and external stresses, such as hypoxia, DNA damage, nutrient deprivation, and activation of oncogenes. However, the mutation in *TP53* is prevalent in tumorigenesis, compromising its role in regulating tumour growth and stopping damaged DNA from replicating [5, 33]. Simultaneously, *ATRX* is a chromatin remodelling protein that acts by remodelling chromatin with the requirement of ATP energy to regulate gene expression. The importance of *ATRX* in transcription regulation, nucleosome reorganization, and DNA repair has also been reported. Nevertheless, some articles suggest that a mutation in *ATRX* triggers an abnormal telomere maintenance mechanism, characterized by an alternative lengthening of telomeres, in which cancer cells prevent the telomere shortening [5, 34, 35]. Interestingly, *ATRX* mutations are both mutually exclusive with 1p/19q codeletions and mutations in the promoter of the *TERT* gene, both class-defining modifications of oligodendrogliomas [34, 36].

In addition to the previously mentioned markers, the *Cyclin Dependent Kinase 2 A/B (CDKN2A/B)* homozygous deletion has also been reported to reduce overall survival in patients with astrocytoma, leading to the promotion of tumour proliferation and the loss of cell cycle control [36, 37]. This deletion has also been reported to inhibit the regulation of *Cyclin Dependant Kinase (CDK)* 4 and 6, crucial in cancer cell proliferation through the G1/S phase [37, 38].

2.2.2 Oligodendroglioma

Oligodendrogliomas have a distinctive molecular outline characterized by IDH mutation and the co-deletion of chromosome arms 1p and 19q. The latter is, arguably, a predictor not only for a positive prognostic but also for an excellent response to chemotherapy [5]. Besides, the co-deletion of chromosome arms 1p and 19q has, consequently, a decreased ability to repair the DNA damage capability of tumour cells, which can explain the reason for chemotherapy success: once the tumour cells, after the irradiation, are less likely to survive and proliferate [5, 39]. It is also relevant to note that only a total arm loss of 1p and 19q constitutes a characteristic of oligodendrogliomas since other CNS tumours like GBM can exhibit a partial deletion of either or both arms [40].

Additionally, oligodendrogliomas commonly display mutations in the *TERT* promotor. This promoter is essential in genome stability, as it regulates the telomerase activity. Malfunctions in the normal function of this promotor, specifically an upregulation of *TERT*, increase telomerase activity. This results in an unlimited expansion of telomeres

and, consequently, the immortalization of cells, a crucial attribute for the survival capacity of cancer cells [41, 42]. As referred in astrocytoma, oligodendrogliomas rarely exhibit *ATRX* mutations attending their mutual exclusivity with *TERT* promotor mutations.

Despite the molecular markers previously mentioned, oligodendrogliomas also present mutations in *Capicua Transcriptional Repressor (CIC)* and *Far Upstream Element Binding Protein 1 (FUBP1)*, occurring on 19q and 1p locations of the chromosome, respectively. Moreover, mutations in either of these genes are associated with shorter survival [5, 40]. The *CIC* gene acts by repressing the transcription of genes essential for growth and metabolism, induced by *Tyrosine Kinase (RTK)* signalling. Mutations affecting the *CIC* gene imply mitigation of the repression of *RTK*-induced gene expression, enhancing the proliferation and survival of cancer cells. In turn, *FUBP1* encodes a central regulator of tumour suppression and oncogene alternative splicing beyond playing an essential role in regulating the *MYC* oncogene. A loss of function of the *FUBP1* gene results in modifications regarding RNA splicing besides increased *MYC* expression [40].

2.2.3 Glioblastoma

GBM is the most aggressive type of glioma, being highly heterogeneous and having a high phenotypic and genetic variation, culminating in significant therapeutical challenges [5]. According to the latest WHO classification, GBM is distinguished by the IDH wildtype, related with the highest grade 4, besides being highly proliferative and associated with a very poor outcome [40]. Some morphological features are still used to diagnose GBM, namely microvascular proliferation and necrosis. Some other molecular features like the *TERT* promoter mutation, *EGFR* gene amplification, as well as combined whole chromosome 7 gain and whole chromosome 10 loss and the O^6 *-Methylguanine-DNA-Methyltransferase (MGMT)* promoter methylation status, also characterize this subtype. In addition, some other molecular markers such as *CDK4* amplification, *MET* gene amplification, *Platelet-Derived Growth Factor Receptor α (PDGFRA)* amplification/mutation, and *Fibroblast Growth Factor Receptor 3 (FGFR3)* fusion are also suggestive of GBM prognosis [5, 21].

The *TERT* promotor is very important in the maintenance of genome stability, and the consequences of a mutation in it have already been discussed in astrocytoma and oligodendroglioma. The particular role of *TERT* promoter in GBM is not consensual, as recent research has argued that the negative impact of *TERT* promoter mutation is correlated with the presence of confounding factors like IDH wildtype status, unmethylated *MGMT* promoter status and older age [43].

Otherwise, *EGFR* gene amplification is one of the most relevant markers for GBM prognosis and diagnosis [40, 44]. *EGFR* is a transmembrane tyrosine kinase that is activated by the *Epidermal Growth Factor (EGF)* and *Transforming Growth Factor α (TGF-α)* and other ligands [44]. This gene is part of a family of *RTK*, whose role is vital in signal transduction, thus directly affecting cell differentiation, migration, proliferation, and survival [45]. When amplified, *EGFR* can activate several pathways simultaneously,

directly influencing GBM tumorigenesis [45, 46]. These include the recruitment and activation of *Phosphatidylinositol-3-Kinase (PI3K)* signalling network, reported to account for tumour growth, apoptosis, metabolism and metastasis, as a result of downstream activation of *Rapamycin (mTOR)* and *Serine/Threonine Kinase (AKT)*. Withal, it is relevant to note that the *Phosphatase and Tensin Homolog (PTEN)* gene negatively regulates this pathway [47]. Another relevant pathway in the GBM pathogenesis activated by *EGFR* amplification is the *RAS/MAPK* pathway, which is necessary for cellular growth and division. Although *GTPase Activating Proteins (GAP)*, like *Neurofibromin 1 (NF1)*, negatively regulate the pathway mentioned above [45, 48]. In addition, it is essential to point out that other amplification and/or mutation events relative to *RTK* genes, other than *EGFR*, are pertinent in GBM, such as *PDGFRA*, *MET*, and *FGFR3* [40].

The co-occurrence of the gain of chromosome 7 and the loss of chromosome 10 adds to the intricate complexity adjacent to GBM. This phenomenon is not yet well elucidated, although some studies argue that the gain of chromosome 7 amplifies the effect of *EGFR*, whereas the loss of chromosome 10 affects *PTEN* gene activity [49].

The *MGMT* gene promoter methylation status is also an important predictive marker related to GBM, allowing the prediction of the reaction to chemotherapeutic drugs like Temozolomide (TMZ) [50, 51]. This gene is a critical enzyme in DNA repair and is resistant to the genotoxic effects of alkylating agents [50, 52]. The methylation status of its promoter epigenetically regulates the transcription of this gene. The hypermethylation of the MGMT promoter results in transcription silencing of the MGMT gene, leading to reported survival benefits from TMZ therapy [50–52].

2.3 Undirected network discovery approaches

Undirected network approaches have an essential contribution to disease, not only in unravelling interactions between biological variables, such as relations between genes in a transcriptomic dataset but also in identifying hub genes [53]. Several methods have already been proven to retrieve relevant information from the computation of undirected networks. For instance, the graphical lasso [54], a technique that estimates sparse undirected graphs, is a valuable resource for finding relevant subnetworks and eliminating redundant links between nodes, hence providing a sparse depiction of the interactions. While it does not directly reduce the dimensionality of the data, this approach highlights isolated nodes, which can subsequently be eliminated. With that, the amount of variables that are considered for further analysis is considerably reduced [53, 54]. This method has been proven to be robust when applied to omics data, disclosing relevant networks, such as gene interactions, and dealing with the big dimensionality inherent to omics data [55, 56].

In addition, indirect network analysis can be instrumental to disclose Differentially Expressed Genes (DEG) in different disease conditions. By only considering DEG, relevant information may be discarded regarding co-expressed features in common regulated pathways [53]. With that, Differential Network Analysis in Genomics (DINGO) [57]

was proposed to fill this gap by estimating group-specific and differential networks by calculating group-specific conditional independencies, breaking them into global and local components. This method has already been applied to retrieve differential networks between long-term and short-term survivors in GBM through the analysis of TCGA multiple omics platforms, namely mRNA expression, DNA copy number, methylation and microRNA expression [53, 57]. It was also extended to allow the integration of expression data from multiple platforms, with hierarchical order, on the same samples.

2.3.1 Graphical lasso

Graphical lasso is an efficient method to estimate sparse inverse covariance matrices in high dimension setting [56]. Given a graph G with $X = (X_1, ..., X_p)$ nodes, assuming a multivariate normal distribution $X \sim N_p(\mu, \Sigma)$, with mean μ and theoretical covariance Σ, the graphical lasso finds the precision matrix $\Theta = \Sigma^{-1}$, by solving the log-likelihood problem [54]:

$$\max_{\Theta} \left\{ \log(\det \Theta) - \text{tr}(S\Theta) - \rho \|\Theta\|_1 \right\}, \tag{2.1}$$

where tr is the trace operator, det is the determinant of the precision matrix (Θ) and S the empirical covariance matrix. In addition, the regularization parameter ρ serves as a penalty that induces sparsity and provides a weight on the elements of the Θ matrix. The value of the parameter ρ determines the level of sparsity. Since the graphical lasso is a Gaussian Graphical Model (GGM), the partial correlations are expressed in terms of the precision matrix $\max_{\Theta}$. If $\Theta_{ij} = 0$, it means that there is no partial correlation between X_i and X_j, thus, X_i and X_j are conditionally independent given every other variable in X. It is also relevant to note that every nonzero entry of the Θ matrix is considered as an edge in the undirected network structure [58].

2.3.2 DINGO

DINGO is a correlation-based method that decomposes the conditional dependencies among biological variables firstly into a global component, which represents the interactions that are shared among the groups/conditions in question, and then into a local component that is specific to a particular condition/group [57]. The main objective of this method is to calculate the conditional GGM of given variables. The result is a positive precision matrix $\sigma(x)$, which is the GGM of the data matrix given a vector of covariates concerning the different groups/conditions taken into consideration.

After the calculation of the group-specific GGM, the differential network is constructed. For such, it is set a threshold to measure the differences in conditional dependencies between pairs of variables across different groups/conditions. A differential score is computed with basis on a bootstrap estimate of the standard error to assess the significance of these changes. Considering the edges whose differential score was bigger than the

imposed cutoff, they can be further regarded as conserved or differential. If the conditional dependencies between two variables are consistent across conditions, the edge is labeled as conserved. Otherwise, if the conditional dependencies have different signs in the different groups/conditions, than the edge is considered differential.

2.4 Directed network discovery approaches

The most common way to infer causal relations is to perform interventions or randomized control trials. However, such methods can be impractical to apply due to ethical issues and high costs [59, 60]. On that account, the inference of causal relations through observational data has become increasingly popular for over twenty years [59, 61]. Two different tasks naturally arose, given the potentiality of applying causality to observational data: causal discovery and causal inference. The former tries to infer the causal interactions across the different variables on a given set of data, not assuming, beforehand, any relationship between the variables that are considered. Inversely, the latter tests the relatedness between two variables and further quantifies the effect on each other [59, 62].

These two different approaches can be applied to different biological problems. On the one hand, causal inference can be helpful in verifying causal relations between biological variables with their interaction wholly or partially known, besides providing the possibility of testing interventions on specific variables and their effects on the remainder of the network [59], with Pearl's do calculus [63]. On the other hand, causal discovery is instrumental in unveiling a network of direct relations between biological variables, with no prior knowledge of the nature of their interaction, solely based on molecular expression data [8, 59, 62].

The initial approaches that aimed to discover the causal structure from biological data uncovered the network structure behind gene expression profiles resulting from microarray data dates from more than twenty years ago [64, 65]. From then, the development of high-throughput technologies for sequencing genetic material provided a larger quantity and quality of observational data to apply causal discovery algorithms [66].

Public access to omics data in databases such as the TCGA provides the opportunity to apply causal algorithms not only to single omics but also to integrate genetic information from multiple sources to obtain a broader view of genetic regulation mechanisms [8, 66]. Usually, transcriptomic data is the type of genetic data that is more resorted to infer causal relations from biomedical input, being especially analyzed to infer GRN. A GRN is a hierarchical direct network comprising target genes, regulatory proteins and interactions among them, fundamental in signalling pathways and cellular functionality [60, 67]. When using disease-related data to infer GRN, one can discover disease-associated pathways and genes that can be fundamental for tracking potential targets for drug treatment [67]. In addition, to obtain a more accurate inference of GRN, other types of data can be added to complement the information retrieved from expression data, such as proteomics data, methylation and chromatin accessibility, or transcription factor binding [67].

The study of GRN has already been applied to multiple disease contexts, namely the integration of gene expression and epigenomic data to study causal Interactions in ovarian cancer [68], the analysis of single-cell gene expression data to model GRN related to colorectal cancer [69], or the integration of miRNA and mRNA to investigate GRN in a pan-cancer context comprising seven different cancers [70].

GRN can be modelled with Bayesian Networks (BN), which are graphical representations of probabilistic relations among variables based on Bayesian inference, where the nodes represent random variables and the edges display the conditional independencies [8, 62]. BN are Directed Acyclic Graphs (DAGs), meaning that in a set of variables, there is no directed cycle between them, which conveys that there is no pair of vertices $A \neq B$ comprising a direct edge from A to B and from B to A [62]. Considering the Pearl's definition of DAG, it can be stated as a graphical representation concerning a constrained joint probability distribution of a set of given random variables [63]. Considering p random variables $X = (X_1, ..., X_p)$ with joint distribution $P(X)$, and the marginal distribution of X_i conditional to any subset of variables from X ($S \subseteq X$), being defined as $P(X_i \mid S)$, the random variables X constitute a Bayesian network concerning a DAG $G = (X, E)$, with E being the set of edges, if:

$$P(X) = \prod_{X_i \in X} P(X_i \mid Pa(X_i)) \tag{2.2}$$

where $Pa(X_i)$ is the set of parents of X_i. Moreover, if a probability function P admits the factorization of the equation 2.2 in respect to the DAG G, it can be stated that G represents P and that P and G are compatible, thus it is considered that P is Markov equivalent in relation to G [63]. This Markov condition can be stated as follows [62]:

Definition 2.1 ***Markov Condition:*** *Given a DAG $G(X, E)$, the random variable X_i satisfies the Markov Condition if for every $X_i \in X$, $X_i \perp\!\!\!\perp X \setminus (Des(X_i) \cup Pa(X_i)) \mid Pa(X)$.*

in which Des represents the descendants of X_i, meaning the nodes that are causally influenced by X_i, and Pa the parents of X_i, meaning the nodes that exert a causal influence on X_i. Thus, a variable establishes a conditional independent relation to each of its non-descendants, given its parents. However, the Markov condition by itself cannot discard the occurrence of arbitrary conditional independencies. For that purpose, the concept of d-separation complements this idea [63]. It can be stated as follows:

Definition 2.2 ***d-Separation:*** *Considering three disjoint sets of nodes A, B, C, where i, j, k represent nodes in the graph G, a path $\mathcal{P}$ can be d-separated (or blocked) by a set of nodes C if and only if:*

1. *$\mathcal{P}$ contains a chain $i \rightarrow j \rightarrow k$ or a fork $i \leftarrow j \rightarrow k$ such that the middle node j is in C.*
2. *$\mathcal{P}$ contains an inverted fork (or collider) $i \rightarrow j \leftarrow k$ such that the middle node j is not in C, and there is no descendant of j in C.*

A set C is said to d-separate A from B if and only if C blocks every path from A to B.

The d-Separation concept is intricately connected to the notion of conditional independence. In the case where A and B are d-separated by C in a DAG G, subsequently, A is independent of B conditional on C in every distribution that is congruent with the DAG G. Otherwise, if A and B are not d-separated by C in a DAG G, then A and B are dependent conditional on C in at least one distribution compatible with G [63]. This concept can be formulated as follows:

Definition 2.3 ***d-Separation applicability to probability:*** *For any three disjoint subsets of nodes (A, B, C) in a DAG $\mathcal{G}$ and for all probability functions P:*

1. $(A \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C})_G \implies (\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C})_P$ *whenever $\mathcal{G}$ and P are compatible.*
2. *If $(\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C})_P$ holds in all distributions compatible with $\mathcal{G}$, it follows that $(\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C})_G$.*

where $(\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C})_G$ illustrates the graphical notion of d-separation and $(\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C})_P$ represents the graphical notion of conditional independence. The concept of faithfulness is also relevant in the context of BN, since by reversing the direction of the above implication, the conditionally independent variables become d-separated in the graph [62]. This can be stated in the following manner:

Definition 2.4 ***Faithfulness:*** *For any three disjoint subsets of nodes (A, B, C) in a DAG G and for all probability functions P, the faithfulness assumption is satisfied if: $(\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C})_P \implies (\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C})_G$, whenever $\mathcal{G}$ and P are compatible.*

Lastly, the Causal Sufficiency assumption states that there is no unobserved variables that can directly influence more than one variable in the graph. It means that all the common causes of a given pair of nodes are measured. Although most causal discovery methods rely on this assumption, it is hard to fulfil in practice [71].

Definition 2.5 ***Causal sufficiency:*** *Considering a pair of observed variables X_i, X_j, all their common causes must be observed and modelled in a graph G.*

Bearing in mind the assumptions previously stated, there are two main types of algorithms that perform causal discovery by modelling BN from data, namely constraint-based algorithms and score-based algorithms.

Constraint-based algorithms aim to recover the causal graph by performing multiple conditional independence tests, converting conditional probability independence into a causal graph due to the faithfulness assumption [72, 73]. One example of a constraint-based algorithm is the Peter-Clark (PC), one the most studied causal discovery algorithms, that performs multiple rules to perform a search over the conditional independencies that are possible in all DAG in an efficient way [61]. Taking into consideration the causal sufficiency restriction in the PC algorithm, the FCI algorithm is proposed to broaden the

search space over the DAG by considering unobserved confounders, relying, therefore, on causal insufficiency [74]. One positive aspect of this type of causal search algorithm is that, in theory, it does not make any assumptions about the parametric structure of the causal connections. Nonetheless, the most known statistical tests to evaluate conditional independencies often require Gaussianity or multinomial distributions of the data [72]. In addition, to select the best-suited statistical test to evaluate the conditional independencies, one must take into consideration background knowledge about the distribution of data, which is very hard to obtain in the high-dimensionality context of omics data [60, 72].

Conversely, score-based algorithms are formulated around the maximization of a measure of fitness considering a graph G in a pool of all possible graphs $\mathbf{G}$ for the observed samples $\mathcal{D}$, that follows a given scoring criterion $\mathcal{S}(G, \mathcal{D})$ [72, 73, 75]:

$$G^* = \underset{G \in \mathbf{G}}{\operatorname{argmax}} \mathcal{S}(G, \mathcal{D}), \tag{2.3}$$

with G^* being the graph with the best fitness measure. This model of fitness is usually a scoring criterion, being the most commonly used the Bayesian Information Criterion (BIC) [76]. Given the unfeasibility of performing an exhaustive scoring and iteration over all possible models relative to a DAG, a greedy method must be applied to search in a heuristic and local way for the best global solution [75]. Besides, to achieve the best global solution, the scoring criteria relative to the score-based algorithms must verify some properties [73, 75]:

Definition 2.6 ***Decomposable Score:*** *A scoring criterion $\mathcal{S}(G, \mathcal{D})$ is considered to be decomposable if it can be represented as a sum of the scores considering a node (X_i) and its parents ($Pa(X_i)$), allowing for the computation of the score to be calculated from the local differences of the causal graph [73, 75]:*

$$S(G, \mathcal{D}) = \sum_{X_i \in X} \mathcal{S}(X_i, Pa(X_i), \mathcal{D}). \tag{2.4}$$

Additionally, when considering two DAG G and H, to compare the scores between both, one should only consider the vertices that have different parent sets [73, 75]:

Definition 2.7 ***Equivalent Score:*** *A scoring criterion $\mathcal{S}(G, \mathcal{D})$ is considered to be* ***score equivalent*** *if $\mathcal{S}(G, \mathcal{D}) = \mathcal{S}(H, \mathcal{D})$, for each pair of graphs G and H in the same Markov Equivalent Class (MEC).*

On top of that, it is important to evaluate the local consistency of a score. It is relevant to point out that the concept of a graph G containing a probability distribution P is employed when there exists an independence model related to the graph G that exactly constitutes P [73, 75].

Definition 2.8 ***Locally Consistent Score:*** *Considering a graph G and as an outcome of adding an edge $X_i \rightarrow X_j$ to the graph G, a graph G, the local consistency of a scoring criterion $\mathcal{S}(G, \mathcal{D})$ is verified if and only if:*

1. $X_i \perp_P X_j \mid Pa(X_i) \implies \mathcal{S}(G, \mathcal{D}) < \mathcal{S}(H, \mathcal{D})$

2. $X_i \not\perp_P X_j \mid Pa(X_i) \implies \mathcal{S}(G, \mathcal{D}) > \mathcal{S}(H, \mathcal{D})$

Altogether, the local consistency property of a scoring criterion means that the score increases when an edge is added that corrects an invalid independence constraint, and decreases when an unnecessary edge is added. This ensures an optimal greedy search, as removing irrelevant edges will improve the score [73].

One example of a score-based algorithm that refers to a greedy search procedure is the Greedy Equivalent Search (GES) [75]. Despite arguably outperforming constraint-based algorithms in small sample sizes [77, 78], it is usually constrained to more restrictive parametric assumptions, derivative from its scoring algorithms, such as the linear-Gaussian setting required from BIC criterion [72]. Considering this restriction, some research has already been developed to apply different types of scores to leverage the applicability of this algorithm to non-linear and non-Gaussian data [79].

There are some drawbacks to using BN applied to biological expression data. One of them lies in the necessity of having access to prior knowledge of some of the biological variables' interactions. Without it, regulations like activation/inhibition cannot be visualized in the GRN. Nevertheless, information retrieved from differential expression analysis can be important to infer such regulations [66]. Much of the premises related to BN can be very restrictive to modulate biological networks. For instance, the acyclicity restriction regarding DAGs can be damaging to infer GRN, since it is known that gene networks do contain feedback loops [66, 80, 81]. One relevant aspect of the methods earlier mentioned is that they do not necessarily provide the full causal information of the data, given that the output of such models is the (independence) MEC, a set of causal structures that satisfy the same conditional independencies [59, 63].

Both constraint-based models and score-based models are unable to recover the causal direction between two variables [59]. Consequently, another type of causal search algorithm, Functional Causal Models (FCM), can be applied to unveil such interactions. As well, FCM set apart from the models earlier mentioned by considering that causality can be stated as a deterministic functional equation, in opposition to the probability-wise approach in the constraint-based and score-based models [82, 83]. Taking this into account, the FCM can be stated as the effect Y, as a function of the cause X plus some noise or unmeasurable factors, expressed as follows:

$$Y = f(X, \mathcal{E}; \theta_1), \tag{2.5}$$

considering that $\mathcal{E}$ is the noise term presumed to be independent from X, and the function f defines how Y is generated from X. Besides, θ_1 is the parameter set involved in f. It is assumed that the transformation regarding $X, \mathcal{E}$ to X, Y is invertible in a way that the noise term $\mathcal{E}$ can only be retrieved from the observed variables X and Y [59].

FCM are flexible enough to be adapted to approximate the true data-generating process, even when there is no prior knowledge of the data-generating process. Additionally, the independence between noise and cause must be identifiable in most cases once that it implies the existence of causal asymmetry between X and Y, thus providing causal direction. Given these conditions, it is possible to estimate the causal direction between two variables, as long as they are not influenced by any unknown confounder [59, 84].

All things considered, the algorithms that apply FCM to observational data under some assumptions have been able to produce accurate, unique causal directions with some practical applications. One of the most known and resorted FCM is the Linear Non-Gaussian and Acyclic model (LiNGAM) [85], which considers that f is linear and at most one of the noise term $\mathcal{E}$ and cause X is Gaussian. In non-linear settings, the Additive Noise Model (ANM) [86] assumes a non-linear function f with additive noise.

These methods based on FCM can have a few drawbacks when handling biological data, such as strong assumptions on the distributions and conditions of data, which are hard to verify in a high-dimensional setting [59]. One setting in which FCM cannot identify the causal direction is the linear Gaussian distribution. Despite these drawbacks, FCM can recover the causal direction of the variables in hand in the majority of the cases [59, 87].

In the following subsections, each of the previously described causal discovery algorithms will be discussed: the constraint-based algorithms PC and FCI, the score-based algorithm GES, and the FCM algorithms, such as LiNGAM and ANM. Additionally, iDINGO, derived from the DINGO framework, will be covered in this section. While based on an undirected framework, *iDINGO* encodes direct relationships between methylomic sites and genes via conditional independencies in light of the Markov chain graphs property [88].

2.4.1 iDINGO

iDINGO [89] is a method that expands the scope of the DINGO framework [57] to perform the integration of different omics (such as genomics, proteomics, and epigenomics, among others), providing a deeper biological overview to approach different diseases. This method has already been applied to microRNA, mRNA, and protein features from TCGA breast cancer [89], and microRNA and RNA features from hepatocellular carcinoma data [90].

This framework targets a set of interactions between biological variables coming from different platforms that are differentially connected between patient groups, comprising direct connections between platforms and undirected within platforms. This method assumes that the biological variables follow a multivariate normal distribution, extending the approach up to three matrices containing expression data from whichever genetic data expression platform, comprising all the matrices with the same samples.

The integration between ordered platforms in this method follows the Markov property

for chain graphs [88]. Conditional dependencies among nodes within the platforms result in undirected edges, while those between platforms encode directed edges. This orientation is based on the assumption that the order of the integrated platforms is irreversible (e.g., methylation status influencing gene expression).

iDINGO works as follows: if there are two or three platforms to be integrated, DINGO is applied not only to each platform separately (conditional independencies among the variables encode undirect connections) but also to the concatenation of the platforms (conditional dependencies encode direct relations between different platforms) [89].

2.4.2 PC algorithm

The PC algorithm [61] is programmed to estimate the MEC of the underlying DAG that explains the data, assuming causal sufficiency, the Markov condition, and faithfulness. This algorithm entails Completed Partially Directed Acyclic Graphs (CPDAG), graphs in which both undirected and directed edges are present and do not allow cyclicity in any direction from both types of edges involved [91].

With the purpose of estimating a CPDAG, this algorithm initially estimates the skeleton, i.e. an undirected graph, through observational data by applying conditional independencies tests to eliminate edges. This algorithm can only retrieve the MEC of the DAG and not the DAG itself, because different DAGs can entail the same conditional independencies relationships [92].

Applying conditional independent tests recursively could be infeasible in a high-dimension setting [91, 93]. However, it was already proved that calculating partial correlations to estimate the conditional independencies, in the Gaussian case with the variables containing a multivariate normal distribution, can be a proper solution. For such, Fisher's Z transformation is applied to test if a partial correlation is zero, with a parameter α, the only parameter that can be tuned over the course of the algorithm, as a significance level for testing conditional independencies [91]. It should be pointed out that as the PC-algorithm performs many conditional tests, this parameter α should be solely considered as a tuning parameter and not as an overall significance level, where smaller values of α lead to sparser graphs [93].

The order of the variables in the earlier versions of this algorithm directly conditioned the estimation of the skeleton and, as a consequence, the orientation of the directed arrows further [61, 91]. Such can be due to the evidence that, according to the order of the variables, the edge deletions are different, thus generating different separation sets that will affect the orientation of the edges.

To solve this issue, the PC-stable algorithm implements an order-independent version of the PC algorithm by not pruning any edges until all conditional-independence tests are performed in a given conditioning set [93]. With this, an edge deletion that is performed to the first variables, for instance, at the beginning of a given conditioning set, does not further affect which conditional independencies are tested for other pairs of variables in

the same conditioning set [93].

Given that this step in the PC-stable version requires more tests than the order-dependent versions, the computational running time is longer [94]. After the skeleton finding phase that results in an undirected graph, the next step is to find v-structures. V-structures occur when for a set of three variables A, B, C that are present in the skeleton, considering that A, C is not adjacent and the pairs A, B and B, C are adjacent, the edges $A - B - C$ are oriented as $A \rightarrow B \leftarrow C$, considering the information stored in the conditioning sets. To verify the v-structures in an order-independent context, a variation of the Conservative PC algorithm (CPC) [95] named Majority Rule PC algorithm (MPC) can be employed [93].

The CPC algorithm considers that a triple A, B, C is considered as unambiguous if in at least one separating set (in all subsets Y of the adjacent variables of A) and of the adjacent variables of C, such that $A \perp\!\!\!\perp C \mid Y$ exists and either B is in all separating sets or in none of them; otherwise it is considered to be ambiguous. Considering that, a triple can only be oriented as a v-structure if and only if B is in none of the separating sets [95]. Nonetheless, this approach can be quite restrictive taking into consideration that very few triples can be classified as unambiguous.

With that in mind, MPC performs an alteration in the CPC procedure by labelling the triple A, B, C as unambiguous if at least one such separating set is found and B is not in exactly 50% (or another significant cutoff) of the separating sets. Therefore, according to this algorithm, an unambiguous triple can only be oriented as a v-structure if and only if is in less than 50% of the separating sets [93].

According to the PC-stable implementation, some conflicts between two v-structures can arise. When this happens, a bi-directed edge is created [93]. According to the first implementation of the PC algorithm [61], the rule that outlines the orientation of the v-structures is the first one to be followed to orient the direction of the edges after the skeleton phase. Subsequently, there are three other rules that must be followed to orient all possible edges. These rules can be applied in an independent arrangement and are stated as follows:

Rule 1: Orient $B - C$ as $B \rightarrow C$ if there is a directed edge $A \rightarrow B$ considering that A and C are not adjacent in the graph.

Rule 2: Orient $A - C$ as $A \rightarrow C$ where there is a direct path $A \rightarrow B \rightarrow C$.

Rule 3: Orient $A - B$ as $A \rightarrow B$ in the case where there are two directed paths $A - B \rightarrow C$ and $A - D \rightarrow B$ considering that C and D are not adjacent in the graph.

As a result of these orientation rules, the output is a set of DAG that constitute a MEC. All things considered, the resultant Markov class is a CPDAG where the direct edges are retained if all DAG in the same equivalence class have the same direction for this edge. Inversely, the edge is not oriented and remains undirected [73].

It is relevant to note that despite this algorithm sharing the basis on conditional independencies with conditional independence graphs such as graphical lasso [54], the output and the rationale behind it are inherently different. Conditional independent graphs consider a distinct concept of adjacency than the one from the PC algorithm since the former states that two variables are not adjacent if and only if they are conditionally independent of all the remaining variables. For that reason, the edges in conditional independence graphs are undirected, while in the PC algorithm, they are directed [59].

Both the original PC-algorithm and its derivations, such as the PC-stable, have been implemented in multiple open-source libraries, such as TETRAD in Java [96], pcalg [92] and bnlearn [97] in R, and causal-learn [98] in Python. Furthermore, the PC algorithm has already been applied in the biological context. From gene expression data, it was already applied to infer GRN by X. Zhang et al. [99] and S. H. Mahmoodi et al. [100].

2.4.3 FCI

The FCI algorithm [74] is a variation of the PC algorithm that considers causal insufficiency. It allows for and sometimes discovers unknown confounders in the causal model. Considering these specific conditions, a generalization of a DAG, named Maximal Ancestral Graph (MAG) [101], is needed.

The MAG considers causality with multiple representations of edge marks, representing conditional independence relations by m-separation, a generalization of d-separation that allows for numerous edge representations [93, 101]. When several MAG encode the same conditional independencies, it is said that they are Markov equivalent, thus forming a MEC. The Markov equivalent MAG are designated Partial Ancestral Graphs (PAG), which are the output from the FCI algorithm [93, 101, 102].

Since it is derived from the PC algorithm, both the steps and the improvements applied to the PC algorithm are also relevant in the FCI algorithm context. The orientation rules of the skeleton are identical to the ones covered for the PC algorithm in the previous section, and the order-independence strategy can also be applied to this algorithm [93].

Afterwards, the v-structures are computed according to modification of the PC algorithm. Considering any unshielded triple A, B, C, besides the implementation of the MPC, additional tests are conducted to verify if both A and B and B and C are conditionally independent given a separation set $(A, B) \setminus \{B\}$ found in the skeleton phase. In the case where conditional independence is detected, the edge in question is removed, and a separation set is saved, designated as Possible-D-SEP [74, 93].

With the removal of an edge, new unshielded triples can be either created or destroyed, demanding more conditional independent tests to be performed. Given that, in the case where both conditional dependencies hold and B is not in the separating set for (A, C), the triple (A, B, C) is orientated as a v-structure [93]. After finding the v-structures, the ten-orientation rules from J. Zhang [102] are applied to identify as many edge marks as possible.

Despite its name, this algorithm, in its initial concept, is significantly slower than the PC algorithm, given the additional conditional tests performed. With that in mind, the Really Fast Causal Inference (RFCI) [103] considers establishing a maximum size of the conditioning sets in the possible D-SEP (d-separation) setting to improve the computational size.

This algorithm and its different improvements are implemented in the same software available for the PC algorithm. As seen with the PC algorithm, the FCI can be applied to gene expression data to infer GRN. It can sometimes allow us to discover hidden common causes, which can be more realistic in a biological setting. Nonetheless, it can come at the price of the result being somewhat vague and difficult to interpret [74, 104].

2.4.4 GES

The GES algorithm is a score-based approach to perform causal discovery, which takes a different approach than the one considered in constraint-based models.

In opposition to orienting the edges after finding the skeleton, such as in constraint-based algorithms, the GES algorithm starts from an empty graph, adds recursively needed edges and then removes the edges that are not necessary for a pattern [59, 75].

The assumptions of this algorithm are almost the same as the ones from the PC algorithm. Besides assuming causal sufficiency, acyclicity, and returning a CPDAG that represents the MEC of the model, it is stricter regarding the distribution of the data since the scoring criteria most usually require a Linear-Gaussian distribution of the data. However, since it does not require conditional independence tests to retrieve causal information, this algorithm does not follow the faithfulness assumption, allowing for greater flexibility than the constraint-based algorithms [75, 79, 105].

In its original conception, this algorithm comprises a forward and a backward phase, and later on, a tuning phase has been added [75, 106]. The forward phase is characterized by starting with an initial empty graph and sequentially adding one single edge at each step. Considering all the possible single-edge additions at each step, the one that minimizes a given scoring criterion is the one that is chosen. This phase ends when the scoring of the CPDAG cannot be further improved by single-edge additions. On the other hand, the backward phase starts with the result of the forward phase. It sequentially deletes one edge at a time, selecting the optimal edge deletion at each step and stopping when the score cannot be further improved by a single edge deletion [75, 107].

Generally, in both the forward and backward phases, all DAGs in the MEC of the initial CPDAG are listed first. Then, all DAGs that can be derived through single edge additions or deletions are considered. Finally, the CPDAG corresponding to the DAG with the minimum score is selected [107].

The turning phase is a later addition to this algorithm, allowing, after the backward phase, to invert the directions of the edges until the score cannot be improved [106].

Hence, despite the CPDAG only modifying one edge at a time, the orientations of edges can change multiple times [107].

This algorithm can highlight some issues, namely the difficulty in handling high-dimension settings and the high computation time related to the chosen scoring criterion [73, 105]. Given that, the Fast Greedy Equivalent Search (FGES) minimizes these issues by storing intermediate scores and minimizing the number of computations, therefore allowing the application of these approaches in high-dimension settings. This variant of the GES algorithm uses parallelization to perform the more computationally demanding tasks, improving the computation time [73, 105].

To solve the limitation of a linear-Gaussian data distribution respective to this algorithm, a scoring procedure based on a regression problem in Reproducing Kernel Hilbert Space (RKHS) proved to be an exciting solution to extend the GES algorithm to data with non-linear and non-Gaussian distribution [79].

The basic version of this algorithm, also considering the introduction of the turning phase, can be found in R in the package pcalg [92] and in the package causal-learn [98] in Python. Neither of the two packages have the FGES implementation available. Despite that, the Python implementation already has the regression problem applied in RKHS as a scoring criterion.

This algorithm has already been implemented in a biological context to infer high-dimensional GRN from brain gene expression data [108].

2.4.5 LiNGAM

The LiNGAM model is a FCM that aims to find causal directions in a linear distribution context. In this setting, the causal relation between two variables X and Y can be stated as follows [59]:

$$Y = bX + \mathcal{E} \tag{2.6}$$

where b is a coefficient, $\mathcal{E}$ is a noise term and $\mathcal{E} \perp\!\!\!\perp X$.

Besides linearity, the assumption of non-Gaussianity is fundamental for estimating the causal direction between any two variables since the Gaussian distribution cannot produce any asymmetry between any two variables X and Y [59].

Given these propositions, the original LiNGAM model [85] resorts to the Independent Component Analysis (ICA) [109] to access the data-generating process, and alongside the assumptions of causal sufficiency and acyclicity, it estimates the underlying causal order in which the observed variables are arranged, without assuming faithfulness [85]. That being said, the observed variables can be graphically represented by a DAG. This DAG can be stated in a matrix form as an $m \times m$ adjacency matrix $\mathcal{B} = \{b_{ij}\}$, where every b_{ij} constitutes a connection strength from a variable X_j to other X_i in the DAG [110]. Additionally, another important concept to notice is the causal order k of variables X in the DAG, once that is assumed that no later variable can influence any early variable. Taking

that into account, every observed variable X_i with zero mean and non-zero variance can be calculated according to the following formula:

$$X_i = \sum_{k(j)<k(i)} b_{ij} x_j + e_i, \tag{2.7}$$

where e_i represents an external influence, considered to be continuous random variables characterized by non-Gaussian distributions, with zero means and non-zero variances. Moreover, given the causal sufficiency assumption concerning the model, it is compulsory that the external factors are independent of other external influences [85, 110]. When applying the equation 2.7 to a matrix format, it can be formulated as follows:

$$x = \mathcal{B}x + e, \tag{2.8}$$

in which x is a p-dimensional random vector and $\mathcal{B}$ is a matrix that could be permuted to a strict lower triangularity (lower triangular matrix with the diagonal with all zeros) by simultaneous equal row and column permutations considering the acyclicity constraint [111].

The goal of the LiNGAM algorithm is to estimate the matrix $\mathcal{B}$ given only the observational data x. The relevance in the estimation of the matrix $\mathcal{B}$ lies in the elements b_ij representing the direct causal effect of X_j in X_i. The equation 2.8 can be solved for x in the following way:

$$x = \mathcal{A}e, \tag{2.9}$$

where $\mathcal{A} = (I - \mathcal{B})^{-1}$ is a mixing matrix in which its elements are considered to be mixing coefficients and can be permuted to lower triangularity (not strictly lower triangularity because the diagonal elements are different from zero) [85].

The ICA can estimate $\mathcal{A}$, once that it estimates W and $W = \mathcal{A}^{-1} = I - \mathcal{B}$, despite having sign, scaling and permutation undetermined. The matrix W, which is estimated by ICA, is a result of the formula $W_I CA = \mathcal{P}DW$, where $\mathcal{P}$ is an unknown permutation matrix and D is an unknown diagonal matrix. The key to solving these indeterminacies is to figure out the correct P. According to the LiNGAM algorithm, such can be achieved by assuming that the only correct $\mathcal{P}$ is the one where the diagonal DW has no zeros. That is sustained on the restriction imposed on $\mathcal{B}$ to be strictly lower triangular, bearing in mind that $W = I - \mathcal{B}$. Further, the right sign and scaling are discovered using a unit diagonal (diagonal full of 1) in $W = I - \mathcal{B}$. Subsequently, the computation of W is achieved by dividing the rows of DW by their respective diagonal. The connection strength matrix can then be calculated using the equation $\mathcal{B} = I - W$ [85].

The basic LiNGAM algorithm bears some issues, particularly concerning the application of the ICA algorithm. One is that when the initially guessed state is poorly chosen, the ICA algorithm may not converge in a correct solution in a finite number of iterations [112]. Secondly, depending on different scales or normalizations applied to the variables,

causal ordering of the variables may suffer some modifications, once the permutation operations are not scale invariant [110].

Keeping in mind the adversities associated with the application of the ICA algorithm to the LiNGAM concept, a second approach, named Direct LiNGAM [110] was developed. This algorithm estimates the causal order of the observed variables by sequentially repeating regression analysis and performing independence tests between the predictor and the residual [59]. The authors resort to the Darmois-Skitovitch theorem to justify the necessity of the non-Gaussian distribution.

According to the Darmois-Skitovitch theorem [113], two random variables, y_1 and y_2, can be defined as linear sums of independent random variables $s_i (i = 1, ..., q)$ in the following manner:

$$y_1 = \sum_{i=1}^{q} \alpha_i s_i, y_2 = \sum_{i=1}^{q} \beta_i s_i, \tag{2.10}$$

where, if y_1 and y_2 are independent, then all variables s_j are Gaussian, considering that neither of the constants α_j or β_j are zero. This statement can be reversed to explain that if there is a variable s_j with a non-Gaussian distribution, then y_1 and y_2 are dependent [110].

According to this theorem, there is only one correct causal order, specifically when the explanatory variables and the residuals are independent when the regression is performed between both. Contrarily, a correct causal order cannot be retrieved when the explanatory variables and the residuals are dependent. Moreover, the residuals are the result of a linear regression between all pairs of observed variables and are obtained according to the following formula:

$$r_i^{(j)} = x_i - \frac{\text{cov}(X_i, X_j)}{\text{var}(X_j)} X_j \quad (i, j = 1, \ldots, p; \, i \neq j). \tag{2.11}$$

where cov denotes the covariance between (X_i, X_j, and var the variance of X_j.

Regarding the causal order, the first-order variable is the one that is independent from the residuals resultant from any pair of variables, being, with that, considered as an exogenous variable, where $e_j = X_j$ [110]. It is relevant to point out that more than one variable can be an exogenous or first-order variable. Additionally, the exogenous variable is necessarily considered to be the first-order variable in light of the causal sufficiency assumption in the LiNGAM model [114]. Concerning retrieving the causal order, after finding the parent variable, its influence is removed from the rest of the variables through least-squares regression. With that, the regression procedure mentioned earlier can be performed recursively, finding, in the end, the correct causal order [110].

To evaluate the independence between the explanatory variables and the residuals, one needs a metric that goes beyond just analyzing uncorrelatedness. The least squares regression only evaluates uncorrelatedness, not necessarily evaluating the independence relationships [85]. For that purpose, the Direct LiNGAM method proposes the application

of a nonparametric estimator of mutual information between two variables with kernel methods [110, 115]. Nevertheless, regarding the high computational costs that kernel methods demand, applying one-dimensional entropies denotes a much less intensive computational method to estimate mutual information [114, 116]. Hence, according to this methodology, the mutual information $m(X_i, X_j)$ after standardizing the variables X_i and X_j can be stated as follows:

$$m(X_i, X_j) = \left\{ H(\tilde{x}_j) + H\left(\frac{\tilde{r}_i^{(j)}}{\mathrm{sd}(\tilde{r}_i^{(j)})} \right) \right\} - \left\{ H(\tilde{x}_i) + H\left(\frac{\tilde{r}_j^{(i)}}{\mathrm{sd}(\tilde{r}_j^{(i)})} \right) \right\}, \tag{2.12}$$

where the $\tilde{X}_i$, $\tilde{X}_j$ and $\tilde{r}_i^{(j)}$, $\tilde{r}_j^{(i)}$ designate the standardized observed variables and residuals, respectively. The $\mathrm{sd}(\tilde{r})$ represents the standard deviation of the residual, and H denotes the entropy.

The measure stated in the following equation considers different directions of the regression analysis to compare the values of mutual information. These measures can be further summarized regarding an explanatory variable X_i over all the objective variables X_j in the following manner [114, 116]:

$$M(X_i; U) = -\sum_{j \in U} \min(0, m(x_i, x_j))^2, \tag{2.13}$$

where U is the number of variables to be considered and $\min(0, m(X_i, X_j))$ indicates the smaller between 0 and $m(X_i, X_j)$. As for the causal ordering of the variables, a variable X_i that maximizes the measure $M(X_i; U)$ denotes the most probable variable to be first in the causal order [114].

The runtime of the mentioned applications for LiNGAM algorithms grow exponentially as the number of variables increases, which may be prejudicial in the biological data sets concerning its high dimensionality. The algorithm also assumes that the sample size is bigger than the variable size [110], which can be too restrictive when accounting for biological data that often is characterized by a small number of samples and many more variables [60]. Moreover, the linearity assumption can be too constraining regarding the eventual non-linearity of biological interactions [60]. Lastly, it is relevant to note that this algorithm has already been applied to biological datasets to infer causal relations of health indicators [117].

2.4.6 ANM

Assuming that a linear transformation can fully capture the data-generating processing is often too restrictive, especially when dealing with biological data [59]. Instead, using a model that assumes that a variable is generated through a non-linear process, involving other variables, along with some noise, can be a more effective way to model biological data.

Considering that the ANM [86], in the acyclic context, states that each observed variable X_i is related to a node i in a DAG G and that the X_i value is derived from a function of its parents in G, in addition to some independent additive noise n_i. It can be defined in the following way:

$$X_i := f_i(X_{\mathrm{pa}(i)}) + n_i, \tag{2.14}$$

where f_i is a random function that can be possibly different for every i, $X_{\mathrm{pa}(i)}$ is a vector that has the elements X_j, assuming that there is an edge from j to i in G, and the noise variables n_i are jointly independent and hold arbitrary probability distributions [86, 87]. This model can be only applied in the bivariate context, i.e., to only two variables at a time.

First, to estimate the causal direction between the two variables, it is tested whether the two variables X and Y are independent. Usually, for this procedure, the Hilbert Schmidt Independence Criterion (HSIC) test with a Gaussian Kernel is employed [118]. If the variables are not independent, then it will be tested if the model $Y := f(X) + n$ is consonant with the data. This test is performed with a nonlinear regression of Y on X, and with that the residuals are estimated $\hat{n} = Y - \hat{f}(X)$. Afterwards, the previously mentioned independence test is performed to assess if $\hat{n}$ is independent of X. If the independence is verified, the model $Y := f(X) + n$ is accepted; otherwise, it is rejected. The same test is performed on the reverse model $X := g(Y) + n$ to check if it fits the data.

The results that can be drafted from this analysis are the following: where X and Y are mutually independent, there is no causal interaction between the two variables. Contrarily, when the variables are dependent, both models suit the data. It implies that both models are correct but cannot be retrieved from the data. Considering that, the settings that are not identifiable by the ANM fall under the following categories [87]:

1. X is Gaussian, n is Gaussian and the function f is linear
2. X is log-mix-lin-exp, n is log-mix-lin-exp and the function f is linear
3. X is log-mix-lin-exp, n is lone-sided asymptotically exponential and f is strictly monotonic
4. X is log-mix-lin-exp, n is a generalized mixture of two exponentials and f is strictly monotonic
5. X is a generalized mixture of two exponentials, n is a two-sided asymptotically exponential and f is strictly monotonic

If a model is rejected, meaning it does not fit the data, the other model establishes the causal direction between the variables [86].

This model should not be applied in a high-dimensionality setting, given the high computational cost of testing all possible causal relations within pairs of variables [119]. Therefore, this methodology can be combined with conditional-independence methods

like the PC algorithm, which estimates the MEC, and the ANM further orient the edges that cannot be oriented with the previously mentioned approaches [59, 87].

This approach has been applied in a biological setting, specifically in the causal analysis of genetic features in the context of complex disease [120, 121].

3

Materials and Methods

This chapter describes the data and methodology that were applied in this work. Initially, an overview of the workflow is covered, with the respective topics specified in the following subchapters. Succeeding, the extraction and preprocessing of the data are detailed. Subsequently, all the parameters and evaluation of the undirected, hybrid and directed methods will be described. Lastly, the methods to evaluate the stability and robustness of the causal networks, as well as methodologies to assess the biological meaning of the outputs, are explained.

3.1 Thesis workflow

This subsection depicts in Figure 3.1 the workflow of the project, whose details will be further scrutinized in the following sections.

3.2 Data extraction and preprocessing

The transcriptomic and methylomic data were retrieved from the TCGA-LGG [122] and TCGA-GBM [123] datasets available in TCGA, with the application of the GDCquery R function [124] from the TCGAbiolinks package [7, 125]. Considering that the TCGA data grouped the patients according to the 2007 WHO classification, the procedure applied in Mendonça et al. [26] was implemented so that the samples could be classified according to the WHO 2021 guidelines [7]. The code provided in [55] repository was utilized to perform this step.

The RNA-seq data extracted from TCGA (Illumina mRNA-seq) was standardized according to Transcripts Per Million (TPM) and upper quantile normalization. To accommodate the requirement of normal distributed variables from graphical lasso, the nonparanormal normalization with $huge.npn$ function from the huge R package [126] was employed [7]. To assess which variables had a normal distribution, the Jarque-Bera test was applied [127].

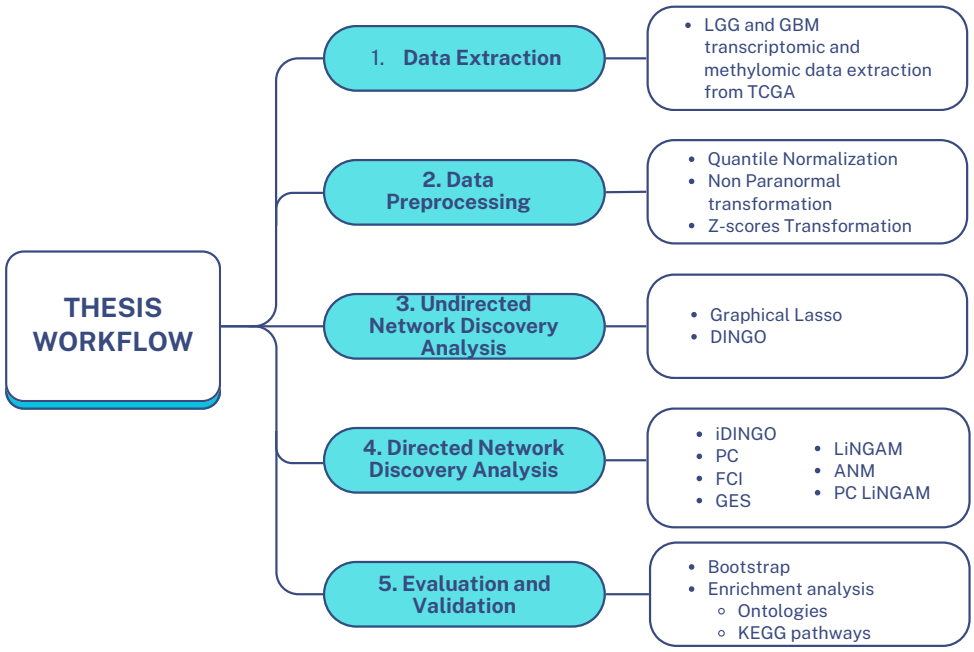


Figure 3.1: **Thesis workflow**: 1. Retrieval of Transcriptomic and Methylomic data from the LGG and GBM projects sourced from the TCGA database. 2. Data standardisation with the Quantile Normalization and the Non-Paranormal transformation. 3. The graphical lasso was employed as a strategy to reduce the data's dimensionality, making the computation of the other algorithms possible. Moreover, besides graphical lasso, another undirected method, DINGO, was employed to perform differential expression analysis according to the different types of samples. 4. Application of the different directed network methods, namely iDINGO, PC-algorithm, FCI algorithm, GES algorithm, LiNGAM, ANM and PC-LiNGAM were applied. 5. A non-parametric bootstrap strategy was performed to assess the causal algorithms' robustness. Lastly, enrichment analysis for ontology and KEGG pathways terms was utilised to assess the biological interpretation of the best-performing causal algorithm.

Concerning the DNA-methylation data (Illumina HumanMethylation450), the probes relative to sexual chromosomes, the ones that had missing (NA) entries and non-valid entries, were removed. No further biological normalization was performed since that the methylomic data comprises values between 0 and 1, representing the unmethylated or methylated status. To harmonize the following analysis of methylomic and gene expression data, only the samples that were common in both datasets for each glioma subtype were considered.

Regarding the astrocytoma individuals, only 249 contained both types of data; in consideration of GBM, scarcely 112 samples shared two kinds of data; and concerning the oligodendroglioma patients, 166 were reported to have gene expression and methylomic data.

3.3 Undirected network discovery analysis

3.3.1 Graphical lasso

Considering the high dimensionality of the data, the graphical lasso [54] was applied to induce sparsity and select the most relevant variables to be analyzed. The network's sparsity degree is controlled by the parameter ρ present in the equation 2.1. This method was applied to the transcriptomic data from TCGA and to the set of CpG sites that were related to the selected transcriptomic dataset.

This methodology was implemented with the $huge.glasso$ function from the huge R package [126]. The regularization parameters were chosen so that around 150 features were selected for each omic layer. The amount of features and the regularization parameters were empirically picked to increase the interpretability of the networks. This step was performed according with [9].

Regarding the RNA samples, all the selected variables were retrieved with a ρ parameter of 0.955. Concerning the methylomics data, for astrocytoma, a ρ value of 0.86 was selected; for GBM, the chosen ρ parameter value was 0.85; while 0.9 was the ρ selected for oligodendroglioma.

3.3.2 DINGO

The DINGO method [57] was utilized to verify the differential networks according to the different subtypes of glioma. Initially, DINGO was applied to assess the differential networks between GBM and LGG samples. For the construction of the bivariate covariate relevant to the calculation of the local component, the GBM samples had a different value (1) than those samples that were classified as astrocytoma and oligodendroglioma (LGG) (-1). In addition, this method was also applied to verify the differential networks between LGG samples, meaning that the astrocytoma samples had a different covariate value (1) than the oligodendroglioma samples (-1).

A bootstrap approach with 100 replicates was performed to retrieve the differential edges. The differential edges were selected based on a p-value of 0.01, meaning that, considering the differential scores in the bootstrap replicate, only if an edge had a p-value inferior to 0.01, was considered differential in the final network. The implementation of this method was executed with the application of the iDINGO R package [89].

One of the steps in this implementation includes the calculation of σ by the application of the graphical lasso [57]. Since that the graphical lasso was already applied to the full dataset, and the inverse covariance matrix was already known, the source code of the iDINGO R package was modified to take into account that particularity. It is relevant to point out that this modification spared computation time. The networks resultant from this method were integrated in the final output of iDINGO, plotted with the Cytoscape software [128].

3.4 Directed network discovery analysis

3.4.1 iDINGO

To perform differential network analysis with the integration of both methylomic data and gene expression data, the iDINGO method was applied [89]. Such as in DINGO method, initially, this method was applied to verify the differential network between GBM samples and LGG samples, and afterwards to assess the differential network within the LGG samples, differentiating astrocytoma samples from the oligodendroglioma samples. The covariates values were the same as the DINGO ones.

This method applies the same rationale regarding the differential score as DINGO, and concomitantly, a p-value of 0.01 was chosen. The iDINGO implementation performs the same steps as DINGO to retrieve the differential networks, the main difference being the application of the workflow to the concatenation of the omics in the study, despite being evaluated the differential networks to each omics separately [89].

For this analysis, it is considered that the CpG sites have a direct influence on the selected genes and not vice-versa. This is applied in consonance with the Markov property of the chain graphs [88]. However, the direction of the edges within each platform cannot be discerned with this approach, and with that, the edges remain undirected. As mentioned in the previous section, the graphical lasso is utilized to estimate the precision matrix. Taking that into consideration, an adaptation of the$idingo$ function from the R package iDINGO [89] was implemented.

To comply with the concatenation of the datasets, a suitable tuning parameter was needed in order to retrieve a sparse network. To select the optimal ρ for the graphical lasso, the Extended Bayesian information criteria (EBIC) criterion [129, 130] was applied. The EBIC can be defined according to the following formula:

$$\text{EBIC} = -2L + E\log(N) + 4\gamma E\log(P), \tag{3.1}$$

where L is the log-likelihood, N represents the sample size, E the number of edges, P the number of nodes, and γ is a hyperparameter (parameter that needs to be set manually and normally regulates other parameters) ranging from 0 to 0.5, with larger values producing sparser graphs [130, 131]. Using a γ value of 0.5, the $huge.select$ function from the $huge$ R package [126] was applied to select the best ρ from 100 candidates generated by the $huge$ function with the $glasso$ method.

The remainder of the implementation followed the approach from the $idingo$ R package [89], and the visualization was performed with the Cytoscape software [128].

3.4.2 PC-algorithm

The PC-algorithm was implemented to retrieve the CPDAG from the different subtypes of glioma, through matrices containing expression data for RNA-seq and methylomics data.

This algorithm assumes that the distribution of the data is faithful to a DAG and estimates the MEC of the DAG, which explains the causal interactions in the data. Despite this assumption, this algorithm does not make any parametric assumption on the distribution of the data. To standardize the distribution of the samples, so that each sample has a mean of 0 and a standard deviation of 1, the Z-scores normalization was applied according to the following formula:

$$Z = \frac{x - \bar{x}}{sd}, \tag{3.2}$$

where x is the observed variable, the $\bar{x}$ the mean of the data and sd the standard deviation. Beyond this algorithm, this standardization was applied to all causal discovery methods that were implemented.

The PC-stable [93] version of the PC-algorithm was employed. This algorithm was implemented with the pc function from the pcalg R package [92].

To meet the requirements of the PC-stable algorithm and to generate an order-independent graph, the $stable$ value for the $skel.method$ parameter was picked. Also, the majority rule was considered according to the argument $maj.rule$ of the pc function.

The PC-stable implementation considers that in situations where conflicts arise in the orientation of v-structures, a bi-directed edge $\leftrightarrow$ must be applied, solving the order-dependence issues [92, 93]. This specificity of the algorithm is taken into consideration in the pc function with the argument $solv.confl$ being set to $TRUE$ and the argument $u2pd$ assuming the value $relaxed$ [92].

The conditional independence test resorted to in this implementation was Fisher's Z transformation, which assumes that the variables follow a Gaussian distribution. The α parameter that controls the sparsity of the graph was set to 0.05.

The pcalg implementation does not provide conditional independencies tests assuming a non-Gaussian data distribution. To account for that setting, this algorithm was also implemented in the Python package causal-learn [98] with the Kernel-based conditional

independence (KCI) test and independence test [132]. Given that this conditional independence test is based on a kernel method, the complexity is cubic in the sample size, so the computation time is expected to be high [98].

3.4.3 FCI-algorithm

The FCI-algorithm has similar principles to the algorithm described previously, but the flexibility in taking into account unobserved variables is valuable in the biological context. This way, this algorithm was applied to extract the PAG from the the expression data relative to the transcriptomics and methylomics datasets.

This algorithm estimates a PAG, which illustrates direct edges that define ancestry relationships and bidirected edges indicating potential unknown confounders. Edges conntaining a circle mark ($* - o$) indicate that within the same MEC, some graphs denote X as an ancestor of Y, while others do not [92, 102]. For interpretability, adjacency matrices usually use 1 to indicate a direct edge, and 0 otherwise. In this situation, circle-marked edges were treated as bidirected edges for further readability and analysis.

The implementation of this algorithm, just as in PC-algorithm is performed in pcalg [92] with the stable variation [93] of the skeleton, resulting in an order-independent graph, and the majority rule is considered to respect this approach. The conditional independence test that was applied, in this implementation was the Fisher's Z test. Similarly to the PC-stable implementation, the α parameter was set to 0.05.

Considering the high dimensionality of the data and the high computational demand of this algorithm, the RFCI [103] was the function chosen from pcalg to apply this procedure. The main difference to the usual FCI implementation from this R package is that it does not compute conditional tests on subsets of Possible-D-SEP sets, reducing the computational complexity [92].

To consider the possibility of a non-Gaussian distribution of the variables, the FCI-algorithm was also implemented with the package causal-learn [98], with the KCI as the conditional independence test.

3.4.4 GES

The GES algorithm estimates the MEC of the DAG which explains the causal relationships in the data by considering a greedy algorithm that maximizes a score function to render the best possible graph in the space of essential graphs [92]. Three phases are applied to orient the edges, namely the forward phase, the backward phase [75] and the turning phase [106].

This method was applied to the transcriptomics and methylation data with the R package pcalg [92]. The chosen scoring criterion was a $\mathbf{l}_0$-penalized Gaussian maximum likelihood estimator. The score was, then, calculated according to the operation $\lambda \times k$. The argument λ corresponds to the BIC criterion, being equal to $\frac{\log(n)}{2}$, with n number of samples. The argument k represents number of parameters of the DAG model [92].

Considering that the pcalg implementation only enables the distribution of the data to be linear-Gaussian (derived from the application of the BIC criterion as a scoring function), this algorithm was also implemented in the package causal-learn [98] to take into account a score that considers a non-linear and non-Gaussian distribution of the data. The scoring function that was chosen was the Cross-Validation likelihood [79], which can be defined as follows:

$$S_{CV}(X_i, \mathrm{PA}_i^{\mathcal{G}_h}) = \frac{1}{Q} \sum_{q=1}^{Q} \ell \left(\widehat{F}_i^{(q)} \mid D_{0,i}^{(q)} \right), \tag{3.3}$$

with $S_{CV}(X_i, \mathrm{PA}_i^{\mathcal{G}_h})$ being the Cross Validation log-likelihood, where X_i is the target variable and $\mathrm{PA}_i^{\mathcal{G}_h}$ the predictor, according to RKHS regression $\mathbf{k}_x = \tilde{F}(z) + \tilde{U}$. $\widehat{F}_i^{(q)}$ is the regression function resultant from the training data $D_{1,i}^{(q)}$. $\ell \left(\widehat{F}_i^{(q)} \mid D_{0,i}^{(q)} \right)$ designates the log-likelihood calculated in the qth test with the learned regression function $\widehat{F}_i^{(q)}$ [79]. For this procedure, a k-fold of 5 was utilized to perform the cross-validation. Taking into account the high dimensionality of the data and the computational costs of utilizing the RKHS, this approach is expected to have poor computational performance.

3.4.5 LiNGAM

The LiNGAM approach that was resorted to retrieve the causal information from the gene expression and methylation data from the different subtypes of glioma was the Direct-LiNGAM algorithm [110]. This approach estimates the causal ordering of the variables by performing, sequentially, regression analysis and evaluating in each step the independence between the explanatory variables and the residuals [114].

The algorithm was implemented with the LiNGAM python package [133]. To perform a linear regression between the variables, the implementation in the package resorts to the Least-Angle Regression (LARS) alongside to the application of a lasso model whose optimal regulation parameter is discovered by an information criterion like Alkaike Informatioon Criterion (AIC), which is the one that is utilized. One issue with this practice is that it only works when the number of variables is superior to the number of samples, which sometimes is not verified in the data in hand. To account for that eventuality, the source code was adapted to consider the context where there are more variables than samples with the application of the standard LARS procedure. An issue of this approach is that it performs an iterative refitting of the residuals based on coordinate descent instead of generating the optimal solution like the model described previously [134].

The evaluation of the independence between the explanatory variables and the residuals, was performed with the pairwise likelihood estimation [116], in detriment to the kernel-based method proposed in the Direct-LiNGAM inception [110].

3.4.6 PC-LiNGAM

The concept of LiNGAM was also applied to infer the causal directions which conditional independence algorithms, such as the PC-algorithm, could not identify [135]. Specifically, the ICA-based version of LiNGAM was employed, as scalling issues usually present in a high-dimensional setting do not apply in the bivariate context. Additionaly, this version was chosen as it was the one tested in the original article that proposed this method [135].

One positive aspect of this hybrid implementation is its effectiveness even when the error variables assume a Gaussian distribution. Considering the non-parametric aspect of the conditional independence algorithms, both Gaussian and non-Gaussian features can be utilized. Hence, the LiNGAM algorithm can be applied only when the variables have non-Gaussian error terms [114].

To discern the orientation of the edge, in the PC-LiNGAM framework, the following approach was adopted: regarding the connection matrix $\mathcal{B}$ containing the connections strengths b_{ij}, if $b_{ij} \neq 0$ and $b_{ji} = 0$, then the orientation of the edge would be $i \rightarrow j$; if $b_{ji} \neq 0$ and $b_{ij} = 0$, then the edge would be oriented as $j \rightarrow i$; otherwise the edge would remain bi-directed. The pcalg implementation of the LiNGAM algorithm was utilized to perform this analysis to each bi-directed edge from the application of the PC-stable algorithm.

3.4.7 ANM

The ANM approach has the limitation of only being applicable in a bivariate context or to a small graph, which does not fit the requirements of the glioma data. Thus, the PC algorithm was, beforehand, utilized to retrieve the skeleton of causal interaction. Afterwards, the ANM was applied to the edges where a causal direction was not clear in the MEC, being that in the form of a bi-directed edge, considering the PC-stable orientation rules.

This procedure was executed with the causal-learn [98] implementation of the PC algorithm in Python, with the KCI conditional independence test. This package also used the Fisher's Z test as a conditional independence assessment to account for the possibility of the KCI implementation not reaching a solution in a reasonable time. The ANM was performed to all the edges that could not be oriented from the method described previously.

To assess the directionality of the edge, a p-value was computed for the forward model ($A \rightarrow B$) and for the backward model ($A \leftarrow B$). Considering a cutoff of 0.05, if the p-value in the forward model was superior to 0.05, and in the backward model was inferior to the same threshold, the edge would be orientated according to the forward model. In the opposite scenario, where the p-value of the forward setting was inferior to 0.05 and the one in the backward model was superior, the edge would be orientated according to the backward model ($A \leftarrow B$). When both p-values were inferior or superior to the threshold, an orientation of the edge could not be verified, therefore remaining bi-directed.

3.5 Evaluation and Validation of the results

3.5.1 Bootstrap

Considering that there is no prior knowledge of the networks being inferred, it is hard to estimate the accuracy of the graphs or how well the causal models fit the data. Hence, a non-parametric bootstrap strategy was applied to assess the stability and robustness of the outputted graphs [104, 136, 137].

The non-parametric bootstrap approach lies in resampling several samples to provoke a 'disturbance' in the data. With that, if the graph remains similar to the original graph after the disruption, it can be stated that the algorithm that outputs that graph produces stable results [136]. Half of the samples in each dataset were randomly chosen, with replacement, in each resampling to perform such analysis. This procedure was replicated 100 times for each causal algorithm [104].

Considering the non-availability of a ground truth graph, the original graph for each causal algorithm was considered the 'true' graph against which the resampled graphs were compared. To estimate the similarity between the graphs, measures that calculate the structural difference between graphs were employed, such as the Structural Hamming Distance (SHD) [138]. Performance measures like precision and recall, the F1-Score, and False Discovery Rate (FDR) were also implemented [73, 104, 139].

Regarding the SHD [138], this measure evaluates the differences between two graphs regarding their edges. Considering two graphs, G and H, and the symmetric difference between the set of edges $E(G)$, $E(U)$) as $E(G, U)$, this metric estimates the number of modifications to transform G into H as follows:

$$\text{SHD}(G, H) = \sum_{(X,Y),X<Y}^{v^2} \begin{cases} 1 & \text{if } (X, Y) \in E(G, U), \\ 1 & \text{if } (Y, X) \in E(U, G), \\ 0 & \text{Otherwise,} \end{cases} \tag{3.4}$$

where v^2 represents the totality of variables from G and H. The modifications that are considered are the addition, removal and reversion of an edge [73]. Altogether, the smaller the overall value of this metric, the more stable the graph in consideration is.

Concerning the performance measurements, the binary classification problem metrics, such as the True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN) can be applied, in the context of this study, to measure the differences between the original graph and the graphs from the bootstrap samples. A TP occurs when a particular edge is present in both the original graph and the bootstrap replicate. Furthermore, an edge is considered a TN if absent in both graphs. Moreover, a FP happens when an edge in the bootstrap sample does not verify in the original graph, and the FN arises if an edge in the original graph is absent in the bootstrap sample [73, 139]. These performance metrics are summarized in 3.1.

Table 3.1: Summary of measures utilized for the performance measure of the nonparametric bootstrap of the original graph. Adapted from Hossin and Sulaiman [140]

Metrics	Formula	Evaluation Focus
Accuracy (Acc)	$\frac{TP+TN}{TP+FP+TN+FN}$	The accuracy metric measures the ratio of correct predicted edges over the total number of edges evaluated. The bigger the value, the more accurate the graph is.
Precision (P)	$\frac{TP}{TP+FP}$	This metric evaluates the edges that are correctly predicted in the pool of both TP and TN. The bigger the value, the more precise the prediction is.
Recall (R)	$\frac{TP}{TP+FN}$	Recall is used to measure the fraction of positive edges that are correctly classified. The bigger the value, the better the classification.
F1-Score (F1)	$\frac{2 \cdot P \cdot R}{P+R}$	This metric denotes the harmonic mean between recall and precision values. A bigger value of F1-Score indicates a more stable graph.
False Discovery Rate (FDR)	$\frac{FP}{FP+TP}$	This measure assesses the rate of false positives, considering the pool of positive classifiers in the bootstrap graphs. The desired outcome is that this rate is low.

For each metric, the mean and standard deviation were calculated considering the performance results from each bootstrap replicate. The algorithm comprising the best overall results was the one that was chosen to represent the causal interaction between the different glioma features. Considering the best overall graph, the edges that had a frequency superior to 60% when comparing the original graph with the bootstrap samples were the ones that were retained. Afterwards, the best graph per subtype of glioma was plotted with the Cytoscape software [128].

3.5.2 Biological assessment of the networks

To evaluate the biological meaning of the causal networks, enrichment analysis based on GO and KEGG pathways [141] were performed.

GO is a valuable resource for analyzing a set of gene products, considering that this tool provides an annotated description of the molecular functions, cellular localization and biological processes of genes across multiple species [142]. With that, Gene Ontology Enrichment Analysis (GOEA) was applied to assess gene ontology terms that would be

enriched in the different gene sets regarding the distinct glioma subtypes. To perform this analysis, the GOATOOLS python library [143] was used, given its better performance in comparison to other well-established tools such as DAVID [144] and GOstats [145].

The goal of GOEA is to verify if certain gene sets are enriched for a functional class of GO. The enrichment of a GO term in a set of genes is measured according to the probability of a group of genes being associated to a particular GO term, considering the background frequency (frequency of genes in the entire genome associated with that particular GO term). This probability results in a p-value, where the lower the P-value, the least probable it is that a particular GO term is associated with a set of genes, thus indicating a bigger enrichment [142]. The p-value was computed based on Fisher's exact test. The α that was considered was 0.05.

Regarding the plethora of p-values calculated in this analysis, each p-value was corrected to avoid the computation of false positives. Despite the variety of p-value correction methods this package displays, the one chosen was the FDR- Benjamini-Hochberg method. The value that was set for the FDR was 0.05, meaning that one should consider that up to 5% of the statistically significant results from the p-value calculation can be, in fact, false positives [143, 146].

The complete set of genes per subtype was provided as input to the GOEA. Given that the result displayed a subset of genes per statistically significant term, those genes were then verified in the networks if they could highlight a sub-network accountable for a given biological function.

The CpG sites that had a differential expression per subtype according to the iDINGO result were tested for the enrichment of Gene Ontology terms [142], and KEGG pathways [141] with the R package missMethyl [147]. The annotation from Illumina HumanMethylation450 was utilized to map each selected CpG site to the corresponding genes. The enrichment of the Gene Ontology or KEGG pathways genes was evaluated according to Wallenius' noncentral hypergeometric distribution. Only the categories ranked in the top 20 lowest p-values associated with the Gene Ontology and KEGG terms were selected.

In order to assess the validity of the gene interactions retrieved from the causal discovery results, databases comprising curated biological interactions, such as the UCSC Genome Browser [148] and STRING [149] were visualized. Lastly, the DrugBank database [150] was consulted to assess drugs that could potentially target selected targets from our analysis.

4

Results and discussion

In this chapter, the findings from both undirected and directed methods are detailed, alongside their biological validation. The discussion focuses on the interpretation of these outcomes, exploring their framing in the context of glioma and cancer research and highlighting their potential implications.

4.1 Hub-genes discovery with undirected methods and iDINGO

The high dimensionality of the data extracted from TCGA was an obstacle for the application of network analysis to the gene expression and methylation data from the glioma subtypes. Hereby, the graphical lasso's ability to induce sparsity in the network according to a particular ρ was sought after to only select the variables that were not isolated, for further analysis.

Initially, the methylation dataset contained 340427 features, while the RNA dataset had 20501 variables. With the graphical lasso, the number of variables retrieved for each subtype was significantly lower. Regarding the gene expression data, 107 variables were retrieved for the astrocytoma samples, 32 features for GBM and 104 were selected for oligodendroglioma. Concerning the methylomics data, 60 features were picked for astrocytoma, 63 for GBM, and for oligodendroglioma, 67 features were selected.

A summary of the number of initial variables and the ones retained after applying graphical lasso with distinct ρ parameters for each sample is illustrated in table 4.1. Moreover, the complete list of all the variables that were selected per subtype is detailed in A.1.

Later on, the distinctive features identified from graphical lasso were further analyzed using iDINGO for each subtype. The iDINGO method was employed not only to assess the differential expression of methylomics and RNA-seq data within LGG samples and between GBM and LGG (DINGO framework), but also to inspect the causal influence of the CpG sites to the RNA-seq features, for each subtype. The results from this approach can be seen in figure 4.1. Edge colors denote partial correlations, where a darker red reflects a higher negative partial correlation, while a more intense degree of green designates

Table 4.1: Summary of the number of initial variables for each sample and the amount of selected features after the application of graphical lasso with the respective ρ parameter

Omic Type	Samples	Initial Variables	Selected Variables	ρ
RNA-seq	Astrocytoma	20501	107	0.955
RNA-seq	GBM	20501	32	0.955
RNA-seq	Oligodendroglioma	20501	104	0.955
Methylomics	Astrocytoma	340427	60	0.86
Methylomics	GBM	340427	63	0.85
Methylomics	Oligodendroglioma	340427	67	0.90

a higher positive partial correlation. Regarding only LGG samples, the result of the implementation of iDINGO framework to astrocytoma samples, against oligodendroglioma samples is present on (4.1a). Diversely, the same approach was applied to assess the differential expression of oligodendroglioma samples against astrocytoma samples (4.1b). This analysis was further used to verify the differential expression of features respective to GBM samples, in comparison with LGG samples (4.1c).

In astrocytoma iDINGO output (4.1a, B.1), it is noticeable that the most connected genes were *COL4A1*, *CYBA* and *HBA2*, *COL4A2* and *HBB*. In addition, the unique CpG sites for astrocytoma, cg08843517 (*CYBA*) and cg04771838 (*COL4A2*) were the methylation sites with the most interactions. Regarding oligodendroglioma (4.1b, B.2), *STX1A*, and *TMEM144* were the ones that contained more edges, with the former being unique to oligodendroglioma, and the latter also showing differential expression in astrocytoma. Both methylomic sites mapped to *LDB3*, cg03303359 and cg01657744, uniquely expressed in the oligodendroglioma samples, were the CpG sites had the highest number of connections. In GBM, the most connected genes were *ANXA2P2*, *ANXA2*, *S100A9*, *S100A8* and *COL4A1*, while the CpG sites cg02072495 and cg14397690, uniquely expressed in GBM, were the ones that share more connections. The nodes previously described as features with a higher degree of connectivity for each glioma subtype can be regarded as hubs for each sample type. GBM (4.1c, B.2) denotes a higher incidence of more intense differential connections than the astrocytoma (4.1a) and oligodendroglioma (4.1b), as verified in the color gradation of the edges. It is also interesting to note that in oligodendroglioma (4.1b) and GBM (4.1c), the strongest differential connections are associated with the hubs, a pattern not observed in astrocytoma (4.1a).

Regarding astrocytoma, the edge with a higher positive differential connection is between *CKAP2L* and *HBB*, while the one with a higher negative differential relation is between *CKAP2L* and *HBA2*. In relation to oligodendroglioma, the edge with the highest positive differential value relates *STX1A* and *TMEM144*, both previously perceived as hub genes for this subtype, while the higher negative differential edge for this subtype is between *FCGR1B* and *STX1A*. With respect to this subtype, it is relevant to point out that the highest positive differential interaction between a methylomic site and a gene is between cg02571816 and *TMEM144*. In contrast, the highest negative differential edge of

the same type relates cg02571816 (*PPP1R14A*) to *STX1A*. Concerning GBM differential connections, it is alluring to verify that the differential edges with higher values relate CpG sites to the previously referred hub genes associated with this subtype, namely the interaction between cg00765737 (*COL4A2*) and *ANXA2*, the one where cg11342452 (*NKX6-2*) influences *ANXA2P2*, and the connection between cg13572782 (*MBP*) and *ANXA2P2*. On the other side, the influence of cg00765737 (*MBP*) in *ANXA2P2* is the edge with the higher negative differential score.

4.2 Glioma subtypes modelling with causal discovery algorithms

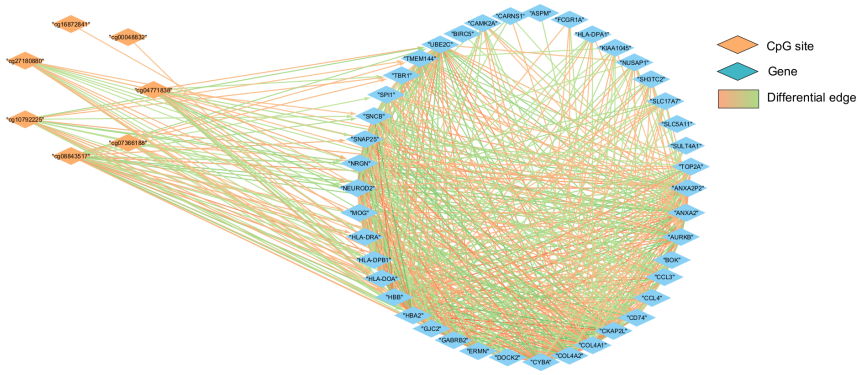
4.2.1 Optimal graphs per omic type

The previous methodology allowed us to infer the direct connections between different platforms, namely from methylomic data to gene expression data. Nevertheless, there was still the need to assess the causal structure within each platform. Considering that, different causal discovery approaches were employed to retrieve the causal structure for each subtype, concerning the gene expression and methylomic features that were previously collected. The causal discovery algorithms that were applied were the PC algorithm, the FCI algorithm, the GES algorithm, the Direct-LiNGAM, the ANM algorithm applied after the PC algorithm retrieved the skeleton and the PC-LiNGAM.

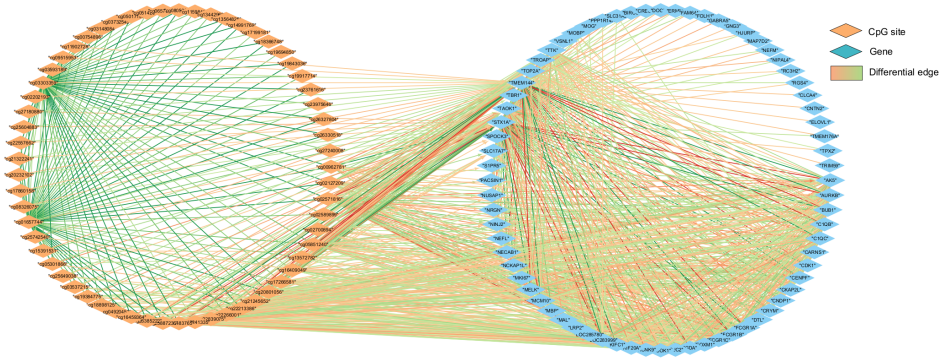
Methods based on RKHS, such as the KCI conditional independence test applied to the PC and FCI algorithms, as well as the Cross Validation log-likelihood, a scoring method employed to the GES algorithm, could not compute the first replicate in the bootstrap implementation within a week for any dataset except GBM RNA. Therefore, these approaches were excluded from further analysis. Alternately, the Fisher's Z test was applied for the evaluation of the conditional independence in PC and FCI algorithms (also affecting ANM and PC-LiNGAM), and GES results were based on the BIC criterion.

To evaluate the robustness and stability of the different causal algorithms, several metrics were applied to the bootstrap implementation. This involved comparing the initial graphs retrieved by the causal algorithm with those generated from each bootstrap replicate. Keeping in mind the similarity of the outcomes of these metrics across the different data types, for matters of visualization, the performance of GBM RNA is depicted in Figure 4.2, while the performance metrics for the causal algorithms applied to astro RNA, astro CpG, GBM CpG, oligo RNA, and oligo CpG are displayed in the annex, from Figure I.1 to I.5.

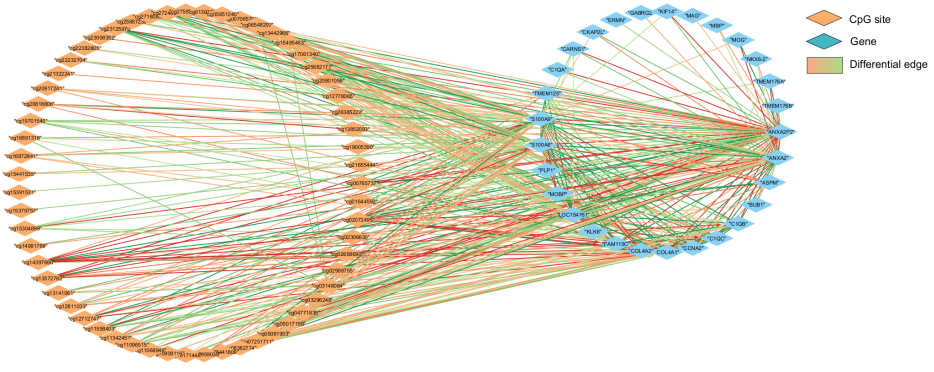
Regarding the stability of the causal algorithms applied to the astrocytoma RNA variables (I.1), the FCI algorithm proved to be more robust. It had the lowest SHD (I.1a, C.1) and FDR (I.1c, C.1) values, while outperforming the other algorithms in F1-score (I.1b, C.1), precision (I.1d, C.1) and recall (I.1e, C.1). For astrocytoma CpG data (I.2,C.2), it is explicit that the FCI was overall the most stable algorithm. It is relevant to point out that



(a) Astrocytoma vs Oligodendroglioma differential expression



(b) Oligodendroglioma vs Astrocytoma differential expression



(c) GBM vs LGG differential expression

Figure 4.1: Integration of methylation and RNA-seq data for different glioma subtypes using iDINGO. Red edges denote negative partial correlations, while green edges represent positive differential connections. Figure 4.1a shows differential expression in astrocytoma, and Figure 4.1b displays differential expression in oligodendroglioma. Both networks represent differential expression analysis in LGG samples. Figure 4.1c contrasts the differential expression between GBM and LGG samples.

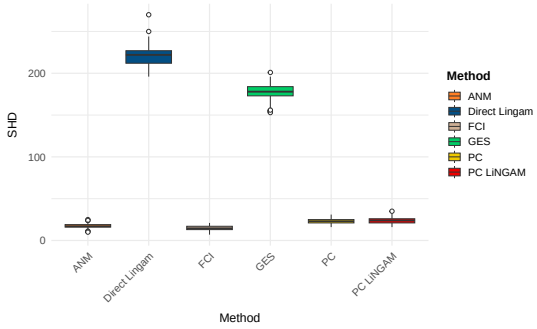
ANM, PC and PC-LiNGAM outperformed FCI in SHD. Nonetheless, FCI still excelled in the remaining evaluation metrics (I.2b,I.2c,I.2d, I.2e, C.2).

The performance evaluation of the causal algorithms applied to oligodendroglioma RNA data exhibited that the FCI algorithm was, without contest, the most robust algorithm (I.4,C.3). Despite this algorithm displaying stable SHD (I.4a, C.3) and FDR (I.4c, C.3) results, the mean of the F-score (I.4b), and precision (I.4d, C.3)) were only slightly above 0.5, while the recall mean was below 0.5 (I.4e) (C.3). Similarly, for oligodendroglioma CpG data, the FCI-algorithm was the overall most robust algorithm (I.5, C.4). However, the ANM method outperformed FCI in SHD (I.5a, C.4). Despite that, the FCI-algorithm excelled in the remaining metrics (I.5b, I.5c, I.5d, I.5e, C.4).

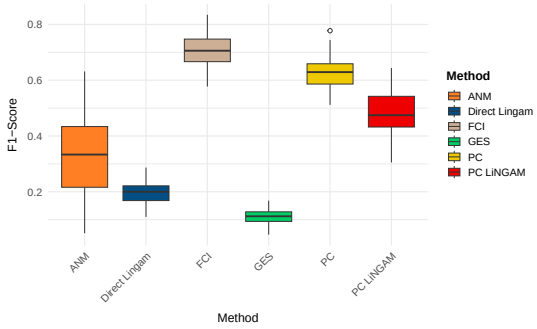
Considering the assessment of the stability of the causal algorithms that were applied to the GBM RNA data, it is clear that the algorithm that had the best performance across all metrics (4.2, C.5) was also the FCI algorithm, denoting even better results across all metrics than those concerning astro RNA, and oligo RNA C.5. Likewise, for GBM CpG data, it is perceptible that the FCI yields the best overall performance across all metrics (I.3, C.6).

According to the evaluation metrics, the FCI algorithm was the best-performing algorithm across all datasets. Therefore, the graphs outputted from this algorithm were the ones that were selected to model the different data related to glioma. That said, only the edges that, in the bootstrap replicates, had a frequency superior to 60% and were present in the initial full-sample graphs were considered. In the final graphs (4.3,4.4, 4.5), it was evident that a substantial part of the edges was bidirected. Under the FCI assumptions, such might indicate the presence of unknown variables that do not allow to discern the directionality of the edges.

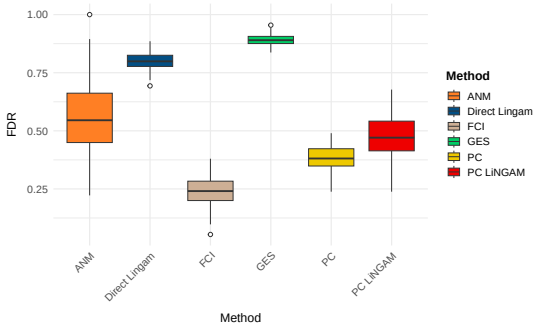
In the astro RNA graph (4.3a), some distinctive subnetworks can be identified. The largest subnetworks comprise 5 nodes. One of them comprehends the unique astrocytoma interactions between *SNAP25*, *SYN1*, *SNCB*, *NRGN*, and the direct influence of *NRGN* on *VSNL1*. Another particular subnetwork of this subtype shows the connections between *HLA-DOA*, *HLA-DPB1*, *HLA-DPA1*, *HLA-DRA* and *CD74*. A third subnetwork highlights interactions between *WDFY4*, *DOCK2* and *SYK*, the connection between *NCKAP1L* and *HAVCR2*, besides the causal influence of *NCKAP1L* in *DOCK2*. Another distinct subnetwork includes the influence of *CNTN2* on *ERMN* and *ERMN* on *ENPP2*. This subnetwork also shares the causal direction of *MOBP* to *ERMN* with the GBM network and with oligo RNA the causal link between *MOBP* and *MBP*. It is relevant to mention that in the astro RNA context, while the edge between *MBP* and *MOBP* is bidirected, in oligo RNA, *MOBP* has a direct effect on *MBP*. Such might suggest that there is an unknown variable affecting this interaction in astro RNA that is non-existent in oligo RNA. Further, there are some four-node unique astro RNA subnetworks to take notice of, including the connections between *LOC283999*, *GPR62*, *S1PR5*, and the causal influence of *GPR62* on *SLC5A11*. Another involves the links between *NUSAP1*, and *TPX2*, *TOP2A*, and *CENPF*, besides the causal interaction between *TPX2* and *TOP2A*. The relatedness between the genes



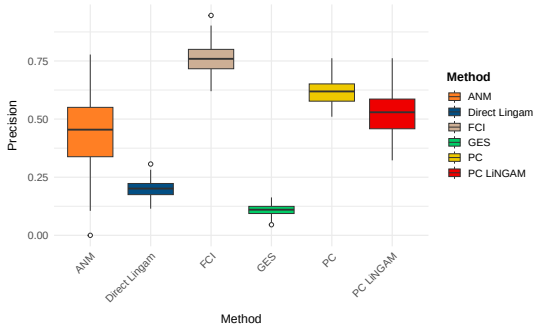
(a) SHD values from the causal discovery algorithms applied to GBM RNA



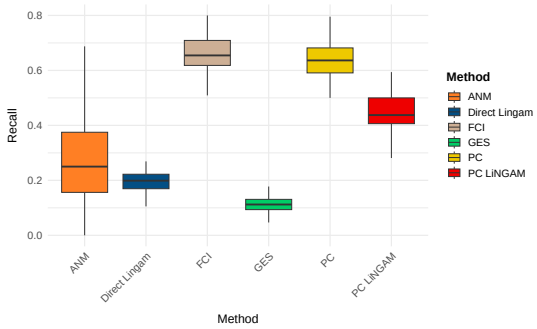
(b) F-score values from the causal discovery algorithms applied to GBM RNA



(c) FDR values from the causal discovery algorithms applied to GBM RNA



(d) Precision values from the causal discovery algorithms applied to GBM RNA



(e) Recall values from the causal discovery algorithms applied to GBM RNA

Figure 4.2: Bar plots describing the evaluation metrics, namely SHD, F-score, FDR, precision and recall, used to assess the robustness of the different causal discovery algorithms applied to the GBM RNA dataset. Each bar plot denotes the distribution of the results for each evaluation metric, comparing each bootstrap replicate to the original graph containing the total number of samples.

CACNG3, *MAL2*, *MAP7D2*, and *KCNC2*, and the subnetwork comprising the interactions between *FERMT3*, *SPI1*, *CYBA*, and *TYROBP*, are also particular distinct astrocytoma subnetworks. In addition, the subnetwork comprising *C1Q* genes (*C1QA*, *C1QB*, *C1QC*) is relevant, with *LAPTM5* uniquely influencing *C1QC*. In the three-node subnetworks, the subnetwork comprising the direct influence of *SULT4A1* in *GABRA1*, which, in turn, shared a bidirected edge with *GABRB2*, had in common the bidirected edge with the oligo RNA graph. The remaining three-node graphs were unique for this subtype. It is also intriguing to verify the link between *COL4A1* and *COL4A2*, and the bidirected edge between *ANXA2* and *ANXA2P2*, shared with GBM.

In astro CpG causal graph (4.3b), the largest subnetwork comprises the unique astrocytoma interactions between probes mapped to *MBP*, namely cg21322214, cg02589899, cg13442966, cg27240008 and cg15391531. The oligodendroglioma CpG graph shares with astrocytoma the causal link between cg27240008 (*MBP*) and cg15391531 *MBP*, which might suggest that an unknown confounder in the astrocytoma context might be responsible for the association between these sites. From the CpG sites that were differentially expressed in astrocytoma according to iDINGO (4.1a), only two of them have a causal connection. the CpG site cg16872841 (*COL4A2*) has a causal influence on cg04771838 (*COL4A2*) through a middle node not differentially expressed in astrocytoma samples, according to the iDINGO output, namely cg00765737 (*COL4A2*). It is relevant to note that this causal interaction is also depicted in the GBM CpG context.

The oligo RNA causal graph (4.4a) reveals some unique subnetworks. The subnetwork with the most edges comprises the unique interactions between *CYB5R2* and *CLCA4*, the causal influence of *CLC4A* in *TRIM59*, the causal effect of *LOC285780* in *CLC4A* and the interaction between *LOC285780* and *EMX1*.A subgraph involving the interactions between *SAMD8*, *UBXN7*, *RC3H2*, and *TAOK1* is noteworthy, besides all edges being bidirected. Further, another distinct subnetwork comprises the interactions between *TBR1* and *LOC286002*, the connection between the previously mentioned node and *SLC6A7*, besides the causal interaction between *SLC6A7* and *NEUROD2*. It is worth noting that, according to FCI, the gene *ERMN* has a causal influence on *TMEM144* and *MOBP*, solely in oligodendroglioma, sharing the presence of the causal influence of *MOBP* in *MBP* with astrocytoma. One curious aspect of the oligodendroglioma causal network is that it does not share many edges with GBM RNA network. The only connection in common is between *TMEM176A* and *TMEM176B*.

The oligodendroglioma CpG causal graph (4.4b), denotes a prevalence of unique oligodendroglioma connections, besides sharing some edges with astrocytoma. In concordance with the oligodendroglioma RNA graph (4.4a), there are not many edges that are shared with GBM, apart from a subnetwork comprising the CpG sites cg27180880 (*MBP*), cg20801056 (*MBP*), and cg25887236 (*MBP*); and the direct link between cg03148084 (*CNDP1*) and cg05851240 (*CNDP1*). One of the unique subnetworks that are noteworthy to mention comprises the CpG site cg23761616 (*TMEM144*) as a source of causal influence to the methylomic sites cg00962781 (*TMEM144*), cg19843036 (*TMEM144*), and

cg18366748 (*TMEM144*). The methylomic site cg01657744 (*LDB3*) interacts with the CpG sites cg00059015 (*HOOK1*), and cg01657744 (*LDB3*), the latter connecting with cg03303359 (*LDB3*). Additionally, cg27240008 (*MBP*) shares with astrocytoma the effect on cg15391531 (*MBP*) while also directly influencing the variable cg00754896 (*LOC283999*), uniquely in oligodendroglioma.

By analyzing the causal graph that represents the interactions between GBM RNA selected variables 4.5a, one might notice the presence of only four identifiable subgraphs. The largest incorporates the unique GBM RNA connections between *FAM115* and *LOC154761*, the connection between *TMEM125* and *NKX6-2*, and the causal influence of *TMEM125* in *LOC154761*. Other unique subnetwork contains the interactions between *ASPM*, *BUB1*, and *CCNA2*. Other subgraph displays the unique causal interaction between *CARNS1* and *MOBP*, which in turn, has a causal effect on *ERMN*, a connection also verified in astrocytoma. The connections between the *C1Q* genes *(C1QA*, *C1QB*, *C1QC*) are also stated in astrocytoma, with the difference of absence of the *LAPTM5* gene that has a connection with *C1QC*, solely in astrocytoma.

Considering the causal interpretation of the connections between the GBM CpG features (4.5b), it is relevant to reference the subnetwork comprised of the methylomic sites cg11342452 (*NKX6-2*), cg02306630 (*NKX6-2*), cg06648277 (*NKX6-2*) and cg08382774 (*NKX6-2*), which is particular to the GBM subtype. Also, the interactions between cg20801056 (*COL4A1*), cg27180880 (*MBP*), and cg25887236 (*MBP*), which have different connections between each other in astrocytoma and oligodendroglioma, might indicate the presence of unknown confounders in GBM that are differentially influencing these variables in the other two subtypes.

4.2.2 Integration of causal discovery and iDINGO graphs

Based on the results obtained from iDINGO which revealed the causal interactions between methylomics data and the selected genes for each subtype, as well as the outcome of the causal discovery graphs, a final integrated graph was constructed for each subtype (4.6,4.8,4.7). As referenced in the legend of each graph, different edge colors represent distinctive relationships between subtypes. Green edges symbolize the unique edges to each subtype; the red edges denote the ones that also occur in oligodendroglioma; the blue edges represent the edges that also appear in astrocytoma; the purple edges highlight the ones that are shared with GBM; and the grey edges are the ones that are represented in all subtypes. Dashed lines denote the cases where CpG sites mapped to a specific gene (D.1,D.3,D.2), have a causal influence in the expression of that gene, as indicated by iDINGO.

4.3 Biological assessment of the networks

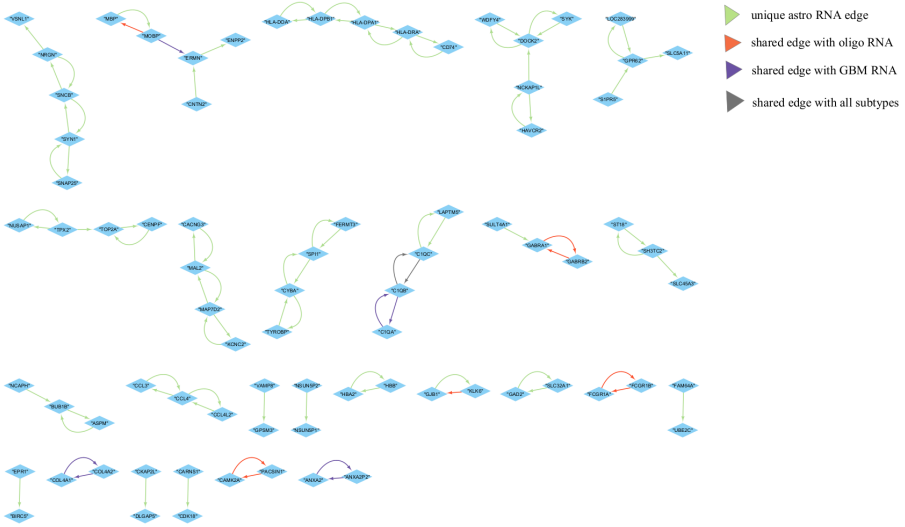
4.3.1 Enrichment analysis

After retrieving the networks that characterize the different glioma subtypes according to the different data types, the subsequent step was to investigate potential biological explanations for the network variations, which could help explain the dissimilarities between subtypes. To this end, a GOEA was performed on the methylomic sites that, according to the iDINGO result, were differentially expressed in each subtype. Regarding the astrocytoma CpG sites that were selected (4.9, E.1), it is evident that the methylomic sites mapped to the genes *BIRC5* and *KIFC1* were enriched for several mitotic events, such as the microtubule cytoskeleton organization, mitotic spindle organization, spindle assembly and the metaphase plate congression, besides also being associated with the protein-containing complex. Furthermore, the CpG sites mapped to *CYBA* and *MBP* are enriched for terms related to interleukin-6 production. The enrichment analysis performed on the GBM methylomic sites (4.11,E.2) highlights the association of collagen-related terms to the CpG sites mapped to *COL4A1* and *COL4A2*. The terms associated with myelin were related to the methylomic sites mapped to *ANXA2*, *MBP*, and *ERMN* genes. The CpG sites mapped to *TMEM176B*, *MBP*, and *NKX6-2* were also enriched for the negative regulation of both multicellular organismal processes and cell development. The same procedure was applied to the methylomic sites designated for oligodendroglioma (4.10,E.3), and it was found that the most enriched terms were related to the myelin and the process of myelination, associated to the CpG sites related to *MAG*, *MAL*, *MBP*, and *CNTN2* genes. In addition, the methylomic sites associated with *STX1A*, *SYN2*, and *SV2B* are connected to synaptic processes.

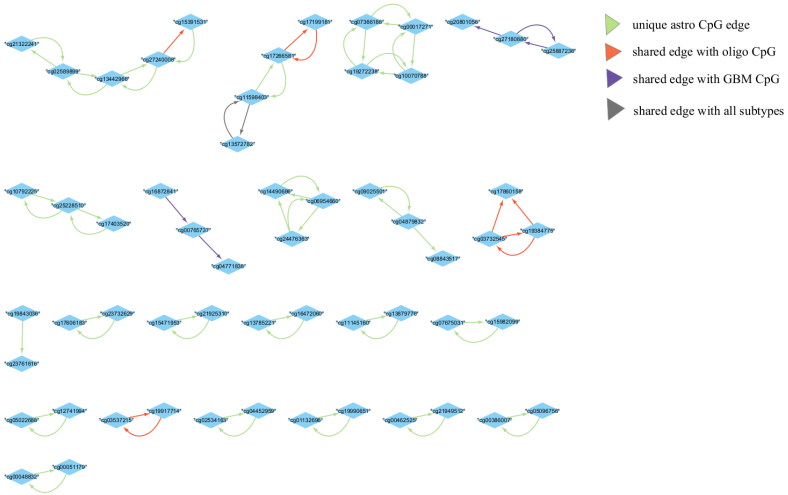
Considering the enrichment of pathways prevalent in KEGG pathways for astrocytoma-selected CpG sites (4.12,E.4), it is relevant to point out the association between the methylomic site mapped to *BIRC5* and pathways in cancer, apoptosis and colorectal cancer. The methylomic sites mapped to *COL4A2*, namely cg04771838 and cg16872841, were enriched for the terms pathways in cancer, small cell lung cancer, ECM-receptor interaction, and AGE-RAGE signalling pathway in diabetic complications. The methylomic site related to *CYBA* was enriched for terms like NOD-like receptor signalling pathway and neutrophil extracellular trap formation. The pathways enriched for GBM (4.14,E.5) are primarily associated with the CpG sites related to *COL4A1* and *COL4A2*, ranging from pathways in cancer, the Pl3K-Akt signalling pathway, and the relaxin signalling pathway, to small-cell lung cancer. Also, the methylomic sites mapped to *CNDP1* are enriched for some metabolic pathways, such as arginine proline and histidine metabolism. Regarding the oligodendroglioma methylomic sites pathway enrichment analysis (4.13,E.6), it is interesting to note the enrichment of the methylomic sites mapped to *MAG* and *CNTN2* to cell adhesion molecules. The methylomic sites related to *CNDP1* and *AK5* were enriched

for different metabolism types, specifically beta-alanine, arginine and proline, and nucleotide metabolism, respectively. Additionally, the CpG sites associated with *STX1A* were enriched for neurodegeneration pathways and synaptic-associated processes.

A gene ontology enrichment analysis, with the GOATOOLS package, was also applied to the selected genes for the different subtypes to assess if some subnetworks could be highlighted according to enriched biological processes. Concerning the terms that were enriched for astrocytoma RNA (4.15,E.7), it is clear the association between the *HLA* genes (*HLA-DOA*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*) and *CD74* with the processes related to antigen processing and assembly with MHC class II. Several genes were associated with the positive regulation of calcium-mediated signalling, namely *SYK*, *GPR62*, *CCL3*, *CCL4*, and *NEUROD2*. Moreover, *CCL3* and *CCL4* also participate in the response to toxic substances, sharing that role with the genes *NEFL*, *GJC2*, *CCL4*, *CDK1*, and *MBP*. The genes *AURKB*, *BIRC5*, and *KIF23* are also enriched for mitotic processes. Further, the enrichment in synapse pruning is enriched for the set of genes *C1QA*, *C1QB*, and *C1QC*. According to this enrichment analysis, the *SYK* gene also participates in the collagen-activated tyrosine kinase receptor signalling pathway alongside *COL4A1* and *COL4A2*. Regarding the enrichment analysis performed to GBM genes (4.17,E.8) it is interesting to note that the most enriched term associates the genes *C1QA*, *C1QB*, and *C1QC* to synapse pruning. These genes are also enriched for the complement activation, classical pathway term. Furthermore, the genes *S100A9* and *S100A8* are enriched for a plethora of biological processes, specifically astrocyte development, neutrophil aggregation, sequestering of zinc ions, and peptidyl-cysteine S-nitrosylation. The *TMEM176A* and *TMEM176B* genes were enriched for negative regulation of dendritic cell differentiation. Interestingly, besides *COL4A1* and *COL4A2*, also *S100A8*, *S100A9*, *C1QA*, *C1QB*, and *C1QC* were enriched for the collagen extracellular matrix component. The results from applying this approach to oligodendroglioma genes highlight the enrichment of the term myelin sheath regarding the genes *MAG*, *CNTN2*, *MBP*, *ERMN*, *MOBP* and *MAL*. Notably, the same genes enriched for the GABA-A receptor complex-related genes, such as *GABRB2*, *GABRA1*, and *GABRA5*, were also enriched for the chloride channel complex alongside *SLC17A7*.

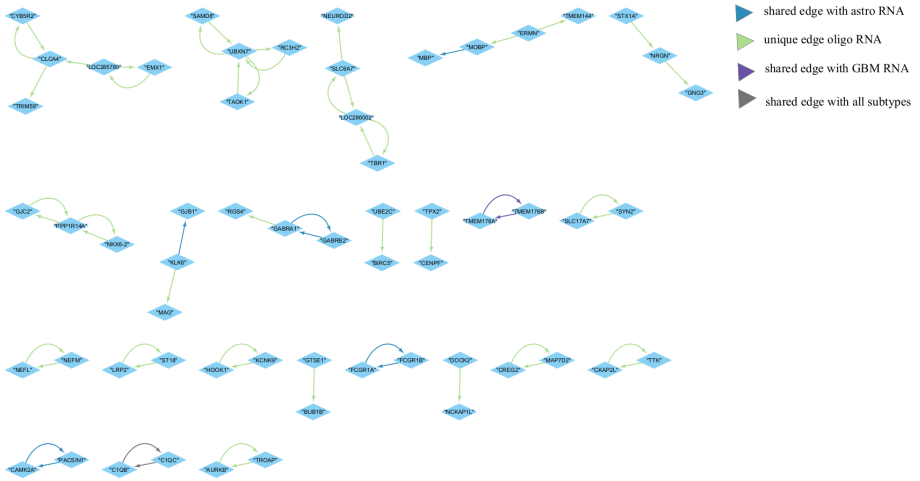


(a) Selected graph from FCI that depicts the causal structure of astro RNA

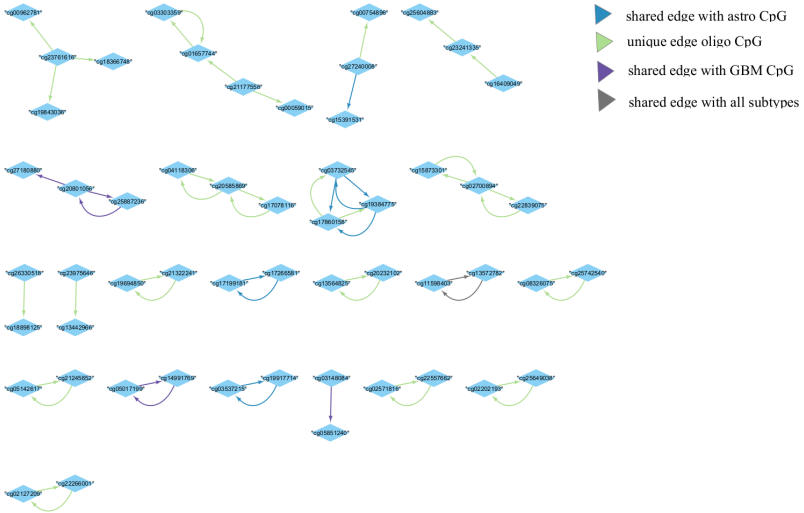


(b) Selected graph from FCI that depicts the causal structure of astro CpG

Figure 4.3: Selected graphs from FCI algorithm for astro RNA (4.3a) and astro CpG (4.3b), where each edge as a frequency superior to 60 % in the bootstrap replicates.

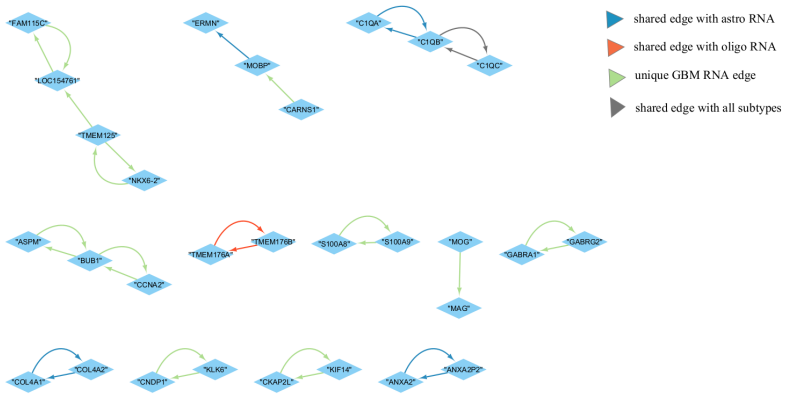


(a) Selected graph from FCI that depicts the causal structure of oligo RNA

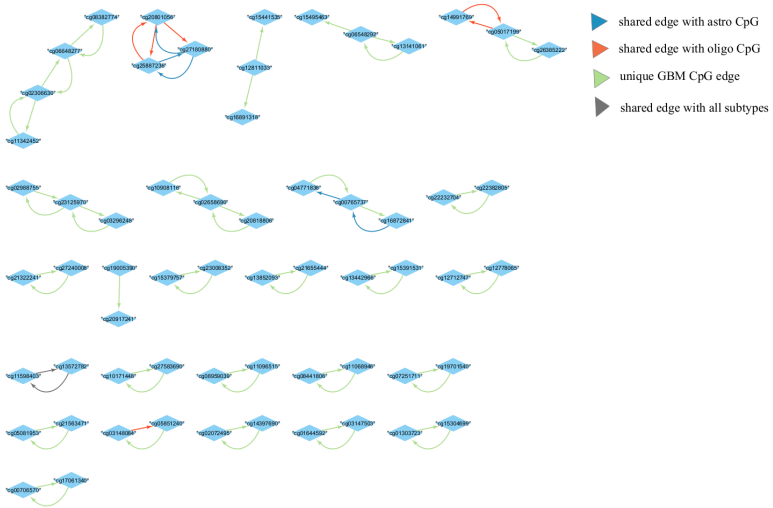


(b) Selected graph from FCI that depicts the causal structure of oligo CpG

Figure 4.4: Selected graphs from FCI algorithm for oligo RNA (4.4a) and oligo CpG (4.5b), where each edge as a frequency superior to 60 % in the bootstrap replicates.



(a) Selected graph from FCI that depicts the causal structure of GBM RNA



(b) Selected graph from FCI that depicts the causal structure of GBM CpG

Figure 4.5: Selected graphs from FCI algorithm for GBM RNA (4.5a) and GBM CpG (4.5b), where each edge as a frequency superior to 60 % in the bootstrap replicates.

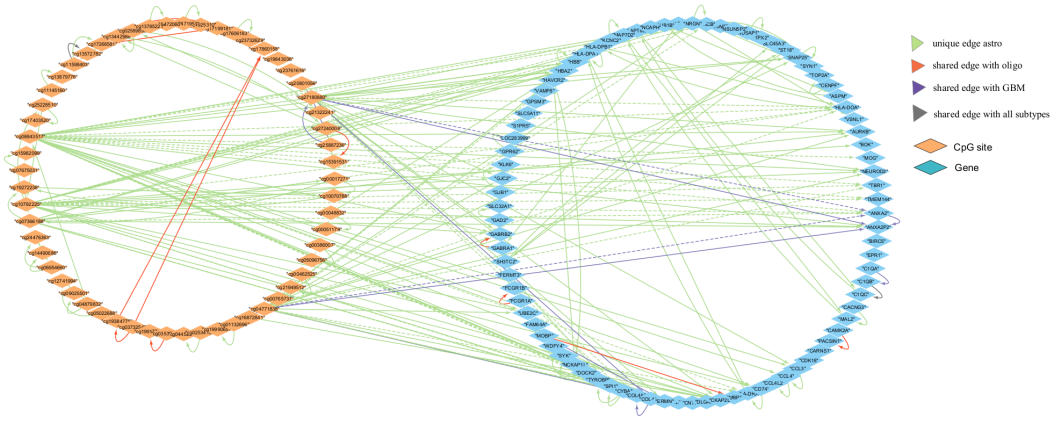


Figure 4.6: Causal graph characterizing the connections between methylomics and transcriptomics features selected for astrocytoma, retrieved from iDINGO and FCI-algorithm

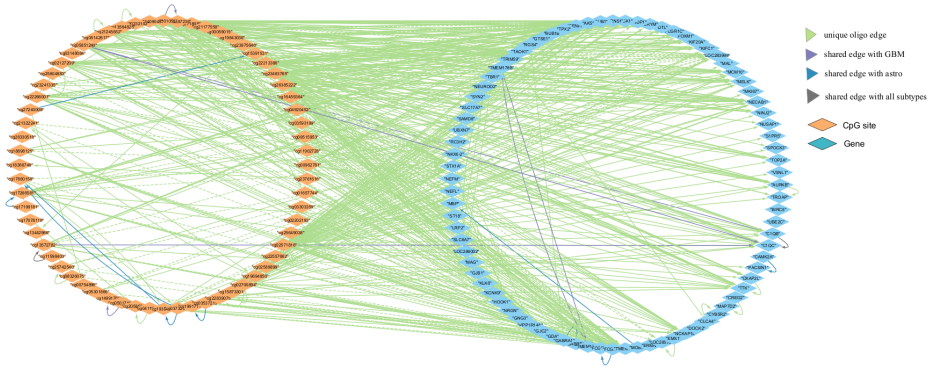


Figure 4.7: Causal graph characterizing the connections between methylomics and transcriptomics features selected for oligodendroglioma, retrieved from iDINGO and FCI-algorithm

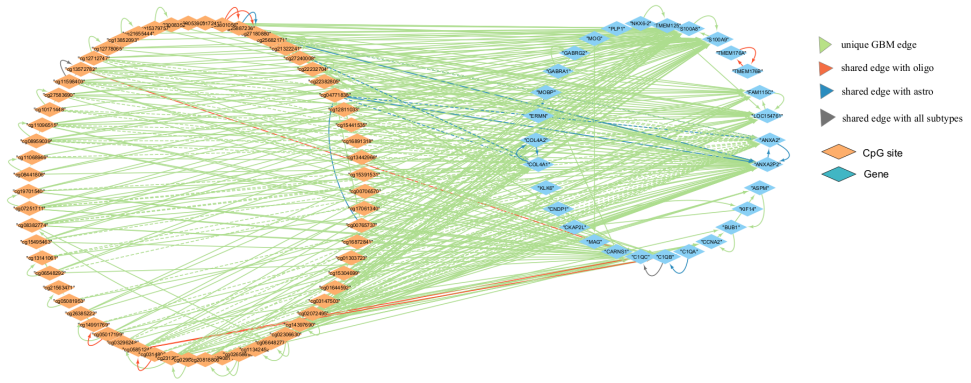


Figure 4.8: Causal graph characterizing the connections between methylomics and transcriptomics features selected for GBM, retrieved from iDINGO and FCI-algorithm

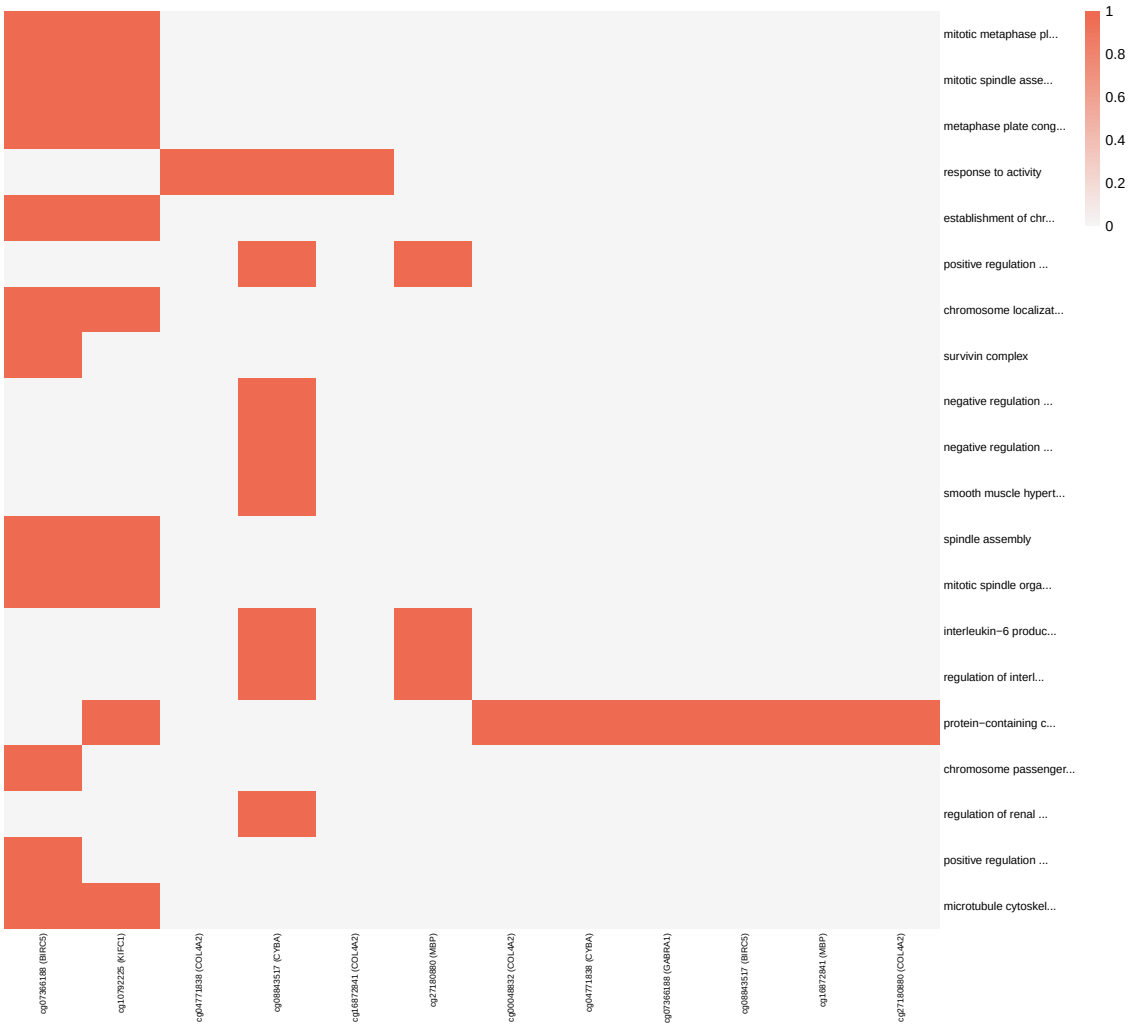


Figure 4.9: Astrocytoma CpG Gene Ontology enrichment analysis

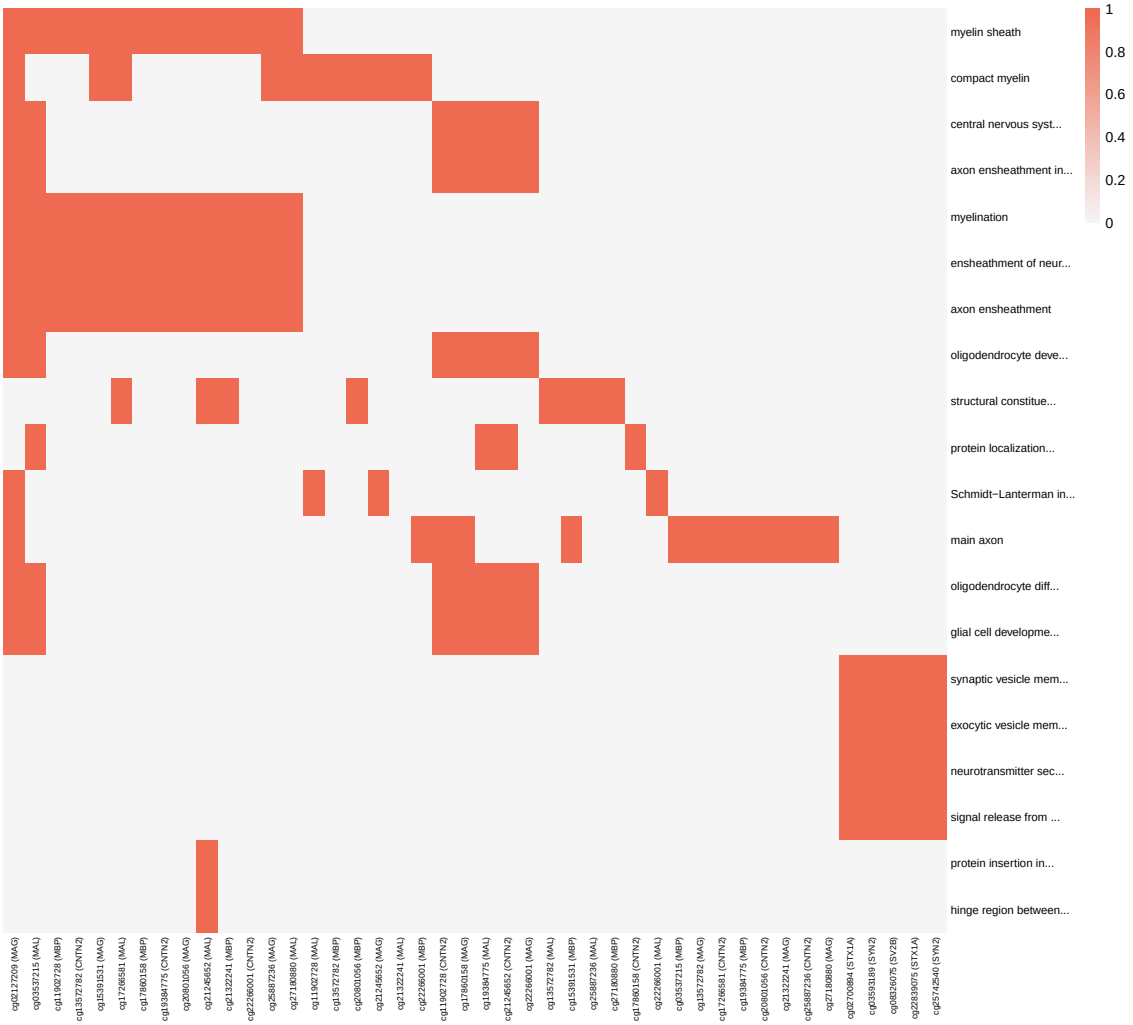


Figure 4.10: Oligodendroglioma CpG Gene Ontology enrichment analysis

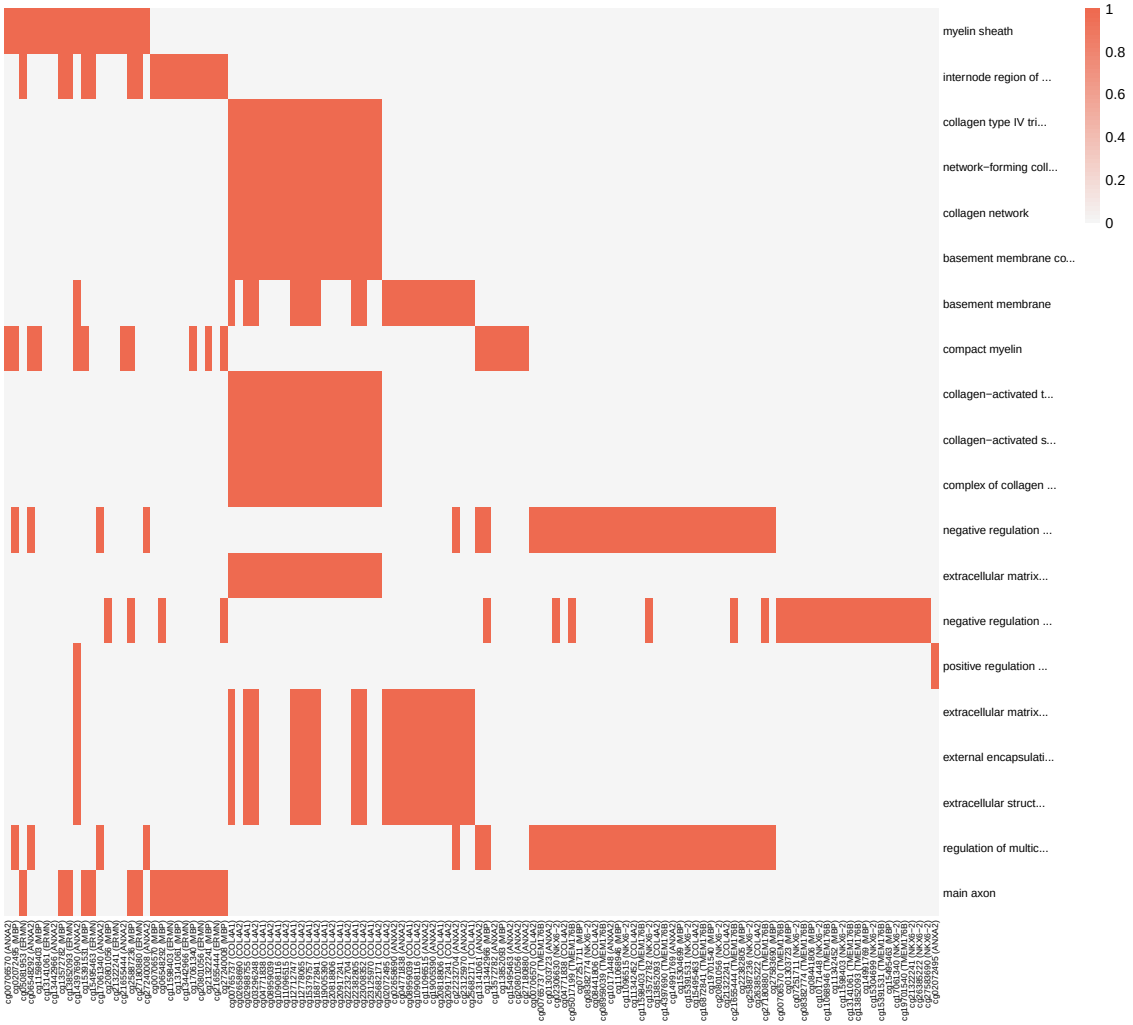


Figure 4.11: GBM CpG Gene Ontology enrichment analysis

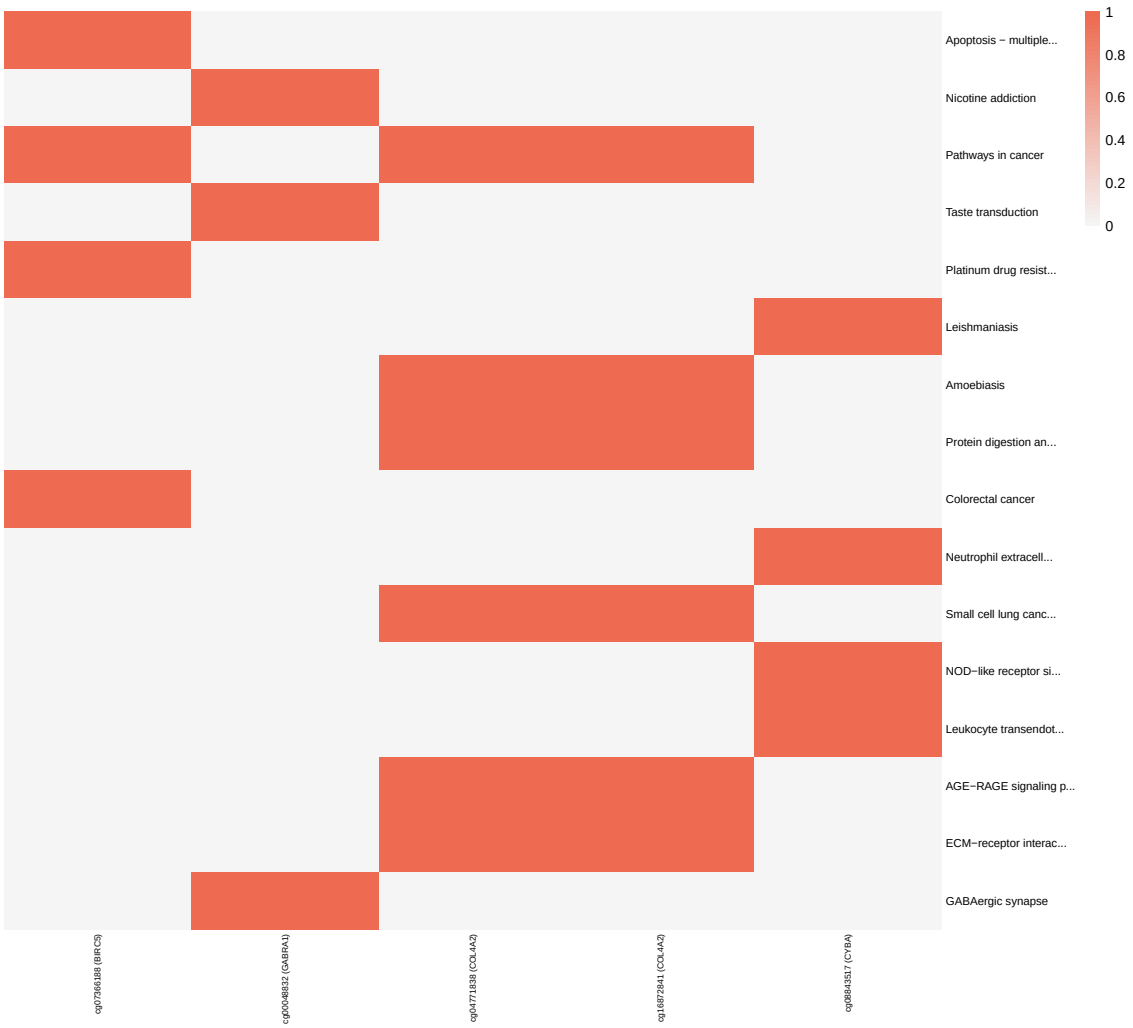


Figure 4.12: Astrocytoma CpG KEGG pathways enrichment analysis

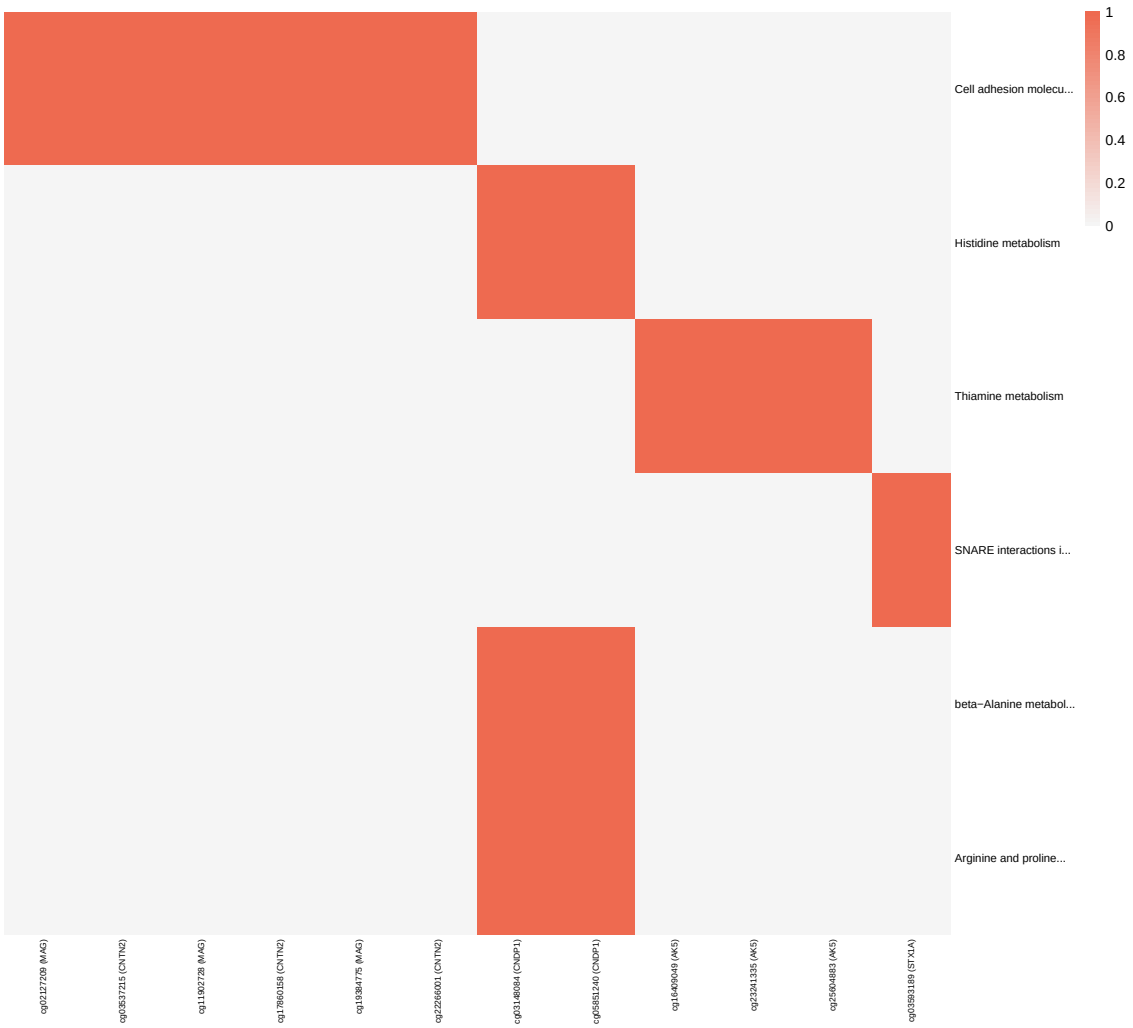


Figure 4.13: Oligodendroglioma CpG KEGG pathways enrichment analysis

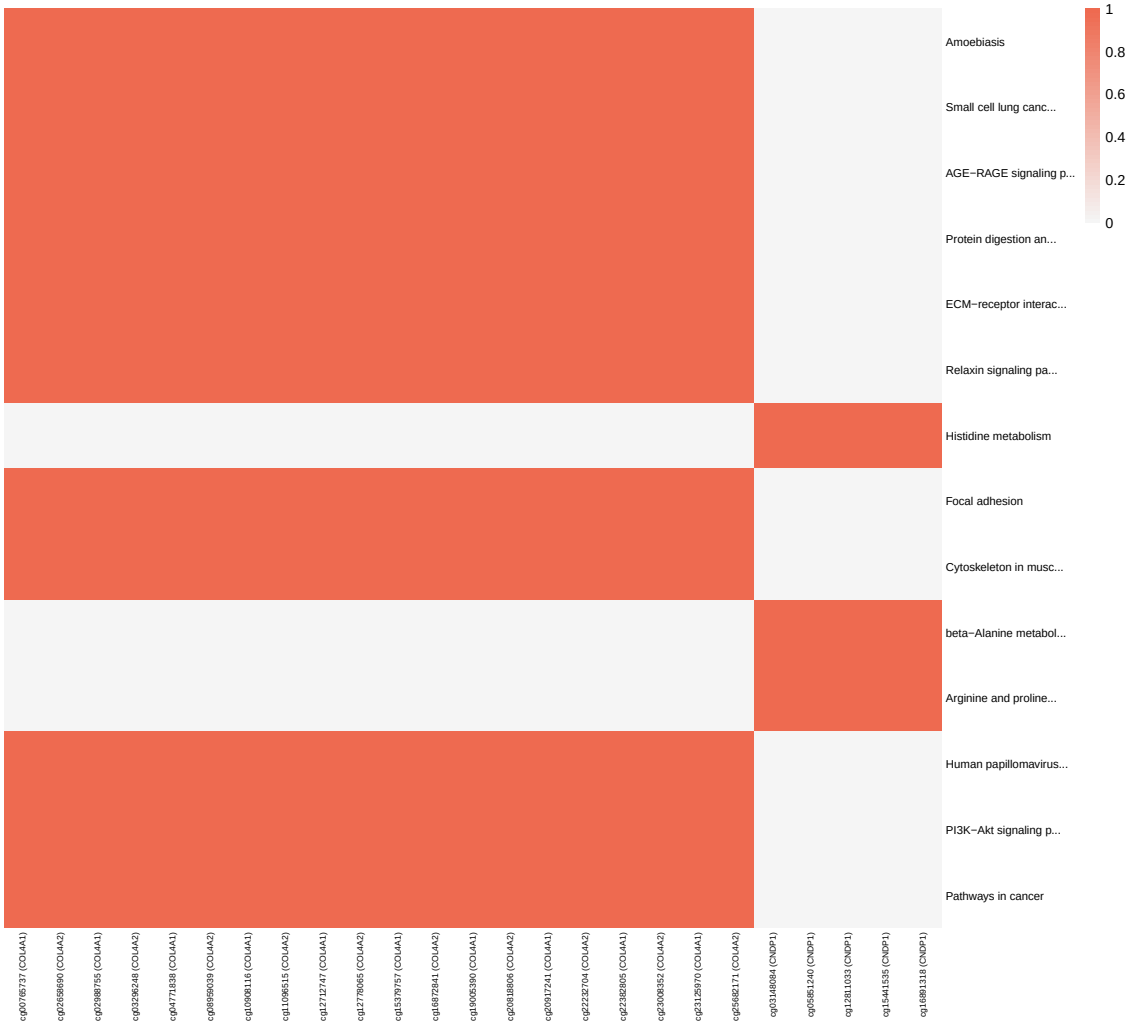


Figure 4.14: GBM CpG KEGG pathways enrichment analysis

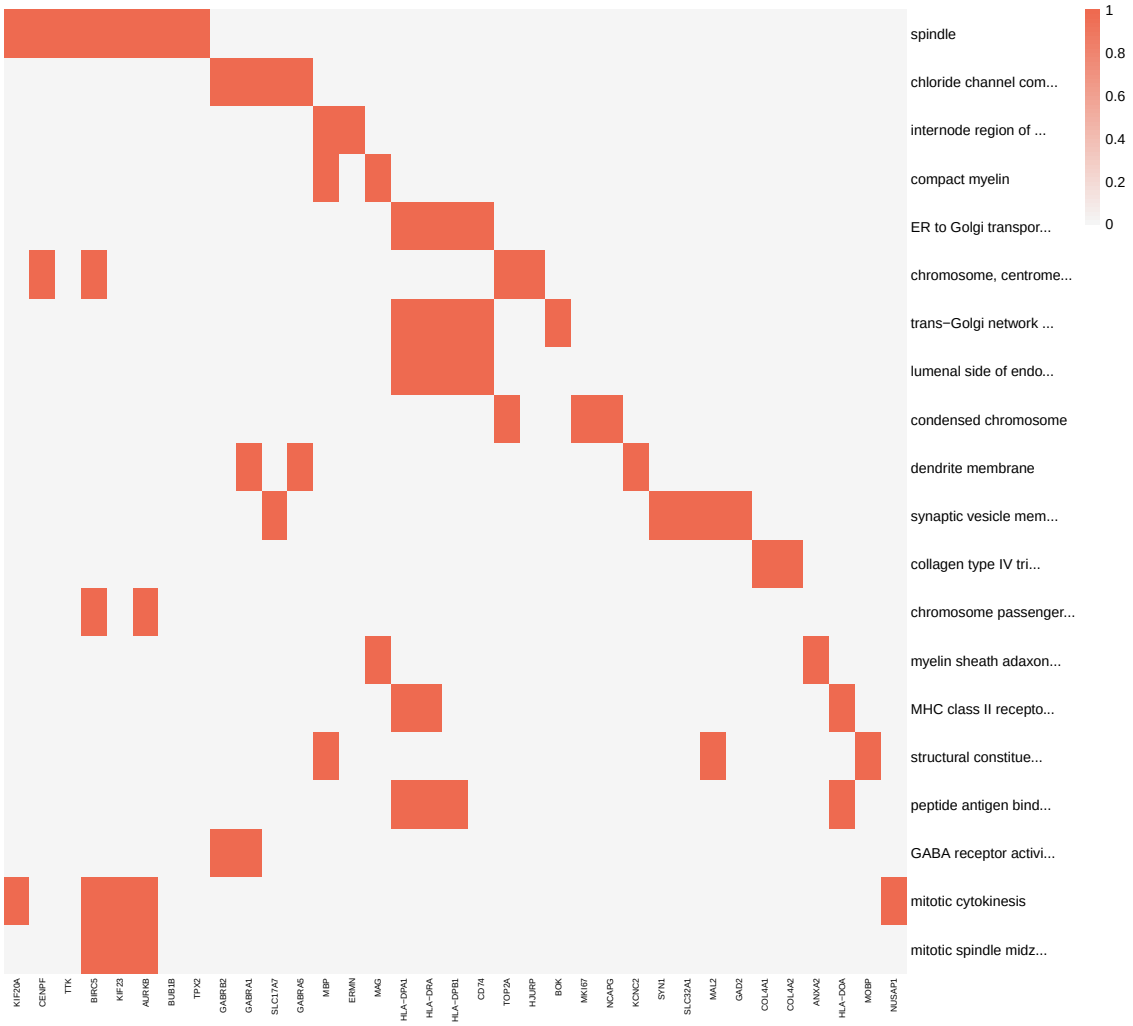


Figure 4.15: Astrocytoma genes Gene Ontology enrichment analysis

4.3.2 Curated assessment of the networks

To validate the biological relevance of the networks derived from the causal discovery algorithms, the identified subgraphs were compared against curated gene interaction databases. This comparison aimed to assess the validity of the causal relationships between genes in the subtypes.

In the astrocytoma network, the subgraph involving interactions between *HLA* genes (*HLA-DOA*, *HLA-DPB1*, *HLA-DPA1*, *HLA-DRA*) and *CD74* was analyzed using the UCSC genome browser. It was found that all *HLA* genes have a causal influence in *CD74*, as confirmed by the REACTOME database. Nonetheless, no curated information concerning the causal directions between the *HLA* genes was found. For the subnetwork comprising the genes *CCL3*, *CCL4*, and *CCL4L2*, the database confirmed an interaction between the genes *CCL3* and *CCL4*, despite the directionality being unclear. These two genes also share a connection with *CCR5*, which is causally influenced by *CCL4L2*. Regarding the

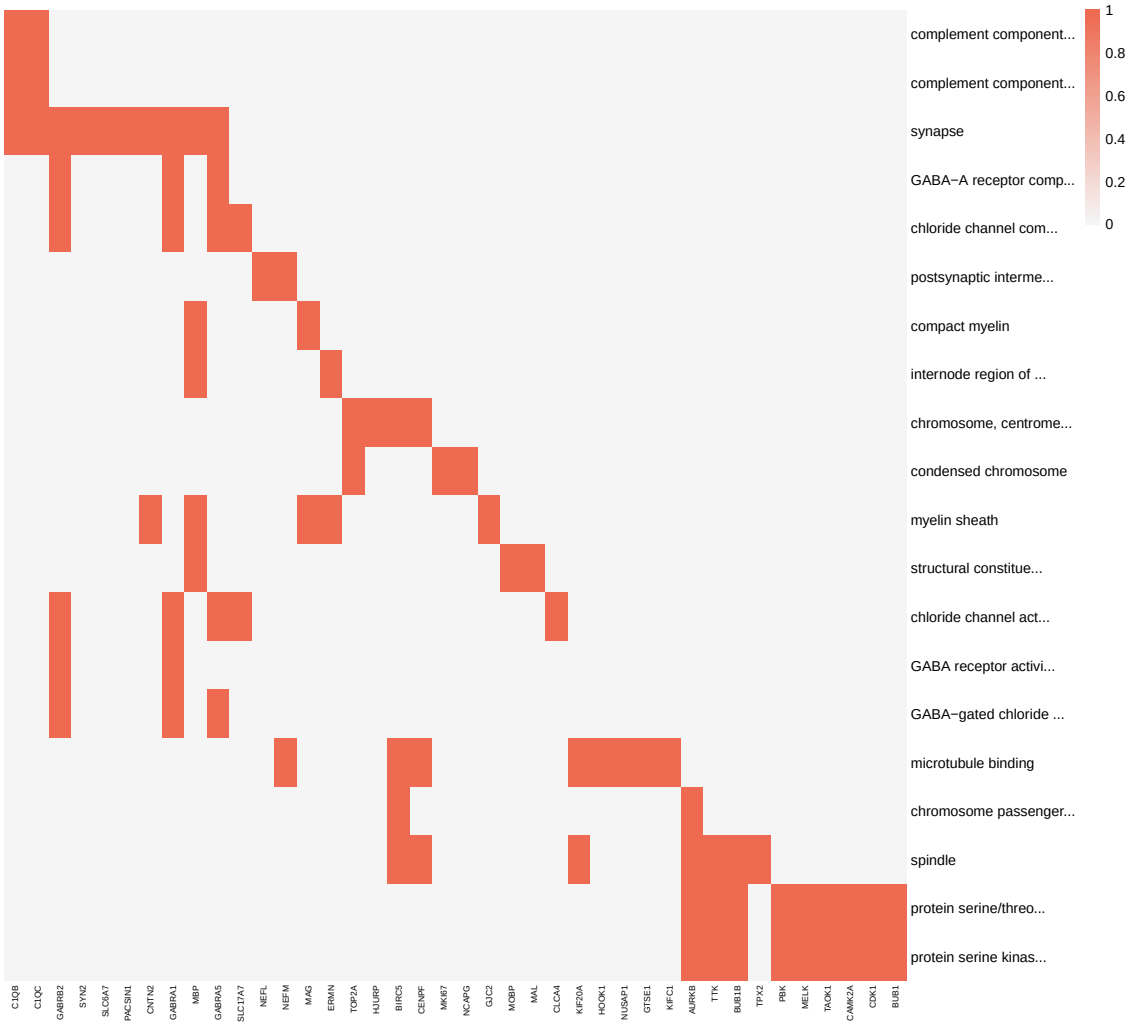


Figure 4.16: Oligodendroglioma genes Gene Ontology enrichment analysis

subnetwork comprised of *TYROBP*, *CYBA*, *SPI1* and *FERMT3*, the STRING database confirmed a co-expression between these features, despite not providing information on causal relationships. Furthermore, the STRING database confirmed an interaction between *MBP* and *MOBP*, possibly through the gene *APP*, despite the latter not being included in the context of this analysis. It is also worth pointing out that despite not being related in the output causal graph, curated information states that the gene *BIRC5* has a causal influence on both *CENPF* and *BUB1B*. *BIRC5* also has a verified causal interaction with *EPR1* according to the STRING database.

The subgraph entailing the interactions between *C1QA*, *C1QB*, and *C1QC* was validated via the UCSC genome browser. Accordingly, *C1QA* exerts a causal influence on both *C1QB* and *C1QC*, while *C1QB* directly influences *C1QC*. However, the role of *LAPTM5* in this subnetwork remains unclear. Despite the relationship between the genes *COL4A1* and *COL4A2* could not be validated, the database previously mentioned identified interactions between these genes and *APP*, suggesting that *APP* might also act as an unknown confounder. It was also reported from the database that there is an interaction between

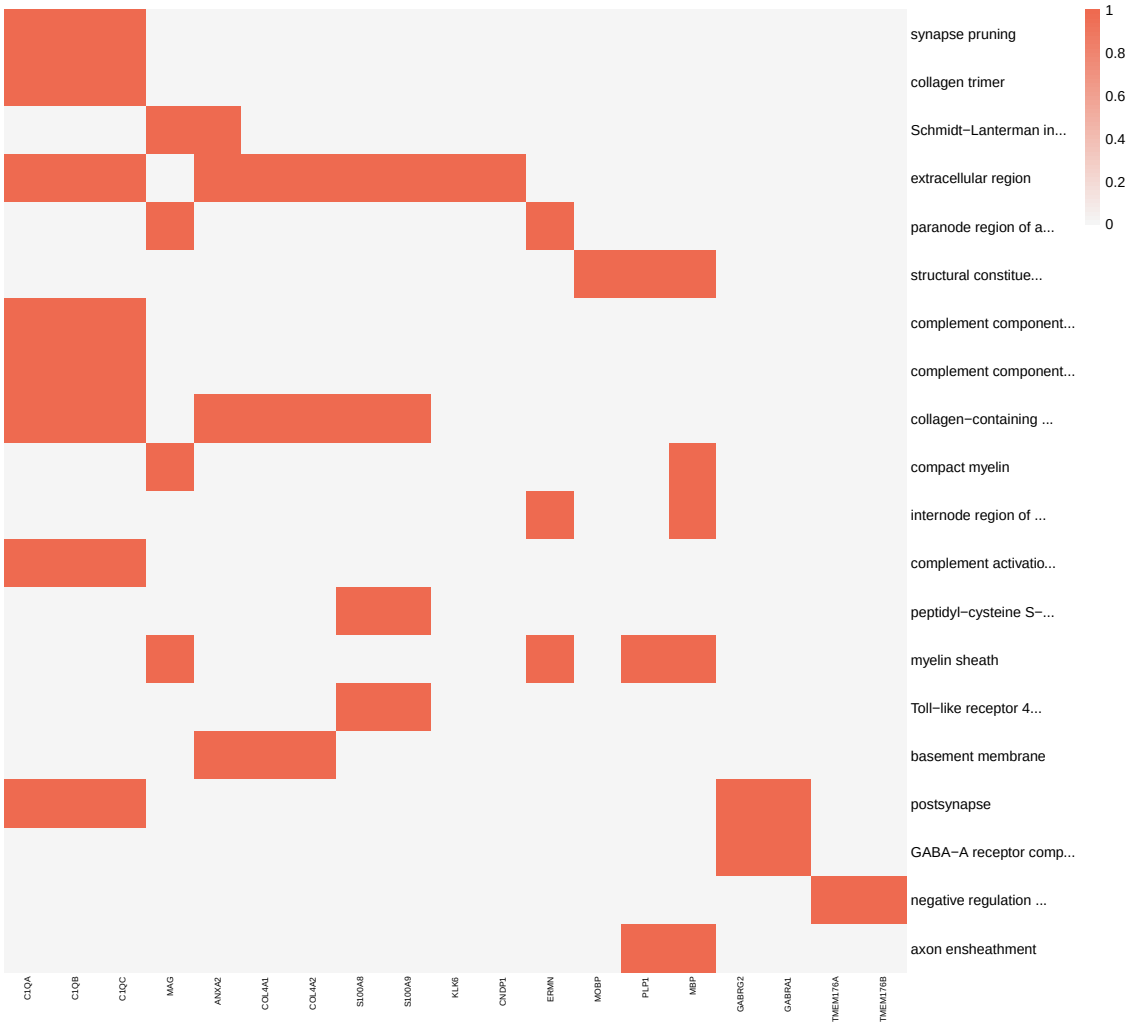


Figure 4.17: GBM genes Gene Ontology enrichment analysis

KLK6 and *COL4A1*, again involving *APP* as the common link.

For the GBM network, apart from the previously mentioned interactions, the gene *APP* was again implicated as a possible confounder in the interaction between *CNDP1* and *KLK6*. Additionally, *GABRA1* was found to exert a causal influence on *GABRG2*, with *GABRB1* acting as a potential confounder. The STRING database also confirmed the causal influence of *MOG* in *MAG*. For GBM-specific genes *S100A8* and *S100A9*, the UCSC genome browser showed a shared interaction, though the causal direction remains unclear. Both genes also target *CYBA*, a gene exclusively selected for astrocytoma. On top of that, the previously mentioned database also carries evidence for the interaction between *ANXA2* and both *S100A8* and *S100A9*.

Considering the graph that displays the causal interactions between the selected features for oligodendroglioma, besides the interactions already discussed in the previous subtypes, it could only be verified from the UCSC genome browser dataset that the causal influence of *GNG3* in *GBRB2*. However, this interaction is not exhibited in the final graph.

No significant findings were obtained from the Connectivity Map platform, primarily

due to limited gene availability. In addition, for the unmentioned subnetworks, no significant interactions were identified in the designated databases for any of the subtypes.

4.4 Discussion

The inherent high heterogeneity and proliferative nature of glioma pose significant challenges in applying effective therapeutical strategies and establishing a universal classification criterion for distinguishing the different subtypes. The last classification already emphasises the molecular and genetic properties that differentiate each subtype, which provides an excellent opportunity to apply in-silico methods to assess the discriminating roles of specific genetic features in each subtype. The availability of molecular information data from both GBM and LGG patient data from publicly available platforms like TCGA (TCGA-LGG, TCGA-GBM) grants the chance to apply computational methods to better understand the underlying disease mechanisms, identify potential biomarkers and theorize about novel therapeutical approaches. Taking that into consideration, in this research, we leveraged the last revised glioma classification applied to the TCGA data [7, 26] to uncover the GRN governing different glioma subtypes, through the integration of methylation and gene expression data.

4.4.1 Overall performance of the algorithms

Using iDINGO, differential networks for each glioma subtype were retrieved, integrating methylation and RNA-seq data, with causal interpretations in light of the Markov property for chain graphs [88]. The results clearly demonstrate that there is a distinct differential methylation regulating gene expression across the different subtypes, as evidenced by the prevalence of unique edges (green edges) in the final graphs for each subtype (4.6, 4.8, 4.7). This highlights the potential of iDINGO to capture subtype-specific regulatory mechanisms.

From evaluating the performance of the distinct causal discovery algorithms, the flexibility and robustness of the FCI-algorithm stood out. Attending the adversities associated with the modelling of GRN, particularly the potential existence of unknown confounders [66], it is relevant that the FCI emerged as the most stable algorithm across all data types and subtypes. Such can be likely attributed to its capacity to avoid assumptions of causal sufficiency and parametric data assumptions, making it well-suited for the challenges of biological data.

The PC-stable algorithm also presented consistent results, although the causal sufficiency assumption must have impacted its stability when compared to FCI. In contrast, the GES algorithm performed the worst across all metrics regarding every subtype. This must be mainly due to its reliance on the (BIC) scoring criterion, which assumes a linear-Gaussian data distribution. This assumption does not align well with the high-dimensional and often non-linear nature of biological data. A non-linear and non-Gaussian scoring

criterion was also employed [79]. Still, the complexity of the kernel methods utilized in this approach proved to be prohibitive for the inherent high-dimensionality of the data in hand.

Algorithms based on FCM also did not bear stable results regarding the evaluation metrics. The Direct-LiNGAM had a considerably poor performance across all evaluation metrics for all subtypes. This is likely due to the linearity and causal sufficiency assumptions, which are both problematic when applied to biological systems. Meanwhile, the ANM model produced mixed results. Despite producing results with overall low SHD values, it underperformed on precision, recall and FDR. One potential justification for the performance of ANM models is its reliance on a bivariate framework, which depends heavily on the initial skeleton discovered by the PC algorithm. Given that the PC algorithm is in itself prone to instability attending its causal sufficiency assumption, any inaccuracies in the skeleton can condition the final model. Apart from the reasons previously discussed in ANM that can explain the performance of the PC-LiNGAM algorithm, the inherent Gaussianity assumption in the Fisher Z test might also have contributed to its limitations.

4.4.2 Biological interpretation of the networks

Distinct subnetworks were identified as relevant for each subtype or across different subtypes according to the network discovery approaches and subsequent validation with bioinformatic methods. Unique astrocytoma subnetworks comprising the *HLA* genes (*HLA-DOA*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*) and *CD74*, besides *CCL* genes (*CCL3*, *CCL4* and *CCL4L2*) were identified as potential biomarkers. Similarly, in all subtypes, subnetworks involving *MOBP*, *MBP* and *ERMN*; *C1Q* genes (*C1QA*, *C1QB* and *C1QC*), and *GABR* were spotted. The connection involving the genes *ANXA2* and *ANXA2P2*, as well as the interaction between *COL4A1* and *COL4A2*, were highlighted for both astrocytoma and GBM. In GBM, the interaction between *S100A8* and *S100A9* appeared to be a significant prognostic marker. Furthermore, *CYBA* was found to be relevant in both astrocytoma and GBM, while *STX1A* was highlighted as a key biomarker for oligodendroglioma.

HLA* genes interaction with *CD74 The subnetwork consisting of *HLA* genes (*HLA-DOA*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*) and *CD74* is epigenetically regulated according to iDINGO (4.1a, 4.6). Validation of these interactions from the UCSC Genome Browser confirms that while direct influences between the *HLA* genes are not fully known, all selected *HLA* genes exert a direct causal influence on *CD74*. According to the iDINGO outcome (4.6), the methylation site relative to *CYBA*, cg08843517, directly influences all the genes in this subnetwork, except *HLA-DRA*, which, appears unaffected by any methylomic site in this analysis.

Ontology analysis performed on astrocytoma with GOATOOLS, revealed that the genes in this network are enriched for antigen processing and presentation of exogenous peptide antigen via Major Histocompability Complex (MHC) class II (4.15, E.7). All genes,

except *CD74*, were also enriched for peptide antigen assembly with MHC class II protein complex, positive T cell activation regulation, and positive immune response regulation. As a known MHC class II chaperone, *CD74* plays an important role in antigen presentation, while *HLA* are crucial in the processing of antigens for MHC class II complexes [151, 152].

The expression of both *CD74* and *HLA* genes has been proposed to enhance antitumor immunity, by presenting tumor cell-derived antigens to $CD4^{+}$T cells, leading to tumor cell elimination [153, 154]. Hence, the expression of both genes could serve as possible biomarkers for a positive outcome. It has also been reported that the expression of these genes might be a good marker for a positive response to anti-Programmed Cell Death protein-1 (PD-1) therapies [152, 153]. Given the lack of success of the application of anti-PD-1 to GBM (where safety was demonstrated, but survival has not improved) [155], this could be attributed to the lack of MHC class II related genes expression. Thus, it might be hypothesized that anti-PD-1 therapies could be more effective in astrocytoma patients. In addition, considering the influence of the methylomic site mapped to *CYBA* gene in this subnetwork (cg08843517), it would be valuable to further investigate how methylation at this site affects the expression of these MHC class II related genes. This could provide additional insights into the regulatory mechanisms that drive their expression and their role in antitumor immunity.

***CCL3*, *CCL4*, and *CCL4L2* subnetwork** The subnetwork comprising *CCL3*, *CCL4* and *CCL4L2* is characterized by bidirected edges, identified exclusively in the astrocytoma causal graph (4.3a, 4.6). According to the validation from the UCSC Genome Browser, an interaction exists between *CCL3* and *CCL4*, despite unclear directionality. Interestingly, the three genes converge in their causal influence towards *CCR5*, which might suggest *CCR5* as a potential unknown confounder in this subnetwork. All genes in this subnetwork are chemokines, meaning that they function as chemoattractants, stimulating the migration of leukocytes and mediating inflammatory processes [156]. Given that *CCR5* is a receptor of all these three chemokines [157], it might explain its role as a confounder in this subnetwork and hint a potential regulation role.

The role of both genes in cancer remains dubious. On the one hand, some studies point out the role of *CCL3* and *CCL4* in recruiting immune cells to the tumor microenvironment, and with that leveraging antitumor cellular response [158]. On the other hand, it has been reported that the activity of these chemokines could also enhance the proliferative nature of cancer cells [159, 160].

Regarding functional enrichment, the term positive regulation of calcium-mediated signalling was significantly associated with this subnetwork (4.15, E.7). This suggests potential therapeutical relevance, as increased calcium levels in the tumor microenvironment may help recruit $CD8^{+}$ tumor infiltrating antigens, potentially reversing the immunosuppressive glioma microenvironment [161].

According to iDINGO these genes are targets of the CpG site mapped to *CYBA* (cg08843517) (4.1a,4.6). As suggested for the MHC class II genes subnetwork, it would

be valuable to investigate how methylation at this site could influence the expression of these chemokines and impact their potential role in modulating antitumor processes.

***MOBP*, *MBP* and *ERMN* subnetwork** The causal link between *MBP*, *MOBP*, and *ERMN* reveal different dynamics across the different glioma subtypes. In astrocytoma, the RNA causal graph (4.6, 4.3a) shows that *MBP* and *MOBP* share a bidirected edge, with *MOBP* exerting a causal influence on *ERMN*. Additionally, *ERMN* exerts a causal influence on *ENPP2* and is directly affected by *CNTN2*. From the iDINGO analysis, only *ERMN* was revealed to be a target of causal influence from CpG sites, namely cg04771838 (*COL4A2*), cg08843517 (*CYBA*) and cg27180880 (*MBP*). From the functional enrichment analysis, the site cg27180880 (*MBP*) was enriched for the ontology term interleukin-6 production (4.9, E.1). Interleukin-6 (IL-6) is a pro-inflammatory cytokine which intervenes in the inflammatory microenvironment of tumours, reported to have an essential role in the proliferation, invasion and migration of cancerigenous cells of glioma [162, 163], primarily through the JAK/STAT3 signalling pathway, a known pathway responsible for tumor development and migration in GBM. Although, the role of this pathway in astrocytoma remains to be assessed [162]. Further investigation could elucidate about the methylation of the gene *ERMN* in the context of the activation of this signalling pathway. From the KEGG pathways enrichment analysis (4.12,E.4) it was verified that the term pathways in cancer was enriched. Such might suggest that the methylation of these gene may contribute to a more aggressive tumor phenotype.

In the GBM causal graph, (4.8, 4.5a), the *MBP* gene is absent, though the causal influence of *MOBP* on *ERMN* remains intact. In turn, *MOBP* is subject to a causal influence from *CARNS1*. Despite the absence of *MBP* in GBM causal network, a CpG site mapped to *MBP* (cg02072495), has a direct influence on *MOBP*, according to the iDINGO output (4.1b, 4.8). Ontology enrichment analysis (4.11, E.2) revealed that methylation at this site is enriched for terms such as negative regulation of multicellular organismal process and myelin sheet-related processes. Additionally, methylomic sites located in *ERMN* were also enriched for terms concerning the myelin sheet (E.2). Both *ERMN* and *MOBP* are components of the myelin sheet, and while their role in the context of GBM is not well established, these genes have already been associated with glioma progression [164].

In the oligodendroglioma causal graph (4.7, 4.4a), *ERMN* has a causal influence in *TMEM144*, and in *MOBP*, which in its turn has a direct influence in *MBP*, mirroring the interaction observed in astrocytoma. From iDINGO output it is clear to see that the gene *TMEM144* is heavily influenced by methylomic sites, unlike other genes of this subnetwork, in this subtype (4.1c, 4.7). From enrichment analysis, no relevant conclusions were extracted beyond the expected enrichment for terms related to the myelin sheath (4.10, 4.16,E.3, E.9). Understanding the role of these three genes, and how their interactions vary across different subtypes might be very important in the context of this disease. While their involvement in glioma is yet to be fully detailed, their association with myelin

function and glioma progression may suggest these three genes as potential biomarkers for glioma.

***C1Q* genes subnetwork** The subnetwork composed of *C1QA*, *C1QB*, and *C1QC* appears in all glioma subtypes, despite different interactions along the subtypes. In astrocytoma, the causal algorithm suggests the possibility of hidden confounders attending the bidirected edges between them, except the connection between *LAPTM5* and *C1QC*, which is direct (4.6, 4.3a). According to the UCSC Genome Browser, there is a direct interaction from *C1QA* to *C1QB* to *C1QC*. However, the interaction with *LAPTM5* could not be confirmed, despite the search performed in STRING suggesting a co-expression with the *C1Q* genes. CpG sites did not appear to differentially influence these genes in astrocytoma (4.6, 4.1a), and no significantly enriched terms were found, except for synapse pruning (4.15,E.7).

In contrast, the enrichment analysis in GBM highlights an essential role of this subnetwork for this subtype. According to the causal graph returned (4.8, 4.4a), the subnetwork now lacks *LAPTM5* and remains with no causal directions, suggesting the influence of hidden confounders. In opposition to the iDINGO result from astrocytoma, this subnetwork was considered to be methylated by several CpG sites (4.8, 4.1b), potentially contributing to the aggressiveness of this subtype. From enrichment analysis, it was found out that methylomic sites regulating the genes from this subnetwork were the ones mapped to *COL4A1* and *COL4A2*, highly related to tumorigenic terms such as pathways in cancer, small cell lung cancer, and others highly related to GBM, namely the PI3K-AKT signalling pathway, and the Extracellular matrix (ECM)-receptor interaction (4.14, E.5). The PI3K-AKT signalling pathway is a crucial driver of tumor growth in GBM [47]. Besides, the ECM has been reported to contribute to therapy resistance and creating favorable niches for cancer stem cells, which play a crucial role in glioma progression and self-renewal properties [13, 165].

Further enrichment analysis in GBM revealed the enrichment of biological processes of synapse pruning and complement activation for the entire subnetwork (4.17, E.8). These genes play an essential role in the complement system, a functional component of the innate immune system and innate immunity, which has been linked with tumor cell proliferation and invasion, resistance to therapeutics and angiogenesis, and negatively correlated with glioma patient survival [166]. Given the known role of these genes in synapse elimination [167], the enrichment of synapse pruning in this subtype may allude that this mechanism could contribute to GBM progression.

Interestingly, considering the possible drugs targeting *C1QA* (the most ancestral node in the subnetwork in GBM), from DRUGBANK, the one that yields the best match is Cetuximab, a drug that binds to EGFR, inhibiting the binding of epidermal growth factor, and with that blocking that signalling pathway. Since EGFR is a critical driver in glioma signalling [40, 44], this subnetwork could serve not only as important GBM biomarkers but also a potential therapeutical target for this drug.

In oligodendroglioma, only *C1QB* and *C1QC* are present in the subnetwork (4.4a, 4.7), and no direct connection between them was retrieved, again hinting at a potential hidden confounder influencing this interaction. No terms from enrichment analysis were considered to be enriched for these genes. Although not displaying a significant role for this subtype as they did for GBM, the presence of these complement system genes across every subtype can indicate an important role in glioma.

***GABR* genes subnetwork** The presence of *GABRA1* across all glioma subtypes, with bidirected interaction with *GABRB2* in astrocytoma and oligodendroglioma, and with *GABRAG2* in GBM (4.6, 4.8, 4.7), highlights a potentially significant subnetwork. In astrocytoma, according to the iDINGO graph (4.1a, 4.6), *GABRB2* is targetted by CpG sites, including cg04771838 (*COL4A2*), which is enriched for KEGG pathways terms such as pathways in cancer, small cell lung cancer and ECM-receptor interaction (4.12, E.4). As discussed previously for the *C1Q* subnetwork, the enrichment of these terms might indicate the role of *GABRB2* and the related subnetwork in astrocytoma progression [13, 165].

This subnetwork functional enrichment for GABAergic synapse-related terms across subtypes (4.15,4.17,4.16,E.7,E.9, E.8) underlines its involvement in glioma. From the literature, it was found that the presence of *GABA* related genes is generally associated with a favourable overall prognosis in glioma patients [168] and that the genes identified for each subtype were correlated with improved overall survival, even in GBM [169]. All in all, investigating the role of differential methylation in the regulation of this subnetwork may reveal potential survival mechanisms in glioma and clarify the subgraph's role in different methylation conditions.

Oligodendroglioma subnetworks In the oligodendroglioma graph (4.1c,4.7), aside from the subgraphs shared with other subtypes, no other subnetworks could be validated by curated pathway databases. This could indicate either the novelty of this approach in the characterization of the oligodendroglioma landscape or allude that the causal algorithm applied was less effective for this subtype. The stability results (I.4) showed lower robustness compared to astrocytoma astrocytoma (I.1) and GBM (4.2), which solidifies this hypothesis.

From the iDINGO analysis, *STX1A* is highlighted as a differential target of several methylomic sites in this subtype (4.1b, 4.7). It is noteworthy to verify that the CpG site cg05851240 (*CNDP1*) exerts a widespread direct influence in most genes in this subtype, suggesting a potentially crucial role in regulating gene expression in oligodendroglioma. KEGG pathways enrichment analysis (E.6, 4.13), reveals enrichment of terms related to diverse metabolic pathways, such as beta-Alanine, arginine and proline, thiamine and histidine to this CpG site. Recent studies have proposed the role of different metabolic pathways in the differentiation of the subtypes in glioma, even proposing the role of metabolic shifts in early tumorigenesis [170, 171].

Interestingly, GBM also denotes enrichment for *KEGG* pathways terms related to arginine and histidine; arginine and proline; and beta-Alanine metabolism (E.5, 4.14), while astrocytoma lacks the enrichment of such terms (E.4, 4.12). For being exclusively enriched in oligodendroglioma, thiamine metabolism may be a relevant prognostic marker in oligodendroglioma. The methylation of the CpG site mapped to *CNDP1* (cg05851240) and its associated genes from iDINGO could serve as important biomarkers, along with *STX1A*, given its significant methylomic influence.

Interaction between S100A8 and S100A9 In GBM, a significant bidirectional interaction exists between *S100A8* and *S100A9*, with the potential involvement of *CYBA* as a hidden confounder, suggesting its relevance in GBM. According to iDINGO, these genes are highly differentially methylated by several methylomic sites (4.1b, 4.8), indicating their potential role in GBM aggressiveness. These genes constitute the calprotectin complex [172], whose function is to bind zinc and calcium ions with an essential role in inflammatory response [173].

One of the primary receptors of this complex is the Toll-like receptor 4 (TLR4), a term that was enriched for both genes(4.17, E.8). It has been reported the important role of TLR4 downregulation in chemoresistance and tumor immune invasion in GBM [174]. Thus, the assessment of the differential methylation of these complex might be a key prognostic for TLR4 signalling pathway activation and therapy resistance.

ANXA2 interaction with ANXA2P2 The gene *ANXA2* shares a connection with its pseudogene *ANXA2P2*, in both astrocytoma and GBM (4.6, 4.8). From the iDINGO result, it is clear the negative partial correlation between *ANXA2P2* and *ANXA2* in both astrocytoma and GBM ((4.1a, B.1, B.3)). This might suggest a potential inhibitory role of either *ANXA2P2* or *ANXA2* on each other, possibly mediated by an unknown confounder.

According to the causal relationships identified by iDINGO, the methylomic site cg08843517 (*CYBA*) besides regulating *ANXA2*, also interacts with *ANXAP2* (4.1a, B.1). Notably, this site shows a differential regulatory pattern, exerting a positive influence on *ANXA2* while negatively affecting *ANXA2P2* (4.1a, B.1). This regulation pattern extends to other CpG sites, such as cg07366188 (*BIRC5*), cg10792225 (*KIFC1*), and cg27180880 (*MBP*), which exhibit a positive influence on *ANXA2* while negatively affecting *ANXA2P2* (4.1a, B.1). Interestingly, the causal influence of cg27180880 and cg04771838 on both genes is also present in GBM (4.1b, B.3). These inverse CpG sites' regulation of both genes further adds up to the hypothesis of the negative regulation between *ANXA2P2* and *ANXA2*, which might be influenced by differential methylation in astrocytoma and GBM.

The enrichment analysis revealed that only the CpG sites differentially affecting these genes in GBM were enriched for GO terms (4.11, E.2), namely terms associated with myelin, and negative regulation of multicellular organismal process. Considering that, some articles already noticed the association between the expression of *ANXA2* and *ANXA2P2* with the worst prognosis and survival in glioma [175], besides also highlighting the role

of *ANXA2P2* in the PI3K signalling pathway, a known driver of tumor growth in GBM [176]. The former article also underscores that the knockdown of *ANXA2P2* inhibited the PI3K pathway.

With that, targeting the referenced methylomic sites and leveraging the regulatory balance between *ANXA2* and *ANXA2P2* might be relevant for future therapeutical approaches to GBM and also to further understand the progression and aggressiveness from LGG to GBM.

COL4A1 interaction with COL4A2 The methylomic regulation of *COL4A1* and *COL4A2* shows some relevant parallels to the patterns observed for *ANXA2* and *ANXA2P2*. In astrocytoma, both *COL4A1* and *COL4A2* are regulated by common CpG sites, where *COL4A1* is positively regulated by cg04771838 (*COL4A2*) and negatively influenced by cg08843517 (*CYBA*), while *COL4A2* exhibits an inverse relation (4.6, 4.1a). Noteworthy, cg04771838 also regulates *COL4A1* in GBM, but not *COL4A2*, according to the retrieved causal graph for this subtype (4.8)). Such might suggest a subtype-specific epigenetic influence and a potential role of *COL4A1* in more aggressive phenotypes. According to the causal map of the causal interactions between CpG sites in astrocytoma (4.3b, 4.6), it is visible the direct influence of cg16872841 (*COL4A2*) and cg00765737 (*COL4A2*) in cg04771838 (*COL4A2*), while in GBM the direction of the interaction between these three features is not clear (4.5b, 4.8). According to the TCGA CpG to gene mapping (D.1), both methylomic features that interact with cg04771838 are associated with *COL4A2*, hinting at a possible epigenetic feedback loop influencing the regulation of these genes, especially in GBM, where the directionality is not clear. However, the presence of a hidden confounder may also impact the interaction between *COL4A1* and *COL4A2* and their roles in both subtypes. According to the analysis via UCSC Genome Browser, the *APP* gene was identified, as for *ANXA2* and *ANXA2P2*, as a potential latent variable impacting these interactions. A further study of the role of these genes in the interaction between these genes could explain some of the complex regulatory patterns observed.

Supporting the differential roles of these genes in astrocytoma and in GBM, the enrichment analysis applied to *COL4A1* and *COL4A2*, and their associated methylomic sites reveal interesting subtype-specific differences. Noteworthy, in astrocytoma, the only CpG site that was enriched for the term pathways in cancer, namely cg04771838, is the one that regulates *COL4A1* in GBM (4.12,E.4). Moreover, cg16872841, another methylomic site associated with the *COL4A1-COL4A2* interaction, is enriched, in astrocytoma and GBM, for the KEGG pathways term ECM-receptor interaction (4.12,E.4,4.14, E.5). This is noteworthy, as the ECM has been reported to contribute to therapy resistance and creating favorable niches for cancer stem cells, which play a crucial role in glioma progression and self-renewal properties [13, 165].

The importance of the methylomic regulation of these genes is even more apparent in GBM. All CpG sites linked to *COL4A1* and *COL4A2* in GBM (4.14,E.5) were enriched for the KEGG term PI3K-AKT signalling pathway, a crucial driver of tumor growth in GBM

[47]. This enrichment was not observed in astrocytoma (4.12,E.4). In addition, all of these CpG sites were also enriched for the terms pathways in cancer and small cell lung cancer (4.14,E.5). Given the established connection between the PI3K-AKT signalling pathway and glioma aggressiveness, *COL4A1* and *COL4A2*, much like *ANXA2* and *ANXA2P2* may play a role in regulating tumor growth and poor prognosis in glioma [177].

Role of *CYBA* in glioma In astrocytoma, *CYBA* is epigenetically regulated and forms a subgraph with *TYROBP*, *SPI1*, and *FERMT3*, that was only discovered for this subtype. Regarding this subnetwork, their co-expression was validated according to STRING database analysis, suggesting the potential existence of unknown confounders influencing this subgraph (4.6, 4.3a). From functional enrichment analysis, *CYBA* is linked to regulation of IL-6 production in 4.9, through the methylomic site cg04771838 (*COL4A2*), whose differential direct influence was verified in iDINGO analysis (4.1a, B.1). The production of IL-6, a pro-inflammatory cytokine which intervenes in the inflammatory microenvironment of tumours, is reported to have an essential role in the proliferation, invasion and migration of cancerigenous cells of glioma [162, 163]. Its mechanism of action in glioma affects the JAK/STAT3 signalling pathway, a known pathway responsible for tumor development and migration in GBM, with unclear role in astrocytoma [162]. The methylomic site that was enriched for this term, was also differentially expressed in GBM.

In addition, *CYBA*'s association with neutrophil extracellular trap formation hints at its involvement in heterogeneity of the tumor microenvironment. It has been reported the relevance of neutrophil extracellular trap formation in tumor growth, metastasis and therapy resistance, and in glioma to promote cell proliferation, invasion and migration through the activation of the PI3K signalling pathway [178, 179].

With that, assessing the role of *CYBA* and its subnetwork in both astrocytoma and GBM could hint about the progression of astrocytoma into more aggresive phenotypes.

5

Conclusion and future work

This chapter describes the conclusions and future work from this research.

5.1 Conclusions

The glioma classification landscape has changed immensely over the last few years, mainly due to a more significant focus on molecular attributes, allowing for a more accurate stratification of the subtypes. Moreover, the availability of public datasets containing molecular data from patients affected by this disease provides an enormous opportunity to research and discover novel biomarkers and suitable therapeutical approaches to the different glioma subtypes. With this, applying causal discovery methods proves to be an essential step towards unveiling signalling pathways impacting the growth, proliferation, survival and resistance to therapeutical approaches that characterize each subtype and improve the chances of patient survival.

In this research, the TCGA-LGG and TCGA-GBM gene expression and methylomic data were utilized, with the patient classification according to the updated classification from the 2021 WHO guidelines [7, 26], to discover regulatory networks that could characterize each subtype, and further give an insight into the pathways differentiating astrocytoma, glioblastoma, and oligodendroglioma. The gene expression data and methylation data were integrated with iDINGO, which allowed us to infer differentially expressed features for each subtype, also providing information regarding the causal influence of CpG sites in the gene expression features. Afterwards, causal discovery methods were applied to discover the causal networks in transcriptomics and methylomics data, where the most robust method was FCI, hence the one used for further analysis.

This pipeline discovered several subnetworks and possible biomarkers for each subtype. Regarding astrocytoma, it was found that the methylomic site cg08843517 (*CYBA*) regulated the majority of the genes and subnetworks in this subtype and that the methylomic site cg04771838 (*COL4A2*), besides being also expressed in GBM, was also associated with the genes that were also expressed in this subtype, hinting a more aggressive outcome related to the methylation from this site. Furthermore, the subnetwork comprising

the *HLA* genes and *CD74* was hypothesized to be associated with a positive result and a favourable response to anti-PD1 therapies. The subgraph containing the genes *CCL3*, *CCL4* and *CCL4L2* was also considered to be a promising biomarker of astrocytoma, given their association with the proliferative activity of the tumor microenvironment. The subgraph comprising the genes *MOBP*, *MBP* and *ERMN* revealed to be potential biomarkers across all subtypes, specially linked to glioma progression. In addition, the *C1Q* genes and the *GABA*-associated receptors were present in all subtypes, with differential methylation and enrichment terms in each subtype, hinting at potential biomarkers in glioma tumorigenesis. In oligodendroglioma, despite the non-validation of the majority of the subnetworks from curated pathways databases, the methylation of the CpG site cg05851240 (*CNDP1*), alongside the gene *STX1A* and the enrichment for thiamine metabolism might be potential unique biomarkers for this subtype. The calprotectin genes (*S100A8*, *S100A9*) were highlighted as potential unique GBM biomarkers, considering the role of differential methylation of these genes in GBM immune escape. From our research, the interactions between *ANXA2* and *ANXAP2*, as well as *COL4A1* and *COL4A2*, and their respective differential methylation, seemed to be reasonable prospects into the mechanisms underlying a more aggressive phenotype of this disease, denoting important roles in astrocytoma and GBM. The discovery of hidden confounders affecting these genes' interactions might consolidate their function in each subtype. The subgraph comprising the *CYBA* gene was also found to be relevant in activating the PI3K signalling pathway, a known player in the aggressiveness of glioma. Nevertheless, the exact role of the other genes in this subnetwork could not be elucidated by this study.

Despite good overall results, handling high-dimensional data presented some challenges. To address the well known difficulty of both iDINGO and causal algorithms methods in handling high-dimensional data [59, 89], the number of variables per subtype was substantially reduced. The significant number of discarded variables may have influenced the discovery of additional relevant subnetworks, and might explain the number of suggested hidden confounders in the retrieved causal graphs. In future research, a bigger pool of variables may hint at an even more significant amount of relevant networks per subtype. In addition, it is relevant to point out that the number of replicates chosen for the bootstrap implementation may have also impacted the stability performance of some algorithms. Hence, the robustness of proposed framework can benefit from an increased number of replicates.

5.2 Future work

Taking into consideration the advent of deep learning to retrieve causal information from data [180], such field could be explored in this context. The implementation of non-linear models to the full dataset instead of just a bivariate context may also benefit the analysis [181]. To tackle the possibility of the existence of feedback loops in GRN [66], incorporating the σ-separation instead of d-separation criterion to FCI could give a better insight into

the dynamics of the variables [182].

Building on this thesis, future research could resort to Pearl's do-calculus [63] to estimate causal effects within the networks, exploring potential interventions in the networks, essential to model candidate therapeutical approaches. Incorporating more omics, such as micro-RNA, proteomics, or metabolomics, could provide more insights into the dynamics of the different networks and unveil new suited biomarkers.

Bibliography

[1] J. M. Lourenço. *The NOVAthesis LaTeX Template User's Manual*. NOVA University Lisbon. 2021. URL: https://github.com/joaomlourenco/novathesis/raw/main/template.pdf (cit. on p. i).

[2] J. G. Nicholson and H. A. Fine. "Diffuse Glioma Heterogeneity and Its Therapeutic Implications". In: *Cancer Discovery* 11.3 (2021-03), pp. 575–590. ISSN: 2159-8274. DOI: 10.1158/2159-8290.CD-20-1474 (cit. on pp. 1, 4).

[3] D. N. Louis et al. "The 2021 WHO Classification of Tumors of the Central Nervous System: a summary". In: *Neuro-Oncology* 23.8 (2021-08), pp. 1231–1251. ISSN: 1522-8517. DOI: 10.1093/neuonc/noab106 (cit. on p. 1).

[4] M. Weller et al. "Glioma". In: *Nature Reviews Disease Primers* 1.1 (2015-07), p. 15017. ISSN: 2056-676X. DOI: 10.1038/nrdp.2015.17 (cit. on p. 1).

[5] R. Gue and D. A. Lakhani. "The 2021 World Health Organization Central Nervous System Tumor Classification: The Spectrum of Diffuse Gliomas". In: *Biomedicines* 12.6 (2024-06), p. 1349. ISSN: 2227-9059. DOI: 10.3390/biomedicines12061349 (cit. on pp. 1, 5, 7, 8).

[6] T. Nejo, A. Mende, and H. Okada. "The current state of immunotherapy for primary and secondary brain tumors: similarities and differences". In: *Japanese Journal of Clinical Oncology* 50.11 (2020-10), pp. 1231–1245. ISSN: 1465-3621. DOI: 10.1093/jjco/hyaa164 (cit. on pp. 1, 4).

[7] S. Martins, R. Coletti, and M. B. Lopes. "Disclosing transcriptomics network-based signatures of glioma heterogeneity using sparse methods". In: *BioData Mining* 16.1 (2023-09), p. 26. ISSN: 1756-0381. DOI: 10.1186/s13040-023-00341-1 (cit. on pp. 1, 6, 27, 61, 70).

[8] J. Kelly et al. "A review of causal discovery methods for molecular network analysis". In: *Molecular Genetics & Genomic Medicine* 10.10 (2022-10). ISSN: 2324-9269. DOI: 10.1002/mgg3.2055 (cit. on pp. 2, 11, 12).

[9] R. Coletti et al. "Inferring diagnostic and prognostic gene expression signatures across WHO glioma classifications: A network-based approach". In: (2023-05) (cit. on pp. 2, 29).

[10] M. Mendonça et al. "Updating TCGA glioma classification through integration of molecular profiling data following the 2016 and 2021 WHO guidelines". In: *bioRxiv (Cold Spring Harbor Laboratory)* (2023). DOI: https://doi.org/10.1101/2023.02.19.529134 (cit. on p. 2).

[11] A. Finch et al. "Advances in Research of Adult Gliomas". In: *International Journal of Molecular Sciences* 22.2 (2021-01), p. 924. ISSN: 1422-0067. DOI: 10.3390/ijms22020924 (cit. on pp. 4, 5).

[12] S. Lapointe, A. Perry, and N. A. Butowski. "Primary brain tumours in adults". In: *The Lancet* 392.10145 (2018-08), pp. 432–446. ISSN: 01406736. DOI: 10.1016/S0140-6736(18)30990-5 (cit. on p. 4).

[13] L. Barthel et al. "Glioma: molecular signature and crossroads with tumor microenvironment". In: *Cancer and Metastasis Reviews* 41.1 (2022-03), pp. 53–75. ISSN: 0167-7659. DOI: 10.1007/s10555-021-09997-9 (cit. on pp. 4, 65, 66, 68).

[14] D. N. Louis et al. "The 2007 WHO Classification of Tumours of the Central Nervous System". In: *Acta Neuropathologica* 114.2 (2007-08), pp. 97–109. ISSN: 0001-6322. DOI: 10.1007/s00401-007-0243-4 (cit. on p. 4).

[15] D. N. Louis et al. "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary". In: *Acta Neuropathologica* 131.6 (2016-06), pp. 803–820. ISSN: 0001-6322. DOI: 10.1007/s00401-016-1545-1 (cit. on p. 5).

[16] M. Ceccarelli et al. "Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma". In: *Cell* 164.3 (2016-01), pp. 550–563. ISSN: 00928674. DOI: 10.1016/j.cell.2015.12.028 (cit. on p. 5).

[17] M. J. van den Bent et al. "A clinical perspective on the 2016 WHO brain tumor classification and routine molecular diagnostics". In: *Neuro-Oncology* 19.5 (2017-05), pp. 614–624. ISSN: 1522-8517. DOI: 10.1093/neuonc/now277 (cit. on p. 5).

[18] D. Reuss. "Updates on the WHO diagnosis of IDH-mutant glioma". In: *Journal of Neuro-Oncology* 162.3 (2023-05), pp. 461–469. ISSN: 0167-594X. DOI: 10.1007/s11060-023-04250-5 (cit. on p. 5).

[19] Y. Felistia and P. Y. Wen. "Molecular Profiling and Targeted Therapies in Gliomas". In: *Current Neurology and Neuroscience Reports* 23.10 (2023-10), pp. 627–636. ISSN: 1528-4042. DOI: 10.1007/s11910-023-01299-7 (cit. on p. 5).

[20] D. R. Johnson et al. "A Radiologist's Guide to the 2021 WHO Central Nervous System Tumor Classification: Part I—Key Concepts and the Spectrum of Diffuse Gliomas". In: *Radiology* 304.3 (2022-09), pp. 494–508. ISSN: 0033-8419. DOI: 10.1148/radiol.213063 (cit. on p. 5).

[21] A. Perez and J. T. Huse. "The Evolving Classification of Diffuse Gliomas: World Health Organization Updates for 2021". In: *Current Neurology and Neuroscience Reports* 21.12 (2021-12), p. 67. ISSN: 1528-4042. DOI: 10.1007/s11910-021-01153-8 (cit. on pp. 5, 8).

[22] X. Guo et al. "Histological and molecular glioblastoma, IDH-wildtype: a real-world landscape using the 2021 WHO classification of central nervous system tumors". In: *Frontiers in Oncology* 13 (2023-07). ISSN: 2234-943X. DOI: 10.3389/fonc.2023.1200815 (cit. on pp. 5, 6).

[23] N. F. Brown et al. "Survival Outcomes and Prognostic Factors in Glioblastoma". In: *Cancers* 14.13 (2022-06), p. 3161. ISSN: 2072-6694. DOI: 10.3390/cancers14133161 (cit. on p. 6).

[24] A. Sharma and J. J. Graber. "Overview of prognostic factors in adult gliomas". In: *Annals of Palliative Medicine* 10.1 (2021-01), pp. 863–874. ISSN: 22245820. DOI: 10.21037/apm-20-640 (cit. on p. 6).

[25] K. Tomczak, P. Czerwińska, and M. Wiznerowicz. "Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge". In: *Współczesna Onkologia* 1A (2015), pp. 68–77. ISSN: 1428-2526. DOI: 10.5114/wo.2014.47136 (cit. on p. 6).

[26] M. Mendonça et al. "Updating TCGA glioma classification through integration of molecular profiling data following the 2016 and 2021 WHO guidelines". In: *bioRxiv (Cold Spring Harbor Laboratory)* (2023). DOI: https://doi.org/10.1101/2023.02.19.529134 (cit. on pp. 6, 27, 61, 70).

[27] B. Dewdney et al. "From signalling pathways to targeted therapies: unravelling glioblastoma's secrets and harnessing two decades of progress". In: *Signal Transduction and Targeted Therapy* 8.1 (2023-10), p. 400. ISSN: 2059-3635. DOI: 10.1038/s41392-023-01637-8 (cit. on p. 6).

[28] T. Komori. "Grading of adult diffuse gliomas according to the 2021 WHO Classification of Tumors of the Central Nervous System". In: *Laboratory Investigation* 102.2 (2022-02), pp. 126–133. ISSN: 00236837. DOI: 10.1038/s41374-021-00667-6 (cit. on p. 6).

[29] T. Cadoux-Hudson, C. J. Schofield, and J. S. McCullagh. "Isocitrate dehydrogenase gene variants in cancer and their clinical significance". In: *Biochemical Society Transactions* 49.6 (2021-12), pp. 2561–2572. ISSN: 0300-5127. DOI: 10.1042/BST20210277 (cit. on p. 6).

[30] A. K. Murugan and A. S. Alzahrani. "Isocitrate Dehydrogenase IDH1 and IDH2 Mutations in Human Cancer: Prognostic Implications for Gliomas". In: *British Journal of Biomedical Science* 79 (2022-01). ISSN: 2474-0896. DOI: 10.3389/bjbs.2021.10208 (cit. on p. 7).

[31] X. Du and H. Hu. "The Roles of 2-Hydroxyglutarate". In: *Frontiers in Cell and Developmental Biology* 9 (2021-03). ISSN: 2296-634X. DOI: 10.3389/fcell.2021.651317 (cit. on p. 7).

[32] P. Ježek. "2-Hydroxyglutarate in Cancer Cells". In: *Antioxidants & Redox Signaling* 33.13 (2020-11), pp. 903–926. ISSN: 1523-0864. DOI: 10.1089/ars.2019.7902 (cit. on p. 7).

[33] X. Chen et al. "Mutant p53 in cancer: from molecular mechanism to therapeutic modulation". In: *Cell Death & Disease* 13.11 (2022-11), p. 974. ISSN: 2041-4889. DOI: 10.1038/s41419-022-05408-1 (cit. on p. 7).

[34] Y. Pang et al. "The Chromatin Remodeler ATRX: Role and Mechanism in Biology and Cancer". In: *Cancers* 15.8 (2023-04), p. 2228. ISSN: 2072-6694. DOI: 10.3390/cancers15082228 (cit. on p. 7).

[35] A. J. Cesare and R. R. Reddel. "Alternative lengthening of telomeres: models, mechanisms and implications". In: *Nature Reviews Genetics* 11.5 (2010-05), pp. 319–330. ISSN: 1471-0056. DOI: 10.1038/nrg2763 (cit. on p. 7).

[36] S. Gritsch, T. T. Batchelor, and L. N. Gonzalez Castro. "Diagnostic, therapeutic, and prognostic implications of the 2021 World Health Organization classification of tumors of the central nervous system". In: *Cancer* 128.1 (2022-01), pp. 47–58. ISSN: 0008-543X. DOI: 10.1002/cncr.33918 (cit. on p. 7).

[37] A. Yuile et al. "CDKN2A/B Homozygous Deletions in Astrocytomas: A Literature Review". In: *Current Issues in Molecular Biology* 45.7 (2023-06), pp. 5276–5292. ISSN: 1467-3045. DOI: 10.3390/cimb45070335 (cit. on p. 7).

[38] J. Qi and Z. Ouyang. "Targeting CDK4/6 for Anticancer Therapy". In: *Biomedicines* 10.3 (2022-03), p. 685. ISSN: 2227-9059. DOI: 10.3390/biomedicines10030685 (cit. on p. 7).

[39] J. G. Cairncross et al. "Benefit From Procarbazine, Lomustine, and Vincristine in Oligodendroglial Tumors Is Associated With Mutation of IDH". In: *Journal of Clinical Oncology* 32.8 (2014-03), pp. 783–790. ISSN: 0732-183X. DOI: 10.1200/JCO.2013.49.3726 (cit. on p. 7).

[40] B. T. Whitfield and J. T. Huse. "Classification of adult-type diffuse gliomas: Impact of the World Health Organization 2021 update". In: *Brain Pathology* 32.4 (2022-07). ISSN: 1015-6305. DOI: 10.1111/bpa.13062 (cit. on pp. 7–9, 65).

[41] T. Hasanau et al. "Detection of TERT Promoter Mutations as a Prognostic Biomarker in Gliomas: Methodology, Prospects, and Advances". In: *Biomedicines* 10.3 (2022-03), p. 728. ISSN: 2227-9059. DOI: 10.3390/biomedicines10030728 (cit. on p. 8).

[42] N. K. Terzi, I. Yilmaz, and A. B. Oz. "The place and prognostic value of tert promoter mutation in molecular classification in grade ii-iii glial tumors and primary glioblastomas". In: *Turkish Journal of Pathology* (2021). ISSN: 1018-5615. DOI: 10.5146/tjpath.2021.01555 (cit. on p. 8).

[43] N. Olympios et al. "TERT Promoter Alterations in Glioblastoma: A Systematic Review". In: *Cancers* 13.5 (2021-03), p. 1147. ISSN: 2072-6694. DOI: 10.3390/cancers13051147 (cit. on p. 8).

[44] E. Eskilsson et al. "EGFR heterogeneity and implications for therapeutic intervention in glioblastoma". In: *Neuro-Oncology* 20.6 (2018-05), pp. 743–752. ISSN: 1522-8517. DOI: 10.1093/neuonc/nox191 (cit. on pp. 8, 65).

[45] S. M. B. Rodriguez et al. "An Overview of EGFR Mechanisms and Their Implications in Targeted Therapies for Glioblastoma". In: *International Journal of Molecular Sciences* 24.13 (2023-07), p. 11110. ISSN: 1422-0067. DOI: 10.3390/ijms241311110 (cit. on pp. 8, 9).

[46] X.-P. Li et al. "EGFR alterations in glioblastoma play a role in antitumor immunity regulation". In: *Frontiers in Oncology* 13 (2023-08). ISSN: 2234-943X. DOI: 10.3389/fonc.2023.1236246 (cit. on p. 9).

[47] M. Hashemi et al. "Progress in targeting PTEN/PI3K/Akt axis in glioblastoma therapy: Revisiting molecular interactions". In: *Biomedicine & Pharmacotherapy* 158 (2023-02), p. 114204. ISSN: 07533322. DOI: 10.1016/j.biopha.2022.114204 (cit. on pp. 9, 65, 69).

[48] H. Liu and T. Tang. "MAPK signaling pathway-based glioma subtypes, machine-learning risk model, and key hub proteins identification". In: *Scientific Reports* 13.1 (2023-11), p. 19055. ISSN: 2045-2322. DOI: 10.1038/s41598-023-45774-0 (cit. on p. 9).

[49] D. Stichel et al. "Distribution of EGFR amplification, combined chromosome 7 gain and chromosome 10 loss, and TERT promoter mutation in brain tumors and their potential for the reclassification of IDHwt astrocytoma to glioblastoma". In: *Acta Neuropathologica* 136.5 (2018-11), pp. 793–803. ISSN: 0001-6322. DOI: 10.1007/s00401-018-1905-0 (cit. on p. 9).

[50] T. Kanazawa et al. "Predictive markers for MGMT promoter methylation in glioblastomas". In: *Neurosurgical Review* 42.4 (2019-12), pp. 867–876. ISSN: 0344-5607. DOI: 10.1007/s10143-018-01061-5 (cit. on p. 9).

[51] R. Chai et al. "Predictive value of MGMT promoter methylation on the survival of TMZ treated IDH-mutant glioblastoma". In: *Cancer Biology and Medicine* 18.1 (2021), pp. 271–282. ISSN: 2095-3941. DOI: 10.20892/j.issn.2095-3941.2020.0179 (cit. on p. 9).

[52] K. Zappe et al. "Association between MGMT Enhancer Methylation and MGMT Promoter Methylation, MGMT Protein Expression, and Overall Survival in Glioblastoma". In: *Cells* 12.12 (2023-06), p. 1639. ISSN: 2073-4409. DOI: 10.3390/cells12121639 (cit. on p. 9).

[53] M. B. Lopes et al. "The Role of Network Science in Glioblastoma". In: *Cancers* 13.5 (2021-03), p. 1045. ISSN: 2072-6694. DOI: 10.3390/cancers13051045 (cit. on pp. 9, 10).

[54] J. Friedman, T. Hastie, and R. Tibshirani. "Sparse inverse covariance estimation with the graphical lasso". In: *Biostatistics* 9.3 (2008-07), pp. 432–441. ISSN: 1468-4357. DOI: 10.1093/biostatistics/kxm045 (cit. on pp. 9, 10, 19, 29).

[55] R. Coletti et al. "Inferring diagnostic and prognostic gene expression signatures across WHO glioma classifications: A network-based approach". In: (2023-05). DOI: https://doi.org/10.48550/arXiv.2305.12207 (cit. on pp. 9, 27).

[56] Y.-J. Huang, T.-P. Lu, and C. K. Hsiao. "Application of graphical lasso in estimating network structure in gene set". In: *Annals of Translational Medicine* 8.23 (2020-12), pp. 1556–1556. ISSN: 23055839. DOI: 10.21037/atm-20-6490 (cit. on pp. 9, 10).

[57] M. J. Ha, V. Baladandayuthapani, and K.-A. Do. "DINGO: differential network analysis in genomics". In: *Bioinformatics* 31.21 (2015-11), pp. 3413–3420. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btv406 (cit. on pp. 9, 10, 16, 29, 30).

[58] K. H. Shutta et al. "Gaussian graphical models with applications to omics analyses". In: *Statistics in Medicine* 41.25 (2022-11), pp. 5150–5187. ISSN: 0277-6715. DOI: 10.1002/sim.9546 (cit. on p. 10).

[59] C. Glymour, K. Zhang, and P. Spirtes. "Review of Causal Discovery Methods Based on Graphical Models". In: *Frontiers in Genetics* 10 (2019-06). ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00524 (cit. on pp. 11, 15, 16, 19–21, 23, 24, 26, 71).

[60] A. White and M. Vignes. "Causal Queries from Observational Data in Biological Systems via Bayesian Networks: An Empirical Study in Small Networks". In: 2019, pp. 111–142. DOI: 10.1007/978-1-4939-8882-2{_}5 (cit. on pp. 11, 14, 24).

[61] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. Ed. by MIT press. 2nd Edition. Vol. 81. New York, NY: Springer New York, 2000, pp. 116–124. ISBN: 978-1-4612-7650-0. DOI: 10.1007/978-1-4612-2748-9 (cit. on pp. 11, 13, 17, 18).

[62] A. R. Nogueira et al. "Methods and tools for causal discovery and causal inference". In: *WIREs Data Mining and Knowledge Discovery* 12.2 (2022-03). ISSN: 1942-4787. DOI: 10.1002/widm.1449 (cit. on pp. 11–13).

[63] J. Pearl. *Causality*. Second. Cambridge University Press, 2009-09, pp. 85–86. ISBN: 9780511803161. DOI: 10.1017/CBO9780511803161 (cit. on pp. 11–13, 15, 72).

[64] N. Friedman et al. "Using Bayesian Networks to Analyze Expression Data". In: *Journal of Computational Biology* 7.3-4 (2000-08), pp. 601–620. ISSN: 1066-5277. DOI: 10.1089/106652700750050961 (cit. on p. 11).

[65] P. Spirtes et al. "Constructing Bayesian network models of gene expression networks from microarray data". In: *Carnegie Mellon University* (2000) (cit. on p. 11).

[66] E. Liu, L. Li, and L. Cheng. "Gene Regulatory Network Review". In: *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, 2019, pp. 155–164. DOI: 10.1016/B978-0-12-809633-8.20218-5 (cit. on pp. 11, 15, 61, 71).

[67] X. Chen and J. Xuan. "Bayesian Inference of Gene Regulatory Network". In: *Bayesian Inference on Complicated Data*. IntechOpen, 2020-07. DOI: 10.5772/intechopen.88799 (cit. on p. 11).

[68] Q. Zhang, J. E. Burdette, and J.-P. Wang. "Integrative network analysis of TCGA data for ovarian cancer". In: *BMC Systems Biology* 8.1 (2014-12), p. 1338. ISSN: 1752-0509. DOI: 10.1186/s12918-014-0136-9 (cit. on p. 12).

[69] D. Osorio et al. "Population-level comparisons of gene regulatory networks modeled on high-throughput single-cell transcriptomics data". In: *Nature Computational Science* 4.3 (2024-03), pp. 237–250. ISSN: 2662-8457. DOI: 10.1038/s43588-024-00597-5 (cit. on p. 12).

[70] R. Ji et al. "Construction of pan-cancer regulatory networks based on causal inference". In: *BioSystems* 243 (2024-09), p. 105279. ISSN: 03032647. DOI: 10.1016/j.biosystems.2024.105279 (cit. on p. 12).

[71] P. Spirtes. "Introduction to causal inference". In: *Journal of Machine Learning Research* 11.1532-4435 (2010-01), pp. 1643–1662 (cit. on p. 13).

[72] D. Malinsky and D. Danks. "Causal discovery algorithms: A practical guide". In: *Philosophy Compass* 13.1 (2018-01). ISSN: 1747-9991. DOI: 10.1111/phc3.12470 (cit. on pp. 13–15).

[73] A. Zanga, E. Ozkirimli, and F. Stella. "A Survey on Causal Discovery: Theory and Practice". In: *International Journal of Approximate Reasoning* 151 (2022-12), pp. 101–129. ISSN: 0888613X. DOI: 10.1016/j.ijar.2022.09.004 (cit. on pp. 13–15, 18, 21, 35).

[74] P. Spirtes. "An Anytime Algorithm for Causal Inference". In: *International Conference on Artificial Intelligence and Statistics* (2001) (cit. on pp. 14, 19, 20).

[75] D. M. Chickering. "Learning equivalence classes of Bayesian-network structures". In: *Journal of Machine Learning Research* 2.1533-7928 (2002), pp. 445–498. ISSN: 0003-6951. DOI: 10.1162/153244302760200696 (cit. on pp. 14, 15, 20, 32).

[76] G. Schwarz. "Estimating the Dimension of a Model". In: *The Annals of Statistics* 6.2 (1978-03). ISSN: 0090-5364. DOI: 10.1214/aos/1176344136 (cit. on p. 14).

[77] M. Scutari, C. E. Graafland, and J. M. Gutiérrez. "Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms". In: *International Journal of Approximate Reasoning* 115 (2019-12), pp. 235–253. ISSN: 0888613X. DOI: 10.1016/j.ijar.2019.10.003 (cit. on p. 15).

[78] D. M. Chickering. "Optimal Structure Identification With Greedy Search". In: *Journal of Machine Learning Research* 3 (2003), pp. 507–554 (cit. on p. 15).

[79] B. Huang et al. "Generalized Score Functions for Causal Discovery". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: ACM, 2018-07, pp. 1551–1560. ISBN: 9781450355520. DOI: 10.1145/3219819.3220104 (cit. on pp. 15, 20, 21, 33, 62).

[80] A. Shojaie et al. "Inferring Regulatory Networks by Combining Perturbation Screens and Steady State Gene Expression Profiles". In: *PLoS ONE* 9.2 (2014-02), e82393. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0082393 (cit. on p. 15).

[81] N. R. Zabet. "Negative feedback and physical limits of genes". In: *Journal of Theoretical Biology* 284.1 (2011-09), pp. 82–91. ISSN: 00225193. DOI: 10.1016/j.jtbi.2011.06.021 (cit. on p. 15).

[82] L. Wang et al. "A survey of causal discovery based on functional causal model". In: *Engineering Applications of Artificial Intelligence* 133 (2024-07), p. 108258. ISSN: 09521976. DOI: 10.1016/j.engappai.2024.108258 (cit. on p. 15).

[83] B. Zhao et al. "Coresets for fast causal discovery with the additive noise model". In: *Pattern Recognition* 148 (2024-04), p. 110149. ISSN: 00313203. DOI: 10.1016/j.patcog.2023.110149 (cit. on p. 15).

[84] K. Zhang et al. "On Estimation of Functional Causal Models". In: *ACM Transactions on Intelligent Systems and Technology* 7.2 (2016-01), pp. 1–22. ISSN: 2157-6904. DOI: 10.1145/2700476 (cit. on p. 16).

[85] S. Shimizu et al. " A linear Non-Gaussian acyclic model for causal discovery". In: *Journal of Machine Learning Research* 7 (2006), pp. 2003–2030 (cit. on pp. 16, 21–23).

[86] P. O. Hoyer et al. "Nonlinear causal discovery with additive noise models". In: *In Proceedings of the 21st International Conference on Neural Information Processing Systems (NIPS'08)* (2008), pp. 689–696 (cit. on pp. 16, 25).

[87] A. Zhang Kun. Hyvärinen. "On the identifiability of the post-nonlinear causal model". In: *In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI '09)* (2009), pp. 647–655 (cit. on pp. 16, 25, 26).

[88] M. Frydenberg. "The chain graph Markov property". In: *Scandinavian Journal of Statistics* 17 (1990), pp. 333–353 (cit. on pp. 16, 17, 30, 61).

[89] C. A. Class et al. "iDINGO—integrative differential network analysis in genomics with Shiny application". In: *Bioinformatics* 34.7 (2018-04), pp. 1243–1245. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btx750 (cit. on pp. 16, 17, 30, 31, 71).

[90] M. Denis et al. "A Bayesian two-step integrative procedure incorporating prior knowledge for the identification of miRNA-mRNAs involved in hepatocellular carcinoma". In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022-07, pp. 81–86. ISBN: 978-1-7281-2782-8. DOI: 10.1109/EMBC48229.2022.9871330 (cit. on p. 16).

[91] M. Kalisch and P. Bühlmann. "Estimating high-dimensional directed acyclic graphs with the PC-algorithm". In: *Journal of Machine Learning Research* 8 (2007), pp. 613–636 (cit. on p. 17).

[92] M. Kalisch et al. "Causal Inference Using Graphical Models with the R Package pcalg". In: *Journal of Statistical Software* 47.11 (2012). ISSN: 1548-7660. DOI: 10.18637/jss.v047.i11 (cit. on pp. 17, 19, 21, 31, 32).

[93] D. Colombo and M. H. Maathuis. "Order-Independent Constraint-Based Causal Structure Learning". In: *Journal of Machine Learning Research* 15 (2014), pp. 3921–3962 (cit. on pp. 17–19, 31, 32).

[94] E. Giudice, J. Kuipers, and G. Moffa. "The dual PC algorithm and the role of Gaussianity for structure learning of Bayesian networks". In: *International Journal of Approximate Reasoning* 161 (2023-10), p. 108975. ISSN: 0888613X. DOI: 10.1016/j.ijar.2023.108975 (cit. on p. 18).

[95] J. Ramsey, J. Zhang, and P. Spirtes. "Adjacency-Faithfulness and Conservative Causal Inference". In: *Proceedings of the Twenty-Second Conference Conference on Uncertainty in Artificial Intelligence (2006)* (2006), pp. 401–408 (cit. on p. 18).

[96] J. D. Ramsey et al. "Tetrad—a toolbox for causal discovery". In: *8th international workshop on climate in-formatics* (2018) (cit. on p. 19).

[97] M. Scutari. "Learning Bayesian Networks with the bnlearn R Package". In: *Journal of Statistical Software* 35.3 (2010). ISSN: 1548-7660. DOI: 10.18637/jss.v035.i03 (cit. on p. 19).

[98] Y. Zheng et al. "Causal-learn: Causal discovery in python". In: *Journal of Machine Learning Research* 25 (2024), pp. 1–8 (cit. on pp. 19, 21, 31–34).

[99] X. Zhang et al. "Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information". In: *Bioinformatics* 28.1 (2012-01), pp. 98–104. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btr626 (cit. on p. 19).

[100] S. H. Mahmoodi, R. Aghdam, and C. Eslahchi. "An order independent algorithm for inferring gene regulatory network using quantile value for conditional independence tests". In: *Scientific Reports* 11.1 (2021-04), p. 7605. ISSN: 2045-2322. DOI: 10.1038/s41598-021-87074-5 (cit. on p. 19).

[101] T. Richardson and P. Spirtes. "Ancestral graph Markov models". In: *The Annals of Statistics* 30.4 (2002-08). ISSN: 0090-5364. DOI: 10.1214/aos/1031689015 (cit. on p. 19).

[102] J. Zhang. "Causal Reasoning with Ancestral Graphs". In: *J. Mach. Learn. Res* 9.1532-4435 (2008-06), pp. 1437–1474 (cit. on pp. 19, 32).

[103] D. Colombo et al. "Learning high-dimensional directed acyclic graphs with latent and selection variables". In: *The Annals of Statistics* 40.1 (2012-02). ISSN: 0090-5364. DOI: 10.1214/11-AOS940 (cit. on pp. 20, 32).

[104] R. Foraita et al. "Causal Discovery of Gene Regulation with Incomplete Data". In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 183.4 (2020-10), pp. 1747–1775. ISSN: 0964-1998. DOI: 10.1111/rssa.12565 (cit. on pp. 20, 35).

[105] J. Ramsey et al. "A million variables and more: the Fast Greedy Equivalence Search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images". In: *International Journal of Data Science and Analytics* 3.2 (2017-03), pp. 121–129. ISSN: 2364-415X. DOI: 10.1007/s41060-016-0032-z (cit. on pp. 20, 21).

[106] A. Hauser and P. Bühlmann. "Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs". In: *Journal of Machine Learning Research* 13.1532-4435 (2012), pp. 2409–2464 (cit. on pp. 20, 32).

[107] P. Nandy, A. Hauser, and M. H. Maathuis. "High-dimensional consistency in score-based and hybrid structure learning". In: *The Annals of Statistics* 46.6A (2018-12). ISSN: 0090-5364. DOI: 10.1214/17-AOS1654 (cit. on pp. 20, 21).

[108] N. Bernaola et al. "Learning massive interpretable gene regulatory networks of the human brain by merging Bayesian networks". In: *PLOS Computational Biology* 19.12 (2023-12), e1011443. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1011443 (cit. on p. 21).

[109] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, 2001-05. ISBN: 9780471405405. DOI: 10.1002/0471221317 (cit. on p. 21).

[110] S. Shimizu et al. "DirectLiNGAM: A Direct Method for learning a linear Non-Gaussian Structural equation model". In: *Journal of Machine Learning Research* 12 (2011), pp. 1225–1248 (cit. on pp. 21–24, 33).

[111] K. A. Bollen. *Structural Equations with Latent Variables*. Wiley, 1989-04. ISBN: 9780471011712. DOI: 10.1002/9781118619179 (cit. on p. 22).

[112] J. Himberg, A. Hyvärinen, and F. Esposito. "Validating the independent components of neuroimaging time series via clustering and visualization". In: *NeuroImage* 22.3 (2004-07), pp. 1214–1222. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2004.03.027 (cit. on p. 22).

[113] G. Darmois. "Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire". In: *Revue de l'Institut International de Statistique / Review of the International Statistical Institute* 21.1/2 (1953), p. 2. ISSN: 03731138. DOI: 10.2307/1401511 (cit. on p. 23).

[114] S. Shimizu. "Lingam: Non-Gaussian Methods for Estimating Causal Structures". In: *Behaviormetrika* 41.1 (2014-01), pp. 65–98. ISSN: 0385-7417. DOI: 10.2333/bhmk.41.65 (cit. on pp. 23, 24, 33, 34).

[115] Bach, Francis and Jordan, Michael. "Kernel Independent Component Analysis". In: *Journal of Machine Learning Research* 3 (2003), pp. 1–48 (cit. on p. 24).

[116] A. Hyvärinen and S. M. Smith. "Pairwise likelihood ratios for estimation of non-Gaussian structural equation models". In: *Journal of Machine Learning Research* 14.1 (2013), pp. 111–152 (cit. on pp. 24, 33).

[117] J. Kotoku et al. "Causal relations of health indices inferred statistically using the DirectLiNGAM algorithm from big data of Osaka prefecture health checkups". In: *PLOS ONE* 15.12 (2020-12), e0243229. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0243229 (cit. on p. 24).

[118] A. Gretton et al. "Kernel Methods for Measuring Independence". In: *Journal of Machine Learning Research* 6 (2005), pp. 2075–2129 (cit. on p. 25).

[119] J. Peters et al. "Causal discovery with continuous additive noise models". In: *J. Mach. Learn. Res* 15 (2013), pp. 2009–2053 (cit. on p. 25).

[120] P. Hu et al. "Application of Causal Inference to Genomic Analysis: Advances in Methodology". In: *Frontiers in Genetics* 9 (2018-07). ISSN: 1664-8021. DOI: 10.3389/fgene.2018.00238 (cit. on p. 26).

[121] J. Peters, D. Janzing, and B. Scholkopf. "Causal Inference on Discrete Data Using Additive Noise Models". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.12 (2011-12), pp. 2436–2450. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.71 (cit. on p. 26).

[122] "Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas". In: *New England Journal of Medicine* 372.26 (2015-06), pp. 2481–2498. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1402121 (cit. on p. 27).

[123] "Comprehensive genomic characterization defines human glioblastoma genes and core pathways". In: *Nature* 455.7216 (2008-10), pp. 1061–1068. ISSN: 0028-0836. DOI: 10.1038/nature07385 (cit. on p. 27).

[124] M. Mounir et al. "New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx". In: *PLOS Computational Biology* 15.3 (2019-03), e1006701. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006701 (cit. on p. 27).

[125] A. Colaprico et al. "TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data". In: *Nucleic Acids Research* 44.8 (2016-05), e71–e71. ISSN: 1362-4962. DOI: 10.1093/nar/gkv1507 (cit. on p. 27).

[126] T. Zhao et al. "The huge Package for High-dimensional Undirected Graph Estimation in R". In: *Journal of Machine Learning Research* 13 (2012), pp. 1059–1062 (cit. on pp. 27, 29, 31).

[127] C. M. Jarque and A. K. Bera. "Efficient tests for normality, homoscedasticity and serial independence of regression residuals". In: *Economics Letters* 6.3 (1980-01), pp. 255–259. ISSN: 01651765. DOI: 10.1016/0165-1765(80)90024-5 (cit. on p. 27).

[128] P. Shannon et al. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks". In: *Genome Research* 13.11 (2003-11), pp. 2498–2504. ISSN: 1088-9051. DOI: 10.1101/gr.1239303 (cit. on pp. 30, 31, 36).

[129] J. Chen and Z. Chen. "Extended Bayesian information criteria for model selection with large model spaces". In: *Biometrika* 95(3) (2008), pp. 759–771 (cit. on p. 30).

[130] R. Foygel and M. Drton. "Extended Bayesian information criteria for Gaussian graphical models". In: *Advances in Neural Information Processing Systems* 23 (2010), pp. 2020–2028 (cit. on pp. 30, 31).

[131] S. Epskamp and E. I. Fried. "A tutorial on regularized partial correlation networks". In: *Psychological Methods* 23(4) (2018), pp. 617–634 (cit. on p. 31).

[132] K. Zhang et al. "Kernel-based Conditional Independence Test and Application in Causal Discovery". In: *In 27th Conference on Uncertainty in Artificial Intelligence (UAI 2011)* (2011-07), pp. 804–813 (cit. on p. 32).

[133] Ikeuchi, T. and Ide, M. and Zeng, Y. and Maeda, T. N. and Shimizu, S. "Python package for causal discovery based on LiNGAM". In: *Journal of Machine Learning Research* 24(14) (2023), pp. 1–8 (cit. on p. 33).

[134] F. Pedregosa et al. "Scikit-learn: Machine learning in Python". In: *Journal of Machine Learning Research* 12 (2011-10), pp. 2825–2830 (cit. on p. 33).

[135] P. O. Hoyer et al. "Causal discovery of linear acyclic models with arbitrary distributions". In: *Uncertainty in Artificial Intelligence* (2008), pp. 282–289 (cit. on p. 34).

[136] N. Friedman, M. Goldszmidt, and A. Wyner. "Data analysis with bayesian networks: a bootstrap approach". In: *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* 1558606149 (1999), pp. 196–205 (cit. on p. 35).

[137] I. Pigeot et al. "The uncertainty of a selected graphical model". In: *Journal of Applied Statistics* 42.11 (2015-11), pp. 2335–2352. ISSN: 0266-4763. DOI: 10.1080/02664763.2015.1030368 (cit. on p. 35).

[138] I. Tsamardinos, L. E. Brown, and C. F. Aliferis. "The max-min hill-climbing Bayesian network structure learning algorithm". In: *Machine Learning* 65.1 (2006-10), pp. 31–78. ISSN: 0885-6125. DOI: 10.1007/s10994-006-6889-7 (cit. on p. 35).

[139] L. Cheng et al. "Evaluation Methods and Measures for Causal Learning Algorithms". In: *IEEE Transactions on Artificial Intelligence* 3.6 (2022-12), pp. 924–943. ISSN: 2691-4581. DOI: 10.1109/TAI.2022.3150264 (cit. on p. 35).

[140] H. M and S. M.N. "A Review on Evaluation Metrics for Data Classification Evaluations". In: *International Journal of Data Mining & Knowledge Management Process* 5.2 (2015-03), pp. 01–11. ISSN: 2231007X. DOI: 10.5121/ijdkp.2015.5201 (cit. on p. 36).

[141] M. Kanehisa. "KEGG: Kyoto Encyclopedia of Genes and Genomes". In: *Nucleic Acids Research* 28.1 (2000-01), pp. 27–30. ISSN: 13624962. DOI: 10.1093/nar/28.1.27 (cit. on pp. 36, 37).

[142] M. Ashburner et al. "Gene Ontology: tool for the unification of biology". In: *Nature Genetics* 25.1 (2000-05), pp. 25–29. ISSN: 1061-4036. DOI: 10.1038/75556 (cit. on pp. 36, 37).

[143] D. V. Klopfenstein et al. "GOATOOLS: A Python library for Gene Ontology analyses". In: *Scientific Reports* 8.1 (2018-07), p. 10872. ISSN: 2045-2322. DOI: 10.1038/s41598-018-28948-z (cit. on p. 37).

[144] B. T. Sherman et al. "DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)". In: *Nucleic Acids Research* 50.W1 (2022-07), W216–W221. ISSN: 0305-1048. DOI: 10.1093/nar/gkac194 (cit. on p. 37).

[145] S. Falcon and R. Gentleman. "Using GOstats to test gene lists for GO term association". In: *Bioinformatics* 23.2 (2007-01), pp. 257–258. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btl567 (cit. on p. 37).

[146] W. S. Noble. "How does multiple testing correction work?" In: *Nature Biotechnology* 27.12 (2009-12), pp. 1135–1137. ISSN: 1087-0156. DOI: 10.1038/nbt1209-1135 (cit. on p. 37).

[147] B. Phipson, J. Maksimovic, and A. Oshlack. "missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform". In: *Bioinformatics* 32.2 (2016-01), pp. 286–288. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btv560 (cit. on p. 37).

[148] D. Karolchik, A. S. Hinrichs, and W. J. Kent. "The UCSC Genome Browser". In: *Current Protocols in Bioinformatics* 28.1 (2009-12). ISSN: 1934-3396. DOI: 10.1002/0471250953.bi0104s28 (cit. on p. 37).

[149] D. Szklarczyk et al. "The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest". In: *Nucleic Acids Research* 51.D1 (2023-01), pp. D638–D646. ISSN: 0305-1048. DOI: 10.1093/nar/gkac1000 (cit. on p. 37).

[150] D. S. Wishart et al. "DrugBank: a knowledgebase for drugs, drug actions and drug targets". In: *Nucleic Acids Research* 36.suppl_1 (2008-01), pp. D901–D906. ISSN: 0305-1048. DOI: 10.1093/nar/gkm958 (cit. on p. 37).

[151] R. Q. Li et al. "CD74 as a prognostic and M1 macrophage infiltration marker in a comprehensive pan-cancer analysis". In: *Scientific Reports* 14.1 (2024-04), p. 8125. ISSN: 2045-2322. DOI: 10.1038/s41598-024-58899-7 (cit. on p. 63).

[152] E. Schaafsma et al. "Pan-cancer association of HLA gene expression with cancer prognosis and immunotherapy efficacy". In: *British Journal of Cancer* 125.3 (2021-08), pp. 422–432. ISSN: 0007-0920. DOI: 10.1038/s41416-021-01400-2 (cit. on p. 63).

[153] A. M. Johnson et al. "Cancer Cell–Intrinsic Expression of MHC Class II Regulates the Immune Microenvironment and Response to Anti–PD-1 Therapy in Lung Adenocarcinoma". In: *The Journal of Immunology* 204.8 (2020-04), pp. 2295–2307. ISSN: 0022-1767. DOI: 10.4049/jimmunol.1900778 (cit. on p. 63).

[154] E. Bonnin et al. "CD74 supports accumulation and function of regulatory T cells in tumors". In: *Nature Communications* 15.1 (2024-05), p. 3749. ISSN: 2041-1723. DOI: 10.1038/s41467-024-47981-3 (cit. on p. 63).

[155] Y.-F. Zeng et al. "The efficacy and safety of anti-PD-1/PD-L1 in treatment of glioma: a single-arm meta-analysis". In: *Frontiers in Immunology* 14 (2023-04). ISSN: 1664-3224. DOI: 10.3389/fimmu.2023.1168244 (cit. on p. 63).

[156] C. E. Hughes and R. J. B. Nibbs. "A guide to chemokines and their receptors". In: *The FEBS Journal* 285.16 (2018-08), pp. 2944–2971. ISSN: 1742-464X. DOI: 10.1111/febs.14466 (cit. on p. 63).

[157] A.-L. Amati et al. "Chemokines (CCL3, CCL4, and CCL5) Inhibit ATP-Induced Release of IL-1 β by Monocytic Cells". In: *Mediators of Inflammation* 2017 (2017), pp. 1–10. ISSN: 0962-9351. DOI: 10.1155/2017/1434872 (cit. on p. 63).

[158] F. Allen et al. "CCL3 Enhances Antitumor Immune Priming in the Lymph Node via IFNγ with Dependency on Natural Killer Cells". In: *Frontiers in Immunology* 8 (2017-10). ISSN: 1664-3224. DOI: 10.3389/fimmu.2017.01390 (cit. on p. 63).

[159] N. Mukaida, S.-i. Sasaki, and T. Baba. "CCL4 Signaling in the Tumor Microenvironment". In: *Tumor Microenvironment The Role of Chemokines – Part A*. Ed. by Alexander Birbrair. Vol. 1231. Springer New York, 2020. Chap. 3, pp. 23–32. ISBN: 978-3-030-36667-4. DOI: 10.1007/978-3-030-36667-4{_}3 (cit. on p. 63).

[160] I. Ntanasis-Stathopoulos, D. Fotiou, and E. Terpos. "CCL3 Signaling in the Tumor Microenvironment". In: *Tumor Microenvironment The Role of Chemokines – Part A*. Ed. by Alexander Birbrair. Vol. 1231. Springer New York, 2020. Chap. 2, pp. 13–21. ISBN: 978-3-030-36667-4. DOI: 10.1007/978-3-030-36667-4{_}2 (cit. on p. 63).

[161] C. Lin et al. "A Calcium-Related Immune Signature in Prognosis Prediction of Patients With Glioma". In: *Frontiers in Cell and Developmental Biology* 9 (2021-09). ISSN: 2296-634X. DOI: 10.3389/fcell.2021.723103 (cit. on p. 63).

[162] M. Singh, A. Raghav, and K. A. Gautam. "Role of the circulatory interleukin-6 in the pathogenesis of gliomas: A systematic review". In: *World Journal of Methodology* 12.5 (2022-09), pp. 428–437. ISSN: 2222-0682. DOI: 10.5662/wjm.v12.i5.428 (cit. on pp. 64, 69).

[163] Q. Zhang, N. Tuerxun, and S. Tuerxun. "IL-6 is associated with poor seizure control in low-grade glioma patients undergoing primary resection". In: *iScience* 27.7 (2024-07), p. 110267. ISSN: 25890042. DOI: 10.1016/j.isci.2024.110267 (cit. on pp. 64, 69).

[164] Y. Yang et al. "Four specific biomarkers associated with the progression of glioblastoma multiforme in older adults identified using weighted gene co-expression network analysis". In: *Bioengineered* 12.1 (2021-01), pp. 6643–6654. ISSN: 2165-5979. DOI: 10.1080/21655979.2021.1975980 (cit. on p. 64).

[165] K. Kesh et al. "Therapy Resistance, Cancer Stem Cells and ECM in Cancer: The Matrix Reloaded". In: *Cancers* 12.10 (2020-10), p. 3067. ISSN: 2072-6694. DOI: 10.3390/cancers12103067 (cit. on pp. 65, 66, 68).

[166] H. Zhu et al. "Targeting the Complement Pathway in Malignant Glioma Microenvironments". In: *Frontiers in Cell and Developmental Biology* 9 (2021-04). ISSN: 2296-634X. DOI: 10.3389/fcell.2021.657472 (cit. on p. 65).

[167] S. S. Bohlson and A. J. Tenner. "Complement in the Brain: Contributions to Neuroprotection, Neuronal Plasticity, and Neuroinflammation". In: *Annual Review of Immunology* 41.1 (2023-04), pp. 431–452. ISSN: 0732-0582. DOI: 10.1146/annurev-immunol-101921-035639 (cit. on p. 65).

[168] E. Tantillo et al. "Differential roles of pyramidal and fast-spiking, GABAergic neurons in the control of glioma cell proliferation". In: *Neurobiology of Disease* 141 (2020-07), p. 104942. ISSN: 09699961. DOI: 10.1016/j.nbd.2020.104942 (cit. on p. 66).

[169] R. Badalotti et al. "Gene Expression of GABAA Receptor Subunits and Association with Patient Survival in Glioma". In: *Brain Sciences* 14.3 (2024-03), p. 275. ISSN: 2076-3425. DOI: 10.3390/brainsci14030275 (cit. on p. 66).

[170] S. Löding et al. "Blood based metabolic markers of glioma from pre-diagnosis to surgery". In: *Scientific Reports* 14.1 (2024-09), p. 20680. ISSN: 2045-2322. DOI: 10.1038/s41598-024-71375-6 (cit. on p. 66).

[171] B. Björkblom et al. "Distinct metabolic hallmarks of WHO classified adult glioma subtypes". In: *Neuro-Oncology* 24.9 (2022-09), pp. 1454–1468. ISSN: 1522-8517. DOI: 10.1093/neuonc/noac042 (cit. on p. 66).

[172] V. Garcia, Y. R. Perera, and W. J. Chazin. "A Structural Perspective on Calprotectin as a Ligand of Receptors Mediating Inflammation and Potential Drug Target". In: *Biomolecules* 12.4 (2022-03), p. 519. ISSN: 2218-273X. DOI: 10.3390/biom12040519 (cit. on p. 67).

[173] P. Khosravi et al. "Calprotectin as new potential clinical marker for multiple myeloma". In: *PLOS ONE* 18.3 (2023-03), e0282841. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0282841 (cit. on p. 67).

[174] L. da Cruz et al. "TLR4 expression and functionality are downregulated in glioblastoma cells and in tumor-associated macrophages: A new mechanism of immune evasion?" In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1867.8 (2021-08), p. 166155. ISSN: 09254439. DOI: 10.1016/j.bbadis.2021.166155 (cit. on p. 67).

[175] S. Li et al. "Pseudogenes of annexin A2, novel prognosis biomarkers for diffuse gliomas". In: *Oncotarget* 8.63 (2017-12), pp. 106962–106975. ISSN: 1949-2553. DOI: 10.18632/oncotarget.22197 (cit. on p. 67).

[176] H. Ni et al. "Pseudogene ANXA2P2 knockdown shows tumor-suppressive function by inhibition of the PI3K/PKB pathway in glioblastoma cells". In: *Journal of Biochemical and Molecular Toxicology* 35.8 (2021-08). ISSN: 1095-6670. DOI: 10.1002/jbt.22824 (cit. on p. 68).

[177] H. Wang et al. "COL4A1 as a novel oncogene associated with the clinical characteristics of malignancy predicts poor prognosis in glioma". In: *Experimental and Therapeutic Medicine* 22.5 (2021-08), p. 1224. ISSN: 1792-0981. DOI: 10.3892/etm.2021.10658 (cit. on p. 69).

[178] J. Zhao and J. Jin. "Neutrophil extracellular traps: New players in cancer research". In: *Frontiers in Immunology* 13 (2022-08). ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.937565 (cit. on p. 69).

[179] C. Zha et al. "Neutrophil extracellular traps mediate the crosstalk between glioma progression and the tumor microenvironment via the HMGB1/RAGE/IL-8 axis". In: *Cancer Biology and Medicine* 17.1 (2020), pp. 154–168. ISSN: 2095-3941. DOI: 10.20892/j.issn.2095-3941.2019.0353 (cit. on p. 69).

[180] K. Lagemann et al. "Deep learning of causal structures in high dimensions under data limitations". In: *Nature Machine Intelligence* 5.11 (2023-10), pp. 1306–1316. ISSN: 2522-5839. DOI: 10.1038/s42256-023-00744-z (cit. on p. 71).

[181] P. Bühlmann, J. Peters, and J. Ernest. "CAM: Causal additive models, high-dimensional order search and penalized regression". In: *The Annals of Statistics* 42.6 (2014-12). ISSN: 0090-5364. DOI: 10.1214/14-AOS1260 (cit. on p. 71).

[182] J. M. Mooij and T. Claassen. "Constraint-Based Causal Discovery using Partial Ancestral Graphs in the presence of Cycles". In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)* 124 (2020-08), pp. 1159–1168 (cit. on p. 72).

A

Graphical Lasso results

Table A.1: Table containing the selected features retrieved from Graphical Lasso for each data type in the study with the respective ρ parameter

Data Type	ρ	**Colnames**
astro RNA	0.955	ANXA2P2
astro RNA	0.955	ANXA2
astro RNA	0.955	ASPM
astro RNA	0.955	AURKB
astro RNA	0.955	BIRC5
astro RNA	0.955	BOK
astro RNA	0.955	BUB1B
astro RNA	0.955	C1QA
astro RNA	0.955	C1QB
astro RNA	0.955	C1QC
astro RNA	0.955	CACNG3
astro RNA	0.955	CAMK2A
astro RNA	0.955	CARNS1
astro RNA	0.955	CCL3
astro RNA	0.955	CCL4L2
astro RNA	0.955	CCL4
astro RNA	0.955	CCNB2
astro RNA	0.955	CD74
astro RNA	0.955	CDK18
astro RNA	0.955	CDK1
astro RNA	0.955	CELF4
astro RNA	0.955	CENPF
astro RNA	0.955	CKAP2L
astro RNA	0.955	CLDN11

Continued on next page

Data Type	ρ	Colnames
astro RNA	0.955	CNTN2
astro RNA	0.955	COL4A1
astro RNA	0.955	COL4A2
astro RNA	0.955	CYBA
astro RNA	0.955	DLGAP5
astro RNA	0.955	DOCK2
astro RNA	0.955	ENPP2
astro RNA	0.955	EPR1
astro RNA	0.955	ERMN
astro RNA	0.955	FAM64A
astro RNA	0.955	FCGR1A
astro RNA	0.955	FCGR1B
astro RNA	0.955	FCGR1C
astro RNA	0.955	FERMT3
astro RNA	0.955	FOLH1
astro RNA	0.955	GABRA1
astro RNA	0.955	GABRA5
astro RNA	0.955	GABRB2
astro RNA	0.955	GAD2
astro RNA	0.955	GJB1
astro RNA	0.955	GJC2
astro RNA	0.955	GPR62
astro RNA	0.955	GPSM3
astro RNA	0.955	GTSE1
astro RNA	0.955	HAVCR2
astro RNA	0.955	HBA2
astro RNA	0.955	HBB
astro RNA	0.955	HJURP
astro RNA	0.955	HLA-DOA
astro RNA	0.955	HLA-DPA1
astro RNA	0.955	HLA-DPB1
astro RNA	0.955	HLA-DRA
astro RNA	0.955	KCNC2
astro RNA	0.955	KIAA1045
astro RNA	0.955	KIF20A
astro RNA	0.955	KIF23
astro RNA	0.955	KIFC1
astro RNA	0.955	KLK6

Continued on next page

Data Type	ρ	Colnames
astro RNA	0.955	LAPTM5
astro RNA	0.955	LOC283999
astro RNA	0.955	MAG
astro RNA	0.955	MAL2
astro RNA	0.955	MAP7D2
astro RNA	0.955	MBP
astro RNA	0.955	MELK
astro RNA	0.955	MKI67
astro RNA	0.955	MOBP
astro RNA	0.955	MOG
astro RNA	0.955	NCAPG
astro RNA	0.955	NCAPH
astro RNA	0.955	NCKAP1L
astro RNA	0.955	NEFL
astro RNA	0.955	NEUROD2
astro RNA	0.955	NRGN
astro RNA	0.955	NSUN5P1
astro RNA	0.955	NSUN5P2
astro RNA	0.955	NUSAP1
astro RNA	0.955	PACSIN1
astro RNA	0.955	PBK
astro RNA	0.955	S1PR5
astro RNA	0.955	SH3TC2
astro RNA	0.955	SLC17A7
astro RNA	0.955	SLC32A1
astro RNA	0.955	SLC45A3
astro RNA	0.955	SLC5A11
astro RNA	0.955	SNAP25
astro RNA	0.955	SNCB
astro RNA	0.955	SPI1
astro RNA	0.955	ST18
astro RNA	0.955	SULT4A1
astro RNA	0.955	SYK
astro RNA	0.955	SYN1
astro RNA	0.955	TBR1
astro RNA	0.955	TMEM125
astro RNA	0.955	TMEM144
astro RNA	0.955	TOP2A

Continued on next page

Data Type	ρ	Colnames
astro RNA	0.955	TPX2
astro RNA	0.955	TTK
astro RNA	0.955	TYROBP
astro RNA	0.955	UBE2C
astro RNA	0.955	VAMP8
astro RNA	0.955	VSNL1
astro RNA	0.955	WDFY4
astro cpg	0.860	cg00017271
astro cpg	0.860	cg00048832
astro cpg	0.860	cg00051179
astro cpg	0.860	cg00386007
astro cpg	0.860	cg00462525
astro cpg	0.860	cg00765737
astro cpg	0.860	cg01132696
astro cpg	0.860	cg02534163
astro cpg	0.860	cg02589899
astro cpg	0.860	cg03537215
astro cpg	0.860	cg03732545
astro cpg	0.860	cg04452959
astro cpg	0.860	cg04771838
astro cpg	0.860	cg04879832
astro cpg	0.860	cg05022688
astro cpg	0.860	cg05096756
astro cpg	0.860	cg06023994
astro cpg	0.860	cg06954660
astro cpg	0.860	cg07366188
astro cpg	0.860	cg07675031
astro cpg	0.860	cg08843517
astro cpg	0.860	cg09025501
astro cpg	0.860	cg10070788
astro cpg	0.860	cg10792225
astro cpg	0.860	cg11145160
astro cpg	0.860	cg11598403
astro cpg	0.860	cg12741994
astro cpg	0.860	cg13442966
astro cpg	0.860	cg13572782
astro cpg	0.860	cg13785221
astro cpg	0.860	cg13879776

Continued on next page

Data Type	ρ	Colnames
astro cpg	0.860	cg14490686
astro cpg	0.860	cg15391531
astro cpg	0.860	cg15471953
astro cpg	0.860	cg15982099
astro cpg	0.860	cg16472060
astro cpg	0.860	cg16872841
astro cpg	0.860	cg17199181
astro cpg	0.860	cg17266581
astro cpg	0.860	cg17403520
astro cpg	0.860	cg17606183
astro cpg	0.860	cg17860158
astro cpg	0.860	cg19272238
astro cpg	0.860	cg19384775
astro cpg	0.860	cg19843036
astro cpg	0.860	cg19917714
astro cpg	0.860	cg19990651
astro cpg	0.860	cg20794824
astro cpg	0.860	cg20801056
astro cpg	0.860	cg20960405
astro cpg	0.860	cg21322241
astro cpg	0.860	cg21925310
astro cpg	0.860	cg21949512
astro cpg	0.860	cg23732629
astro cpg	0.860	cg23761616
astro cpg	0.860	cg24476363
astro cpg	0.860	cg25228510
astro cpg	0.860	cg25887236
astro cpg	0.860	cg27180880
astro cpg	0.860	cg27240008
GBM RNA	0.955	ANXA2P2
GBM RNA	0.955	ANXA2
GBM RNA	0.955	ASPM
GBM RNA	0.955	BUB1
GBM RNA	0.955	C1QA
GBM RNA	0.955	C1QB
GBM RNA	0.955	C1QC
GBM RNA	0.955	CARNS1
GBM RNA	0.955	CCNA2

Continued on next page

Data Type	ρ	Colnames
GBM RNA	0.955	CKAP2L
GBM RNA	0.955	CNDP1
GBM RNA	0.955	COL4A1
GBM RNA	0.955	COL4A2
GBM RNA	0.955	ERMN
GBM RNA	0.955	FAM115C
GBM RNA	0.955	GABRA1
GBM RNA	0.955	GABRG2
GBM RNA	0.955	KIF14
GBM RNA	0.955	KLK6
GBM RNA	0.955	LOC154761
GBM RNA	0.955	MAG
GBM RNA	0.955	MBP
GBM RNA	0.955	MOBP
GBM RNA	0.955	MOG
GBM RNA	0.955	NKX6-2
GBM RNA	0.955	PLP1
GBM RNA	0.955	S100A8
GBM RNA	0.955	S100A9
GBM RNA	0.955	TMEM125
GBM RNA	0.955	TMEM176A
GBM RNA	0.955	TMEM176B
GBM RNA	0.955	TOP2A
GBM cpg	0.850	cg00706570
GBM cpg	0.850	cg00765737
GBM cpg	0.850	cg01303723
GBM cpg	0.850	cg01644592
GBM cpg	0.850	cg02072495
GBM cpg	0.850	cg02306630
GBM cpg	0.850	cg02658690
GBM cpg	0.850	cg02988755
GBM cpg	0.850	cg03147503
GBM cpg	0.850	cg03148084
GBM cpg	0.850	cg03296248
GBM cpg	0.850	cg04771838
GBM cpg	0.850	cg05017199
GBM cpg	0.850	cg05081953
GBM cpg	0.850	cg05851240

Continued on next page

Data Type	ρ	**Colnames**
GBM cpg	0.850	cg06548292
GBM cpg	0.850	cg06648277
GBM cpg	0.850	cg07251711
GBM cpg	0.850	cg08382774
GBM cpg	0.850	cg08441806
GBM cpg	0.850	cg08959039
GBM cpg	0.850	cg10171448
GBM cpg	0.850	cg10908116
GBM cpg	0.850	cg11068946
GBM cpg	0.850	cg11096515
GBM cpg	0.850	cg11342452
GBM cpg	0.850	cg11598403
GBM cpg	0.850	cg12712747
GBM cpg	0.850	cg12778065
GBM cpg	0.850	cg12811033
GBM cpg	0.850	cg13141061
GBM cpg	0.850	cg13442966
GBM cpg	0.850	cg13572782
GBM cpg	0.850	cg13852093
GBM cpg	0.850	cg14397690
GBM cpg	0.850	cg14991769
GBM cpg	0.850	cg15304699
GBM cpg	0.850	cg15379757
GBM cpg	0.850	cg15391531
GBM cpg	0.850	cg15441535
GBM cpg	0.850	cg15495463
GBM cpg	0.850	cg16872841
GBM cpg	0.850	cg16891318
GBM cpg	0.850	cg17061340
GBM cpg	0.850	cg17266581
GBM cpg	0.850	cg19005390
GBM cpg	0.850	cg19701540
GBM cpg	0.850	cg20801056
GBM cpg	0.850	cg20818806
GBM cpg	0.850	cg20917241
GBM cpg	0.850	cg21322241
GBM cpg	0.850	cg21563471
GBM cpg	0.850	cg21655444

Continued on next page

Data Type	ρ	Colnames
GBM cpg	0.850	cg22232704
GBM cpg	0.850	cg22382805
GBM cpg	0.850	cg23008352
GBM cpg	0.850	cg23125970
GBM cpg	0.850	cg25682171
GBM cpg	0.850	cg25887236
GBM cpg	0.850	cg26385222
GBM cpg	0.850	cg27180880
GBM cpg	0.850	cg27240008
GBM cpg	0.850	cg27583690
oligo RNA	0.955	AK5
oligo RNA	0.955	AURKB
oligo RNA	0.955	BIRC5
oligo RNA	0.955	BUB1B
oligo RNA	0.955	BUB1
oligo RNA	0.955	C1QB
oligo RNA	0.955	C1QC
oligo RNA	0.955	CAMK2A
oligo RNA	0.955	CARNS1
oligo RNA	0.955	CDK1
oligo RNA	0.955	CENPF
oligo RNA	0.955	CKAP2L
oligo RNA	0.955	CLCA4
oligo RNA	0.955	CNDP1
oligo RNA	0.955	CNTN2
oligo RNA	0.955	CREG2
oligo RNA	0.955	CRYM
oligo RNA	0.955	CYB5R2
oligo RNA	0.955	DOCK2
oligo RNA	0.955	DTL
oligo RNA	0.955	ELOVL1
oligo RNA	0.955	EMX1
oligo RNA	0.955	ENPP2
oligo RNA	0.955	ERMN
oligo RNA	0.955	FAM64A
oligo RNA	0.955	FCGR1A
oligo RNA	0.955	FCGR1B
oligo RNA	0.955	FCGR1C

Continued on next page

Data Type	ρ	Colnames
oligo RNA	0.955	FOLH1
oligo RNA	0.955	FOXM1
oligo RNA	0.955	GABRA1
oligo RNA	0.955	GABRA5
oligo RNA	0.955	GABRB2
oligo RNA	0.955	GDA
oligo RNA	0.955	GJB1
oligo RNA	0.955	GJC2
oligo RNA	0.955	GNG3
oligo RNA	0.955	GPR62
oligo RNA	0.955	GTSE1
oligo RNA	0.955	HJURP
oligo RNA	0.955	HOOK1
oligo RNA	0.955	KCNK9
oligo RNA	0.955	KIF20A
oligo RNA	0.955	KIFC1
oligo RNA	0.955	KLK6
oligo RNA	0.955	LDB3
oligo RNA	0.955	LOC283999
oligo RNA	0.955	LOC285780
oligo RNA	0.955	LOC286002
oligo RNA	0.955	LRP2
oligo RNA	0.955	MAG
oligo RNA	0.955	MAL
oligo RNA	0.955	MAP7D2
oligo RNA	0.955	MBP
oligo RNA	0.955	MCM10
oligo RNA	0.955	MELK
oligo RNA	0.955	MKI67
oligo RNA	0.955	MOBP
oligo RNA	0.955	MOG
oligo RNA	0.955	NCAPG
oligo RNA	0.955	NCKAP1L
oligo RNA	0.955	NECAB1
oligo RNA	0.955	NEFL
oligo RNA	0.955	NEFM
oligo RNA	0.955	NEUROD2
oligo RNA	0.955	NINJ2

Continued on next page

Data Type	ρ	Colnames
oligo RNA	0.955	NIPAL4
oligo RNA	0.955	NKAIN2
oligo RNA	0.955	NKX6-2
oligo RNA	0.955	NRGN
oligo RNA	0.955	NUSAP1
oligo RNA	0.955	PACSIN1
oligo RNA	0.955	PBK
oligo RNA	0.955	PPP1R14A
oligo RNA	0.955	RC3H2
oligo RNA	0.955	RGS4
oligo RNA	0.955	RRM2
oligo RNA	0.955	S1PR5
oligo RNA	0.955	SAMD8
oligo RNA	0.955	SERPINI1
oligo RNA	0.955	SLC17A7
oligo RNA	0.955	SLC31A2
oligo RNA	0.955	SLC45A3
oligo RNA	0.955	SLC5A11
oligo RNA	0.955	SLC6A7
oligo RNA	0.955	SPOCK3
oligo RNA	0.955	ST18
oligo RNA	0.955	STX1A
oligo RNA	0.955	SV2B
oligo RNA	0.955	SYN2
oligo RNA	0.955	TAOK1
oligo RNA	0.955	TBR1
oligo RNA	0.955	TMEM125
oligo RNA	0.955	TMEM144
oligo RNA	0.955	TMEM176A
oligo RNA	0.955	TMEM176B
oligo RNA	0.955	TOP2A
oligo RNA	0.955	TPX2
oligo RNA	0.955	TRIM59
oligo RNA	0.955	TROAP
oligo RNA	0.955	TTK
oligo RNA	0.955	UBE2C
oligo RNA	0.955	UBXN7
oligo RNA	0.955	VSNL1

Continued on next page

Data Type	ρ	Colnames
oligo cpg	0.900	cg00059015
oligo cpg	0.900	cg00754896
oligo cpg	0.900	cg00962781
oligo cpg	0.900	cg01657744
oligo cpg	0.900	cg02127209
oligo cpg	0.900	cg02202193
oligo cpg	0.900	cg02571816
oligo cpg	0.900	cg02589899
oligo cpg	0.900	cg02700894
oligo cpg	0.900	cg03148084
oligo cpg	0.900	cg03303359
oligo cpg	0.900	cg03537215
oligo cpg	0.900	cg03593189
oligo cpg	0.900	cg03732545
oligo cpg	0.900	cg04118306
oligo cpg	0.900	cg04920452
oligo cpg	0.900	cg05017199
oligo cpg	0.900	cg05142617
oligo cpg	0.900	cg05301866
oligo cpg	0.900	cg05851240
oligo cpg	0.900	cg06570930
oligo cpg	0.900	cg08052629
oligo cpg	0.900	cg08326075
oligo cpg	0.900	cg09515953
oligo cpg	0.900	cg11598403
oligo cpg	0.900	cg11902728
oligo cpg	0.900	cg13442966
oligo cpg	0.900	cg13564825
oligo cpg	0.900	cg13572782
oligo cpg	0.900	cg14991769
oligo cpg	0.900	cg15391531
oligo cpg	0.900	cg15873301
oligo cpg	0.900	cg16409049
oligo cpg	0.900	cg16459364
oligo cpg	0.900	cg17078116
oligo cpg	0.900	cg17199181
oligo cpg	0.900	cg17266581
oligo cpg	0.900	cg17860158

Continued on next page

Data Type	ρ	Colnames
oligo cpg	0.900	cg18366748
oligo cpg	0.900	cg18898125
oligo cpg	0.900	cg19384775
oligo cpg	0.900	cg19694850
oligo cpg	0.900	cg19843036
oligo cpg	0.900	cg19917714
oligo cpg	0.900	cg20232102
oligo cpg	0.900	cg20585869
oligo cpg	0.900	cg20801056
oligo cpg	0.900	cg21177558
oligo cpg	0.900	cg21245652
oligo cpg	0.900	cg21322241
oligo cpg	0.900	cg22213386
oligo cpg	0.900	cg22266001
oligo cpg	0.900	cg22557662
oligo cpg	0.900	cg22839075
oligo cpg	0.900	cg23241335
oligo cpg	0.900	cg23483765
oligo cpg	0.900	cg23761616
oligo cpg	0.900	cg23975646
oligo cpg	0.900	cg25604883
oligo cpg	0.900	cg25649038
oligo cpg	0.900	cg25742540
oligo cpg	0.900	cg25887236
oligo cpg	0.900	cg26327804
oligo cpg	0.900	cg26330518
oligo cpg	0.900	cg26385222
oligo cpg	0.900	cg27180880
oligo cpg	0.900	cg27240008

B

iDINGO results

B.1 Astrocytoma vs Oligodendroglioma

Table B.1: Astrocytoma iDINGO summary denoting the differential edges with the respective p-value, differential score and edge type.

Source node	Target node	p-value	differential score	interaction type
cg04771838	ANXA2P2	0.0015	1.94	directed
cg07366188	ANXA2P2	0.0082	-1.61	directed
cg08843517	ANXA2P2	3.5e-05	-2.53	directed
cg10792225	ANXA2P2	0.0007	-2.07	directed
cg27180880	ANXA2P2	0.0052	-1.71	directed
cg04771838	ANXA2	0.0017	-1.92	directed
cg07366188	ANXA2	0.0084	1.62	directed
cg08843517	ANXA2	3.5e-05	2.54	directed
cg10792225	ANXA2	0.0004	2.17	directed
cg27180880	ANXA2	0.0059	1.69	directed
cg04771838	AURKB	0.0071	-1.64	directed
cg07366188	AURKB	0.0096	1.59	directed
cg08843517	AURKB	0.00014	2.34	directed
cg27180880	AURKB	0.005	1.72	directed
cg10792225	BOK	0.0091	1.60	directed
cg08843517	CCL3	0.0028	1.83	directed
cg27180880	CCL3	0.0095	1.59	directed
cg08843517	CCL4	0.0018	-1.90	directed
cg08843517	CD74	0.0018	1.92	directed
cg10792225	CD74	0.0068	1.66	directed
cg04771838	CKAP2L	6.3e-05	2.45	directed
cg07366188	CKAP2L	0.009	-1.59	directed
cg08843517	CKAP2L	1.4e-05	-2.65	directed

cg10792225	CKAP2L	0.00064	-2.08	directed
cg16872841	CKAP2L	0.0072	-1.64	directed
cg27180880	CKAP2L	0.0024	-1.85	directed
cg04771838	COL4A1	0.00066	2.09	directed
cg07366188	COL4A1	0.0063	-1.67	directed
cg08843517	COL4A1	0.00033	-2.19	directed
cg10792225	COL4A1	0.0069	-1.65	directed
cg27180880	COL4A1	0.0037	-1.77	directed
cg04771838	COL4A2	0.00068	-2.07	directed
cg07366188	COL4A2	0.0085	1.61	directed
cg08843517	COL4A2	0.00025	2.25	directed
cg27180880	COL4A2	0.0034	1.80	directed
cg04771838	CYBA	0.00037	-2.17	directed
cg07366188	CYBA	0.0088	1.61	directed
cg08843517	CYBA	0.00021	2.27	directed
cg10792225	CYBA	0.0067	1.66	directed
cg27180880	CYBA	0.0032	1.81	directed
cg08843517	DOCK2	0.0077	-1.63	directed
cg04771838	ERMN	0.0077	1.64	directed
cg08843517	ERMN	0.00069	-2.07	directed
cg27180880	ERMN	0.009	-1.59	directed
cg04771838	GABRB2	0.0079	1.63	directed
cg08843517	GABRB2	0.0019	-1.89	directed
cg10792225	GABRB2	0.0074	-1.63	directed
cg27180880	GABRB2	0.0029	-1.82	directed
cg04771838	GJC2	0.00046	-2.14	directed
cg10792225	GJC2	0.006	1.69	directed
cg27180880	GJC2	0.0036	1.78	directed
cg00048832	HBA2	0.0071	-1.64	directed
cg04771838	HBA2	0.00061	-2.09	directed
cg07366188	HBA2	0.0043	1.75	directed
cg08843517	HBA2	9.6e-05	2.39	directed
cg10792225	HBA2	0.0013	1.97	directed
cg27180880	HBA2	0.0044	1.75	directed
cg04771838	HBB	0.00082	2.05	directed
cg07366188	HBB	0.004	-1.76	directed
cg08843517	HBB	0.00011	-2.36	directed
cg10792225	HBB	0.0017	-1.92	directed
cg27180880	HBB	0.005	-1.71	directed
cg04771838	HLA-DOA	0.0021	1.89	directed

cg07366188	HLA-DOA	0.0091	-1.59	directed
cg08843517	HLA-DOA	0.00022	-2.26	directed
cg10792225	HLA-DOA	0.00085	-2.04	directed
cg27180880	HLA-DOA	0.0053	-1.70	directed
cg04771838	HLA-DPB1	0.00051	-2.12	directed
cg08843517	HLA-DPB1	0.0024	1.86	directed
cg10792225	HLA-DPB1	0.0024	1.86	directed
cg04771838	HLA-DRA	0.009	1.60	directed
cg08843517	HLA-DRA	0.001	-2.01	directed
cg08843517	MOG	0.0062	-1.67	directed
cg04771838	NEUROD2	0.00067	-2.08	directed
cg08843517	NEUROD2	0.00011	2.36	directed
cg10792225	NEUROD2	0.0044	1.75	directed
cg27180880	NEUROD2	0.0074	1.64	directed
cg04771838	NRGN	0.00052	-2.12	directed
cg08843517	NRGN	0.00069	2.08	directed
cg10792225	NRGN	0.00086	2.04	directed
cg27180880	NRGN	0.0045	1.74	directed
cg04771838	SNAP25	0.006	1.69	directed
cg08843517	SNAP25	0.001	-2.01	directed
cg27180880	SNAP25	0.0043	-1.74	directed
cg04771838	SNCB	0.0064	-1.66	directed
cg08843517	SNCB	0.0016	1.93	directed
cg10792225	SNCB	0.0064	1.67	directed
cg08843517	SPI1	0.0074	1.64	directed
cg04771838	TBR1	0.0049	1.72	directed
cg08843517	TBR1	0.0029	-1.81	directed
cg10792225	TBR1	0.0029	-1.82	directed
cg08843517	TMEM144	0.0045	-1.74	directed
cg10792225	TMEM144	0.0066	-1.66	directed
cg04771838	UBE2C	0.0029	-1.82	directed
cg08843517	UBE2C	0.00099	2.02	directed
cg10792225	UBE2C	0.00033	2.20	directed
ANXA2P2	ANXA2	0.0001	-2.37	undirected
ANXA2P2	AURKB	1.9e-06	-2.91	undirected
ANXA2	AURKB	1.3e-06	2.96	undirected
AURKB	BIRC5	0.0096	1.59	undirected
ANXA2P2	CAMK2A	0.0033	1.80	undirected
ANXA2	CAMK2A	0.0027	-1.83	undirected
AURKB	CAMK2A	0.0077	-1.63	undirected

ANXA2	CARNS1	0.0096	1.59	undirected
ANXA2P2	CCL3	4.2e-05	-2.50	undirected
ANXA2	CCL3	5.4e-05	2.48	undirected
AURKB	CCL3	0.00012	2.36	undirected
BOK	CCL3	0.0099	1.58	undirected
ANXA2P2	CCL4	0.0016	1.93	undirected
ANXA2	CCL4	0.0012	-1.97	undirected
AURKB	CCL4	0.00069	-2.07	undirected
ANXA2P2	CD74	0.00055	-2.11	undirected
ANXA2	CD74	0.00082	2.05	undirected
AURKB	CD74	0.0017	1.92	undirected
CCL3	CD74	0.0014	1.96	undirected
CCL4	CD74	0.0023	-1.86	undirected
ANXA2P2	CKAP2L	1e-05	2.70	undirected
ANXA2	CKAP2L	2.6e-05	-2.57	undirected
ASPM	CKAP2L	0.0051	1.72	undirected
AURKB	CKAP2L	2.7e-05	-2.57	undirected
CAMK2A	CKAP2L	0.0041	1.76	undirected
CARNS1	CKAP2L	0.0045	-1.73	undirected
CCL3	CKAP2L	0.00044	-2.15	undirected
CCL4	CKAP2L	0.003	1.82	undirected
CD74	CKAP2L	5.8e-05	-2.46	undirected
ANXA2P2	COL4A1	1.9e-06	2.92	undirected
ANXA2	COL4A1	6e-07	-3.05	undirected
AURKB	COL4A1	1.5e-06	-2.94	undirected
BOK	COL4A1	0.0012	-1.97	undirected
CAMK2A	COL4A1	0.0027	1.84	undirected
CCL3	COL4A1	0.00024	-2.24	undirected
CCL4	COL4A1	0.0014	1.96	undirected
CD74	COL4A1	6.7e-05	-2.44	undirected
CKAP2L	COL4A1	5.6e-06	2.78	undirected
ANXA2P2	COL4A2	1.6e-06	-2.93	undirected
ANXA2	COL4A2	3.1e-07	3.13	undirected
AURKB	COL4A2	6.7e-06	2.76	undirected
BOK	COL4A2	0.00084	2.05	undirected
CAMK2A	COL4A2	0.0013	-1.96	undirected
CCL3	COL4A2	0.00013	2.35	undirected
CCL4	COL4A2	0.00091	-2.02	undirected
CD74	COL4A2	6.1e-05	2.46	undirected
CKAP2L	COL4A2	1.1e-06	-2.97	undirected

COL4A1	COL4A2	5.7e-05	-2.46	undirected
ANXA2P2	CYBA	3.5e-06	-2.83	undirected
ANXA2	CYBA	3.2e-06	2.85	undirected
AURKB	CYBA	7.2e-06	2.75	undirected
BOK	CYBA	0.0031	1.81	undirected
CAMK2A	CYBA	0.0025	-1.84	undirected
CCL3	CYBA	4.6e-05	2.50	undirected
CCL4	CYBA	0.00028	-2.22	undirected
CD74	CYBA	0.00043	2.16	undirected
CKAP2L	CYBA	1.2e-06	-2.97	undirected
COL4A1	CYBA	1.2e-05	-2.68	undirected
COL4A2	CYBA	1.1e-05	2.69	undirected
ANXA2P2	DOCK2	0.0069	1.66	undirected
AURKB	DOCK2	0.0064	-1.66	undirected
COL4A1	DOCK2	0.002	1.90	undirected
COL4A2	DOCK2	0.0011	-1.99	undirected
CYBA	DOCK2	0.0011	-1.99	undirected
ANXA2P2	ERMN	1.2e-06	2.97	undirected
ANXA2	ERMN	5.1e-07	-3.07	undirected
AURKB	ERMN	0.0013	-1.96	undirected
BOK	ERMN	0.0039	-1.76	undirected
CCL3	ERMN	1.3e-05	-2.67	undirected
CCL4	ERMN	0.00014	2.34	undirected
CD74	ERMN	0.00056	-2.11	undirected
CKAP2L	ERMN	5.4e-08	3.33	undirected
COL4A1	ERMN	6.5e-06	2.76	undirected
COL4A2	ERMN	4.1e-05	-2.51	undirected
CYBA	ERMN	1.1e-05	-2.69	undirected
DOCK2	ERMN	0.0016	1.93	undirected
COL4A1	FCGR1A	0.007	1.65	undirected
COL4A2	FCGR1A	0.0083	-1.61	undirected
CYBA	FCGR1A	0.0063	-1.67	undirected
ANXA2P2	GABRB2	8.4e-05	2.41	undirected
ANXA2	GABRB2	7.2e-05	-2.42	undirected
AURKB	GABRB2	1.3e-06	-2.95	undirected
CAMK2A	GABRB2	0.0072	1.65	undirected
CCL3	GABRB2	0.00024	-2.24	undirected
CCL4	GABRB2	0.0013	1.97	undirected
CD74	GABRB2	0.00038	-2.17	undirected
CKAP2L	GABRB2	3e-06	2.86	undirected

COL4A1	GABRB2	4.8e-05	2.49	undirected
COL4A2	GABRB2	4.7e-05	-2.48	undirected
CYBA	GABRB2	8.4e-05	-2.40	undirected
DOCK2	GABRB2	0.0011	2.01	undirected
ERMN	GABRB2	0.00017	2.30	undirected
ANXA2P2	GJC2	2.2e-05	-2.59	undirected
ANXA2	GJC2	2.2e-05	2.60	undirected
AURKB	GJC2	1.9e-05	2.62	undirected
BOK	GJC2	0.00032	2.21	undirected
CCL3	GJC2	0.0033	1.80	undirected
CCL4	GJC2	0.006	-1.68	undirected
CD74	GJC2	3.9e-05	2.52	undirected
CKAP2L	GJC2	0.0002	-2.27	undirected
COL4A1	GJC2	0.00031	-2.20	undirected
COL4A2	GJC2	0.00023	2.26	undirected
CYBA	GJC2	9.7e-05	2.39	undirected
ERMN	GJC2	3.9e-05	-2.51	undirected
GABRB2	GJC2	0.00064	-2.08	undirected
ANXA2P2	HBA2	2.4e-08	-3.41	undirected
ANXA2	HBA2	3e-08	3.40	undirected
AURKB	HBA2	2.7e-06	2.87	undirected
BOK	HBA2	0.004	1.76	undirected
CAMK2A	HBA2	0.0012	-1.98	undirected
CCL3	HBA2	3.1e-06	2.86	undirected
CCL4	HBA2	4.8e-05	-2.48	undirected
CD74	HBA2	0.00048	2.14	undirected
CKAP2L	HBA2	4.2e-10	-3.82	undirected
COL4A1	HBA2	1.3e-07	-3.22	undirected
COL4A2	HBA2	2.4e-07	3.16	undirected
CYBA	HBA2	2.4e-07	3.16	undirected
DOCK2	HBA2	0.00085	-2.04	undirected
ERMN	HBA2	6.6e-06	-2.75	undirected
GABRB2	HBA2	9.8e-07	-2.99	undirected
GJC2	HBA2	3.4e-06	2.85	undirected
ANXA2P2	HBB	4.1e-08	3.36	undirected
ANXA2	HBB	4.7e-08	-3.34	undirected
AURKB	HBB	5.4e-06	-2.78	undirected
BOK	HBB	0.0045	-1.73	undirected
CAMK2A	HBB	0.0014	1.95	undirected
CCL3	HBB	3.4e-06	-2.84	undirected

CCL4	HBB	4.3e-05	2.51	undirected
CD74	HBB	0.00059	-2.10	undirected
CKAP2L	HBB	5.5e-10	3.80	undirected
COL4A1	HBB	1.8e-07	3.20	undirected
COL4A2	HBB	2.6e-07	-3.14	undirected
CYBA	HBB	5.6e-07	-3.06	undirected
DOCK2	HBB	0.00086	2.04	undirected
ERMN	HBB	1.3e-05	2.68	undirected
GABRB2	HBB	1.5e-06	2.95	undirected
GJC2	HBB	2.5e-06	-2.87	undirected
HBA2	HBB	7.7e-07	-3.02	undirected
ANXA2P2	HLA-DOA	7e-05	2.44	undirected
ANXA2	HLA-DOA	7e-05	-2.43	undirected
AURKB	HLA-DOA	3.8e-05	-2.51	undirected
CAMK2A	HLA-DOA	0.0098	1.58	undirected
CCL3	HLA-DOA	0.00037	-2.17	undirected
CCL4	HLA-DOA	0.0049	1.73	undirected
CD74	HLA-DOA	0.00096	-2.02	undirected
CKAP2L	HLA-DOA	0.00023	2.26	undirected
COL4A1	HLA-DOA	8.3e-06	2.73	undirected
COL4A2	HLA-DOA	6.6e-06	-2.75	undirected
CYBA	HLA-DOA	2.5e-05	-2.58	undirected
ERMN	HLA-DOA	5.1e-05	2.48	undirected
GABRB2	HLA-DOA	0.00026	2.24	undirected
GJC2	HLA-DOA	8.8e-05	-2.39	undirected
HBA2	HLA-DOA	3.9e-07	-3.10	undirected
HBB	HLA-DOA	5.7e-07	3.06	undirected
AURKB	HLA-DPA1	0.0045	-1.74	undirected
COL4A1	HLA-DPA1	0.007	1.65	undirected
COL4A2	HLA-DPA1	0.0082	-1.61	undirected
CYBA	HLA-DPA1	0.0025	-1.85	undirected
ERMN	HLA-DPA1	0.0037	1.78	undirected
HBA2	HLA-DPA1	0.0052	-1.70	undirected
HBB	HLA-DPA1	0.0053	1.71	undirected
HLA-DOA	HLA-DPA1	0.0037	1.78	undirected
ANXA2P2	HLA-DPB1	0.001	-2.01	undirected
ANXA2	HLA-DPB1	0.0012	1.99	undirected
AURKB	HLA-DPB1	2.6e-05	2.58	undirected
BOK	HLA-DPB1	0.006	1.69	undirected
CARNS1	HLA-DPB1	0.004	1.76	undirected

CD74	HLA-DPB1	7.8e-07	3.03	undirected
CKAP2L	HLA-DPB1	0.0079	-1.62	undirected
COL4A1	HLA-DPB1	0.00079	-2.05	undirected
COL4A2	HLA-DPB1	0.00042	2.16	undirected
CYBA	HLA-DPB1	1.1e-05	2.70	undirected
ERMN	HLA-DPB1	2.4e-06	-2.88	undirected
FCGR1A	HLA-DPB1	0.0049	-1.72	undirected
GABRB2	HLA-DPB1	0.00039	-2.16	undirected
HBA2	HLA-DPB1	2.7e-06	2.88	undirected
HBB	HLA-DPB1	1.1e-06	-2.98	undirected
HLA-DOA	HLA-DPB1	0.0034	-1.79	undirected
ANXA2P2	HLA-DRA	9.3e-05	2.40	undirected
ANXA2	HLA-DRA	9.5e-05	-2.38	undirected
AURKB	HLA-DRA	0.00093	-2.02	undirected
CCL3	HLA-DRA	0.00073	-2.06	undirected
CCL4	HLA-DRA	0.0031	1.82	undirected
CD74	HLA-DRA	0.0078	-1.62	undirected
CKAP2L	HLA-DRA	7e-05	2.44	undirected
COL4A1	HLA-DRA	0.00011	2.38	undirected
COL4A2	HLA-DRA	0.0001	-2.37	undirected
CYBA	HLA-DRA	8.4e-05	-2.40	undirected
DOCK2	HLA-DRA	0.009	1.60	undirected
ERMN	HLA-DRA	0.00034	2.20	undirected
GABRB2	HLA-DRA	5e-05	2.49	undirected
GJC2	HLA-DRA	0.00018	-2.28	undirected
HBA2	HLA-DRA	0.00014	-2.32	undirected
HBB	HLA-DRA	0.0002	2.28	undirected
HLA-DOA	HLA-DRA	2.9e-05	2.56	undirected
HLA-DPA1	HLA-DRA	0.0052	1.71	undirected
HLA-DPB1	HLA-DRA	0.00078	-2.05	undirected
AURKB	KIAA1045	0.0087	-1.60	undirected
CKAP2L	KIAA1045	0.0068	1.66	undirected
COL4A1	KIAA1045	0.0066	1.67	undirected
COL4A2	KIAA1045	0.0064	-1.66	undirected
CYBA	KIAA1045	0.0027	-1.83	undirected
ERMN	KIAA1045	0.0057	1.69	undirected
HBA2	KIAA1045	0.0017	-1.92	undirected
HBB	KIAA1045	0.002	1.90	undirected
HLA-DRA	KIAA1045	0.0094	1.59	undirected
ANXA2P2	MOG	0.00068	2.08	undirected

ANXA2	MOG	0.00054	-2.11	undirected
AURKB	MOG	0.00041	-2.16	undirected
CCL3	MOG	0.003	-1.81	undirected
CD74	MOG	0.0075	-1.63	undirected
CKAP2L	MOG	0.00057	2.11	undirected
COL4A1	MOG	0.00017	2.30	undirected
COL4A2	MOG	0.00011	-2.36	undirected
CYBA	MOG	1.9e-05	-2.61	undirected
ERMN	MOG	0.00094	2.03	undirected
GABRB2	MOG	0.0038	1.78	undirected
GJC2	MOG	0.0058	-1.69	undirected
HBA2	MOG	1.2e-05	-2.67	undirected
HBB	MOG	1.3e-05	2.68	undirected
HLA-DOA	MOG	0.00067	2.09	undirected
HLA-DRA	MOG	0.0022	1.88	undirected
ANXA2P2	NEUROD2	3e-06	-2.85	undirected
ANXA2	NEUROD2	4.4e-06	2.81	undirected
AURKB	NEUROD2	0.0001	2.38	undirected
BOK	NEUROD2	0.0012	1.99	undirected
CCL3	NEUROD2	0.00039	2.18	undirected
CCL4	NEUROD2	0.0012	-1.98	undirected
CD74	NEUROD2	0.00029	2.22	undirected
CKAP2L	NEUROD2	3.9e-05	-2.51	undirected
COL4A1	NEUROD2	6.2e-06	-2.76	undirected
COL4A2	NEUROD2	1.1e-05	2.70	undirected
CYBA	NEUROD2	2.5e-06	2.88	undirected
ERMN	NEUROD2	0.00013	-2.34	undirected
GABRB2	NEUROD2	2.3e-06	-2.89	undirected
GJC2	NEUROD2	0.00049	2.14	undirected
HBA2	NEUROD2	1e-06	2.99	undirected
HBB	NEUROD2	1.2e-06	-2.96	undirected
HLA-DOA	NEUROD2	1.5e-05	-2.65	undirected
HLA-DPB1	NEUROD2	0.0051	1.72	undirected
HLA-DRA	NEUROD2	0.00011	-2.36	undirected
MOG	NEUROD2	0.0049	-1.72	undirected
ANXA2P2	NRGN	2.6e-05	-2.57	undirected
ANXA2	NRGN	3.6e-05	2.53	undirected
AURKB	NRGN	5.2e-06	2.79	undirected
BOK	NRGN	0.0081	1.63	undirected
CARNS1	NRGN	0.0049	1.73	undirected

CCL3	NRGN	0.00044	2.15	undirected
CCL4	NRGN	0.0021	-1.87	undirected
CD74	NRGN	0.00032	2.21	undirected
CKAP2L	NRGN	0.00088	-2.03	undirected
COL4A1	NRGN	2.8e-05	-2.56	undirected
COL4A2	NRGN	1.4e-05	2.66	undirected
CYBA	NRGN	1.3e-05	2.67	undirected
ERMN	NRGN	5.1e-07	-3.07	undirected
GABRB2	NRGN	8.6e-05	-2.40	undirected
GJC2	NRGN	0.0011	2.01	undirected
HBA2	NRGN	1.5e-07	3.22	undirected
HBB	NRGN	1.8e-07	-3.19	undirected
HLA-DOA	NRGN	0.00011	-2.37	undirected
HLA-DPB1	NRGN	0.0063	1.68	undirected
HLA-DRA	NRGN	0.00016	-2.31	undirected
MOG	NRGN	0.0039	-1.76	undirected
NEUROD2	NRGN	3.2e-05	2.55	undirected
ANXA2P2	NUSAP1	0.0072	1.65	undirected
ANXA2	NUSAP1	0.006	-1.68	undirected
AURKB	NUSAP1	0.0055	-1.70	undirected
CD74	NUSAP1	0.004	-1.76	undirected
COL4A1	NUSAP1	0.0082	1.62	undirected
COL4A2	NUSAP1	0.0073	-1.64	undirected
CYBA	NUSAP1	0.0053	-1.70	undirected
ERMN	NUSAP1	0.0039	1.77	undirected
GABRB2	NUSAP1	0.0064	1.67	undirected
HBA2	NUSAP1	0.0015	-1.94	undirected
HBB	NUSAP1	0.0015	1.94	undirected
ANXA2P2	SH3TC2	0.0054	1.71	undirected
ANXA2	SH3TC2	0.0078	-1.62	undirected
AURKB	SH3TC2	0.0018	-1.90	undirected
CD74	SH3TC2	0.0046	-1.73	undirected
COL4A1	SH3TC2	0.0027	1.84	undirected
COL4A2	SH3TC2	0.0025	-1.85	undirected
CYBA	SH3TC2	0.00063	-2.09	undirected
ERMN	SH3TC2	0.0059	1.69	undirected
HBA2	SH3TC2	0.00056	-2.11	undirected
HBB	SH3TC2	0.00057	2.11	undirected
HLA-DRA	SH3TC2	0.0045	1.74	undirected
ANXA2P2	SLC17A7	0.0031	-1.81	undirected

ANXA2	SLC17A7	0.0035	1.79	undirected
CKAP2L	SLC17A7	0.003	-1.81	undirected
COL4A1	SLC17A7	0.0072	-1.64	undirected
COL4A2	SLC17A7	0.0075	1.64	undirected
CYBA	SLC17A7	0.0082	1.62	undirected
HBA2	SLC17A7	0.002	1.90	undirected
HBB	SLC17A7	0.0023	-1.86	undirected
HLA-DOA	SLC17A7	0.0074	-1.63	undirected
HLA-DPB1	SLC17A7	0.007	1.65	undirected
HLA-DRA	SLC17A7	0.0081	-1.62	undirected
NEUROD2	SLC17A7	0.00062	2.10	undirected
NRGN	SLC17A7	0.002	1.90	undirected
GJC2	SLC5A11	0.0084	1.62	undirected
ANXA2P2	SNAP25	3.5e-06	2.84	undirected
ANXA2	SNAP25	6.3e-06	-2.76	undirected
AURKB	SNAP25	1e-05	-2.69	undirected
CAMK2A	SNAP25	0.0078	1.63	undirected
CCL3	SNAP25	7.9e-05	-2.41	undirected
CCL4	SNAP25	0.00076	2.06	undirected
CD74	SNAP25	0.00073	-2.06	undirected
CKAP2L	SNAP25	1.1e-06	2.99	undirected
COL4A1	SNAP25	5.3e-07	3.07	undirected
COL4A2	SNAP25	6.8e-07	-3.03	undirected
CYBA	SNAP25	2.7e-06	-2.87	undirected
DOCK2	SNAP25	0.007	1.65	undirected
ERMN	SNAP25	0.0001	2.38	undirected
GABRB2	SNAP25	0.0012	1.98	undirected
GJC2	SNAP25	0.0004	-2.16	undirected
HBA2	SNAP25	2.7e-07	-3.14	undirected
HBB	SNAP25	4.9e-07	3.08	undirected
HLA-DOA	SNAP25	4.1e-05	2.51	undirected
HLA-DPB1	SNAP25	0.0015	-1.94	undirected
HLA-DRA	SNAP25	0.00065	2.09	undirected
NEUROD2	SNAP25	6.1e-06	-2.76	undirected
NRGN	SNAP25	0.00013	-2.34	undirected
NUSAP1	SNAP25	0.0076	1.64	undirected
ANXA2P2	SNCB	0.00015	-2.32	undirected
ANXA2	SNCB	9.3e-05	2.39	undirected
AURKB	SNCB	3.6e-05	2.53	undirected
CCL3	SNCB	0.0004	2.17	undirected

CCL4	SNCB	0.0053	-1.70	undirected
CD74	SNCB	0.0013	1.97	undirected
CKAP2L	SNCB	6.4e-06	-2.76	undirected
COL4A1	SNCB	3.3e-06	-2.84	undirected
COL4A2	SNCB	2.7e-06	2.87	undirected
CYBA	SNCB	1.4e-05	2.67	undirected
ERMN	SNCB	0.0003	-2.21	undirected
GABRB2	SNCB	0.00016	-2.30	undirected
GJC2	SNCB	0.0035	1.79	undirected
HBA2	SNCB	1.1e-08	3.50	undirected
HBB	SNCB	1.3e-08	-3.48	undirected
HLA-DOA	SNCB	0.00096	-2.02	undirected
HLA-DPB1	SNCB	0.0081	1.63	undirected
HLA-DRA	SNCB	0.0004	-2.16	undirected
NEUROD2	SNCB	0.00024	2.25	undirected
NRGN	SNCB	0.00013	2.34	undirected
SNAP25	SNCB	2.3e-05	-2.59	undirected
ANXA2P2	SPI1	0.0015	-1.94	undirected
ANXA2	SPI1	0.0019	1.90	undirected
CCL3	SPI1	0.0029	1.83	undirected
CCL4	SPI1	0.0025	-1.84	undirected
CKAP2L	SPI1	0.00039	-2.17	undirected
COL4A1	SPI1	0.0018	-1.91	undirected
COL4A2	SPI1	0.0029	1.83	undirected
CYBA	SPI1	0.00025	2.25	undirected
GABRB2	SPI1	0.00095	-2.02	undirected
GJC2	SPI1	0.00021	2.27	undirected
HBA2	SPI1	0.0087	1.61	undirected
HBB	SPI1	0.0099	-1.57	undirected
HLA-DOA	SPI1	0.0036	-1.77	undirected
HLA-DPB1	SPI1	0.0005	2.13	undirected
NEUROD2	SPI1	0.0067	1.66	undirected
NRGN	SPI1	0.0011	2.00	undirected
SNAP25	SPI1	0.009	-1.59	undirected
SNCB	SPI1	0.0013	1.98	undirected
COL4A1	SULT4A1	0.0097	-1.58	undirected
COL4A2	SULT4A1	0.0092	1.60	undirected
CYBA	SULT4A1	0.0093	1.59	undirected
HBA2	SULT4A1	0.0073	1.65	undirected
HBB	SULT4A1	0.0074	-1.63	undirected

NRGN	SULT4A1	0.0087	1.61	undirected
ANXA2P2	TBR1	0.002	1.90	undirected
ANXA2	TBR1	0.0021	-1.87	undirected
AURKB	TBR1	0.0014	-1.94	undirected
CD74	TBR1	0.0023	-1.86	undirected
COL4A1	TBR1	0.0011	2.00	undirected
COL4A2	TBR1	0.00067	-2.08	undirected
CYBA	TBR1	3.7e-05	-2.52	undirected
ERMN	TBR1	1.7e-05	2.63	undirected
GABRB2	TBR1	8.6e-05	2.41	undirected
HBA2	TBR1	3.6e-05	-2.52	undirected
HBB	TBR1	2.8e-05	2.57	undirected
HLA-DOA	TBR1	0.0077	1.63	undirected
HLA-DRA	TBR1	0.0041	1.76	undirected
SNAP25	TBR1	0.0021	1.88	undirected
SNCB	TBR1	0.007	-1.65	undirected
SPI1	TBR1	0.004	-1.75	undirected
ANXA2P2	TMEM144	0.00022	2.27	undirected
ANXA2	TMEM144	0.00022	-2.26	undirected
AURKB	TMEM144	0.0012	-1.98	undirected
CCL3	TMEM144	0.00067	-2.08	undirected
CCL4	TMEM144	0.0027	1.84	undirected
CD74	TMEM144	0.0084	-1.61	undirected
CKAP2L	TMEM144	0.00059	2.11	undirected
COL4A1	TMEM144	6.5e-05	2.45	undirected
COL4A2	TMEM144	5.1e-05	-2.47	undirected
CYBA	TMEM144	2.3e-05	-2.59	undirected
ERMN	TMEM144	0.00016	2.31	undirected
GABRB2	TMEM144	9.7e-05	2.39	undirected
GJC2	TMEM144	0.00054	-2.11	undirected
HBA2	TMEM144	0.00013	-2.34	undirected
HBB	TMEM144	0.00015	2.32	undirected
HLA-DOA	TMEM144	0.0014	1.96	undirected
HLA-DPB1	TMEM144	0.0046	-1.73	undirected
HLA-DRA	TMEM144	0.0033	1.80	undirected
KIAA1045	TMEM144	0.0098	1.58	undirected
NEUROD2	TMEM144	7.5e-05	-2.42	undirected
NRGN	TMEM144	0.0009	-2.03	undirected
SNAP25	TMEM144	0.00054	2.12	undirected
SNCB	TMEM144	0.0055	-1.69	undirected

ANXA2P2	TOP2A	0.00026	2.24	undirected
ANXA2	TOP2A	0.00018	-2.29	undirected
AURKB	TOP2A	0.00031	-2.20	undirected
BOK	TOP2A	0.008	-1.62	undirected
CD74	TOP2A	0.0012	-1.97	undirected
CKAP2L	TOP2A	0.00039	2.18	undirected
COL4A1	TOP2A	0.0028	1.83	undirected
COL4A2	TOP2A	0.0026	-1.84	undirected
CYBA	TOP2A	0.00066	-2.08	undirected
ERMN	TOP2A	0.0017	1.93	undirected
GABRB2	TOP2A	0.0033	1.80	undirected
HBA2	TOP2A	9.5e-05	-2.38	undirected
HBB	TOP2A	0.0001	2.38	undirected
HLA-DOA	TOP2A	0.00092	2.03	undirected
HLA-DRA	TOP2A	0.00052	2.13	undirected
NEUROD2	TOP2A	0.0015	-1.94	undirected
NRGN	TOP2A	0.003	-1.81	undirected
SNAP25	TOP2A	0.0034	1.80	undirected
SPI1	TOP2A	0.0048	-1.72	undirected
TMEM144	TOP2A	0.002	1.89	undirected
ANXA2P2	UBE2C	5.8e-05	-2.46	undirected
ANXA2	UBE2C	8.2e-05	2.41	undirected
AURKB	UBE2C	3.7e-06	2.83	undirected
CCL3	UBE2C	4.1e-05	2.52	undirected
CCL4	UBE2C	0.00091	-2.02	undirected
CD74	UBE2C	0.0025	1.86	undirected
CKAP2L	UBE2C	7.9e-06	-2.73	undirected
COL4A1	UBE2C	4.1e-06	-2.81	undirected
COL4A2	UBE2C	4.1e-06	2.82	undirected
CYBA	UBE2C	1.1e-05	2.69	undirected
ERMN	UBE2C	1.5e-05	-2.64	undirected
GABRB2	UBE2C	0.00054	-2.11	undirected
GJC2	UBE2C	0.00018	2.29	undirected
HBA2	UBE2C	2e-06	2.91	undirected
HBB	UBE2C	2.5e-06	-2.88	undirected
HLA-DOA	UBE2C	0.00029	-2.21	undirected
HLA-DPB1	UBE2C	0.0021	1.89	undirected
HLA-DRA	UBE2C	0.00017	-2.30	undirected
MOG	UBE2C	0.0095	-1.58	undirected
NEUROD2	UBE2C	3.9e-05	2.52	undirected

NRGN	UBE2C	9e-05	2.40	undirected
NUSAP1	UBE2C	0.0057	-1.69	undirected
SNAP25	UBE2C	0.00033	-2.19	undirected
SNCB	UBE2C	0.00021	2.28	undirected
SPI1	UBE2C	0.0031	1.81	undirected
TBR1	UBE2C	0.0078	-1.62	undirected
TMEM144	UBE2C	0.0025	-1.84	undirected
TOP2A	UBE2C	0.0028	-1.82	undirected

B.2 Oligodendroglioma vs Astrocytoma

Table B.2: Oligodendroglioma iDINGO summary denoting the differential edges with the respective p-value, differential score and edge type.

Source node	Target node	p-value	differential score	interaction type
cg00962781	AK5	0.00045	1.64	directed
cg02127209	AK5	0.00013	-1.83	directed
cg02571816	AK5	6.1e-12	-3.28	directed
cg02589899	AK5	0.01	-1.24	directed
cg02700894	AK5	0.0012	1.52	directed
cg05851240	AK5	1.9e-06	2.24	directed
cg13572782	AK5	0.006	1.28	directed
cg16409049	AK5	3e-05	1.96	directed
cg17266581	AK5	0.0017	-1.51	directed
cg20801056	AK5	8.5e-06	2.09	directed
cg21245652	AK5	7.5e-05	1.86	directed
cg22213386	AK5	8.9e-07	2.31	directed
cg22266001	AK5	8.2e-06	2.09	directed
cg22839075	AK5	0.0055	-1.34	directed
cg23241335	AK5	0.0094	-1.25	directed
cg23483765	AK5	2e-09	-2.86	directed
cg25887236	AK5	4.9e-06	-2.19	directed
cg26385222	AK5	0.0029	1.39	directed
cg02127209	AURKB	0.0072	-1.30	directed
cg02571816	AURKB	1e-06	-2.34	directed
cg02700894	AURKB	0.0082	1.23	directed
cg05851240	AURKB	9.5e-05	1.83	directed
cg16459364	AURKB	0.0055	1.29	directed
cg20801056	AURKB	0.0005	1.63	directed
cg21245652	AURKB	0.0026	1.41	directed

cg22213386	AURKB	0.001	1.54	directed
cg23483765	AURKB	0.0003	-1.73	directed
cg25887236	AURKB	0.0073	-1.29	directed
cg26385222	AURKB	0.0034	1.37	directed
cg00962781	BUB1	0.0094	-1.25	directed
cg02127209	BUB1	7.8e-05	1.85	directed
cg02571816	BUB1	4e-09	2.77	directed
cg02700894	BUB1	0.0031	-1.43	directed
cg05851240	BUB1	0.0005	-1.67	directed
cg16409049	BUB1	0.0051	-1.35	directed
cg17266581	BUB1	0.0039	1.35	directed
cg20801056	BUB1	0.00028	-1.74	directed
cg21245652	BUB1	0.0016	-1.51	directed
cg22213386	BUB1	0.0012	-1.56	directed
cg22266001	BUB1	0.00091	-1.59	directed
cg22839075	BUB1	0.001	1.54	directed
cg23483765	BUB1	0.0013	1.50	directed
cg25887236	BUB1	0.0002	1.74	directed
cg26385222	BUB1	0.00087	-1.60	directed
cg00962781	C1QB	0.0077	-1.28	directed
cg02127209	C1QB	0.00062	1.60	directed
cg02571816	C1QB	1.5e-07	2.47	directed
cg05851240	C1QB	0.0035	-1.41	directed
cg20801056	C1QB	0.0039	-1.39	directed
cg22213386	C1QB	0.00025	-1.76	directed
cg22266001	C1QB	0.00044	-1.69	directed
cg22839075	C1QB	0.0019	1.45	directed
cg23483765	C1QB	0.00094	1.55	directed
cg25887236	C1QB	0.00017	1.76	directed
cg00962781	C1QC	0.0068	1.26	directed
cg02127209	C1QC	0.00099	-1.58	directed
cg02571816	C1QC	7.6e-10	-2.94	directed
cg02700894	C1QC	0.0015	1.48	directed
cg05851240	C1QC	1e-05	2.07	directed
cg13572782	C1QC	0.0071	1.26	directed
cg17266581	C1QC	0.0083	-1.27	directed
cg20801056	C1QC	0.00023	1.73	directed
cg21245652	C1QC	0.0037	1.36	directed
cg22213386	C1QC	0.00023	1.73	directed
cg22266001	C1QC	0.0014	1.49	directed

cg22839075	C1QC	0.0011	-1.57	directed
cg23483765	C1QC	4.1e-05	-1.96	directed
cg25887236	C1QC	0.00033	-1.72	directed
cg02571816	CARNS1	2.6e-05	-2.02	directed
cg05851240	CARNS1	0.0032	1.38	directed
cg20801056	CARNS1	0.005	1.31	directed
cg21245652	CARNS1	0.0039	1.35	directed
cg22213386	CARNS1	0.0015	1.49	directed
cg00962781	CDK1	0.00072	-1.62	directed
cg02127209	CDK1	0.00036	1.67	directed
cg02571816	CDK1	3.6e-12	3.27	directed
cg02700894	CDK1	0.0016	-1.52	directed
cg04920452	CDK1	0.0087	-1.26	directed
cg05851240	CDK1	1.3e-05	-2.09	directed
cg13572782	CDK1	0.0023	-1.47	directed
cg16409049	CDK1	4.5e-05	-1.96	directed
cg16459364	CDK1	0.0094	-1.25	directed
cg17266581	CDK1	0.00039	1.66	directed
cg18898125	CDK1	0.0092	1.21	directed
cg19384775	CDK1	0.0096	1.21	directed
cg20801056	CDK1	2.2e-06	-2.26	directed
cg21245652	CDK1	0.00011	-1.85	directed
cg22213386	CDK1	4.5e-06	-2.19	directed
cg22266001	CDK1	8.6e-05	-1.88	directed
cg22839075	CDK1	0.00019	1.75	directed
cg23241335	CDK1	0.0031	1.38	directed
cg23483765	CDK1	9e-08	2.51	directed
cg25887236	CDK1	1.4e-06	2.27	directed
cg26385222	CDK1	0.00039	-1.70	directed
cg00962781	CENPF	0.0039	1.35	directed
cg02127209	CENPF	0.002	-1.49	directed
cg02571816	CENPF	1e-09	-2.92	directed
cg02700894	CENPF	0.0011	1.53	directed
cg05851240	CENPF	1.2e-07	2.49	directed
cg13572782	CENPF	0.0016	1.47	directed
cg16409049	CENPF	0.0015	1.49	directed
cg16459364	CENPF	0.0069	1.26	directed
cg17266581	CENPF	0.0017	-1.51	directed
cg18898125	CENPF	0.0032	-1.42	directed
cg20801056	CENPF	8.8e-05	1.84	directed

cg21245652	CENPF	6.4e-05	1.87	directed
cg22213386	CENPF	5.5e-06	2.13	directed
cg22266001	CENPF	0.0021	1.44	directed
cg22839075	CENPF	0.0072	-1.29	directed
cg23483765	CENPF	4.5e-06	-2.20	directed
cg25887236	CENPF	1.2e-07	-2.53	directed
cg26385222	CENPF	0.0034	1.37	directed
cg02571816	CKAP2L	4.1e-07	2.38	directed
cg05851240	CKAP2L	0.00016	-1.81	directed
cg16409049	CKAP2L	0.007	-1.30	directed
cg20801056	CKAP2L	0.0059	-1.33	directed
cg21245652	CKAP2L	0.0084	-1.27	directed
cg22213386	CKAP2L	0.0013	-1.54	directed
cg23483765	CKAP2L	0.00073	1.58	directed
cg02571816	CNDP1	3.7e-06	-2.21	directed
cg05851240	CNDP1	0.0013	1.50	directed
cg23483765	CNDP1	0.00057	-1.65	directed
cg02571816	CRYM	5.5e-08	2.56	directed
cg05851240	CRYM	0.0046	-1.36	directed
cg20801056	CRYM	0.00036	-1.71	directed
cg22213386	CRYM	0.003	-1.43	directed
cg22839075	CRYM	0.0048	1.32	directed
cg23483765	CRYM	0.00019	1.75	directed
cg25887236	CRYM	0.0032	1.38	directed
cg02571816	DTL	0.00047	-1.68	directed
cg25887236	DTL	0.0085	-1.27	directed
cg00962781	FCGR1A	0.0041	-1.38	directed
cg02127209	FCGR1A	0.0017	1.47	directed
cg02571816	FCGR1A	3.2e-10	2.96	directed
cg02700894	FCGR1A	0.0043	-1.38	directed
cg05851240	FCGR1A	3.7e-06	-2.21	directed
cg13572782	FCGR1A	0.0081	-1.28	directed
cg16409049	FCGR1A	0.0033	-1.42	directed
cg16459364	FCGR1A	0.0039	-1.39	directed
cg17266581	FCGR1A	0.0024	1.42	directed
cg18898125	FCGR1A	0.0091	1.22	directed
cg20801056	FCGR1A	3.8e-05	-1.97	directed
cg21245652	FCGR1A	0.0019	-1.49	directed
cg22213386	FCGR1A	4.7e-05	-1.95	directed
cg22266001	FCGR1A	0.0059	-1.32	directed

cg22839075	FCGR1A	0.0011	1.52	directed
cg23483765	FCGR1A	2.6e-05	1.97	directed
cg25887236	FCGR1A	9.9e-05	1.83	directed
cg26385222	FCGR1A	0.0072	-1.30	directed
cg02127209	FCGR1B	5.8e-05	-1.93	directed
cg02571816	FCGR1B	6.1e-12	-3.28	directed
cg02700894	FCGR1B	0.00082	1.57	directed
cg03537215	FCGR1B	0.0031	-1.42	directed
cg05851240	FCGR1B	7.3e-05	1.86	directed
cg13572782	FCGR1B	0.0016	1.47	directed
cg16409049	FCGR1B	5.6e-07	2.35	directed
cg16459364	FCGR1B	0.0015	1.49	directed
cg18898125	FCGR1B	0.0075	-1.29	directed
cg19384775	FCGR1B	0.0099	-1.24	directed
cg20801056	FCGR1B	3.8e-05	1.93	directed
cg21245652	FCGR1B	1.5e-05	2.03	directed
cg22213386	FCGR1B	6.6e-06	2.11	directed
cg22266001	FCGR1B	0.0044	1.33	directed
cg22839075	FCGR1B	1.4e-05	-2.08	directed
cg23483765	FCGR1B	1.4e-07	-2.52	directed
cg25649038	FCGR1B	0.0092	-1.25	directed
cg25887236	FCGR1B	3.3e-05	-1.99	directed
cg26385222	FCGR1B	0.0025	1.41	directed
cg02571816	FCGR1C	9.7e-05	-1.87	directed
cg05851240	FCGR1C	0.0025	1.41	directed
cg22213386	FCGR1C	0.0076	1.25	directed
cg02571816	FOXM1	5.9e-07	-2.39	directed
cg17266581	FOXM1	0.0063	-1.31	directed
cg20801056	FOXM1	0.0047	1.32	directed
cg21245652	FOXM1	0.0093	1.21	directed
cg22213386	FOXM1	0.00075	1.58	directed
cg23483765	FOXM1	0.0016	-1.52	directed
cg25887236	FOXM1	0.005	-1.35	directed
cg00962781	GDA	8.6e-06	-2.13	directed
cg02127209	GDA	0.00018	1.75	directed
cg02571816	GDA	2.9e-10	2.97	directed
cg02589899	GDA	0.0023	1.42	directed
cg02700894	GDA	0.0049	-1.36	directed
cg03537215	GDA	0.001	1.54	directed
cg05851240	GDA	4.8e-06	-2.19	directed

cg16409049	GDA	4.2e-06	-2.20	directed
cg16459364	GDA	0.0042	-1.38	directed
cg17266581	GDA	0.00031	1.69	directed
cg18898125	GDA	0.002	1.44	directed
cg19384775	GDA	0.0029	1.39	directed
cg20801056	GDA	9.8e-05	-1.87	directed
cg21245652	GDA	0.0048	-1.36	directed
cg22213386	GDA	0.00083	-1.61	directed
cg22266001	GDA	6.5e-05	-1.91	directed
cg22839075	GDA	5.3e-06	2.14	directed
cg23241335	GDA	0.0099	1.20	directed
cg23483765	GDA	4e-07	2.38	directed
cg25649038	GDA	0.0082	1.23	directed
cg25887236	GDA	2.4e-08	2.62	directed
cg26385222	GDA	1.8e-05	-2.06	directed
cg02571816	GJC2	2.6e-08	-2.66	directed
cg03537215	GJC2	0.0063	-1.32	directed
cg05851240	GJC2	2.4e-05	1.98	directed
cg16409049	GJC2	0.0043	1.33	directed
cg16459364	GJC2	0.0007	1.59	directed
cg20801056	GJC2	0.00021	1.74	directed
cg21245652	GJC2	0.0094	1.21	directed
cg22213386	GJC2	0.002	1.44	directed
cg22266001	GJC2	0.0087	1.22	directed
cg23483765	GJC2	0.00015	-1.82	directed
cg25887236	GJC2	0.0022	-1.47	directed
cg02571816	HOOK1	6.1e-08	2.55	directed
cg02700894	HOOK1	0.00063	-1.64	directed
cg05851240	HOOK1	0.00013	-1.84	directed
cg20801056	HOOK1	0.0014	-1.54	directed
cg22213386	HOOK1	0.0032	-1.42	directed
cg23483765	HOOK1	0.00091	1.55	directed
cg25887236	HOOK1	0.0033	1.37	directed
cg02571816	KCNK9	4.6e-05	1.91	directed
cg22213386	KCNK9	0.0017	-1.50	directed
cg23483765	KCNK9	0.00053	1.62	directed
cg25887236	KCNK9	0.0051	1.31	directed
cg02127209	KIF20A	0.0052	1.30	directed
cg02571816	KIF20A	0.00031	1.69	directed
cg05851240	KIF20A	0.0063	-1.32	directed

cg20801056	KIF20A	0.01	-1.24	directed
cg22213386	KIF20A	0.0085	-1.27	directed
cg23483765	KIF20A	0.00038	1.66	directed
cg25887236	KIF20A	0.0021	1.44	directed
cg02127209	KIFC1	0.0025	-1.45	directed
cg02571816	KIFC1	6.4e-11	-3.12	directed
cg02700894	KIFC1	0.0018	1.46	directed
cg05851240	KIFC1	0.00011	1.81	directed
cg16409049	KIFC1	0.0059	1.29	directed
cg20801056	KIFC1	0.0001	1.82	directed
cg21245652	KIFC1	0.00062	1.60	directed
cg22213386	KIFC1	2.1e-05	1.99	directed
cg22839075	KIFC1	0.0019	-1.49	directed
cg23483765	KIFC1	8e-05	-1.89	directed
cg25887236	KIFC1	0.00028	-1.74	directed
cg00962781	LOC283999	1.3e-06	2.28	directed
cg02127209	LOC283999	1.6e-08	-2.70	directed
cg02571816	LOC283999	0	-4.57	directed
cg02589899	LOC283999	0.0014	-1.53	directed
cg02700894	LOC283999	0.00075	1.58	directed
cg03537215	LOC283999	0.0041	-1.38	directed
cg04920452	LOC283999	0.0052	1.30	directed
cg05301866	LOC283999	0.00034	-1.72	directed
cg05851240	LOC283999	1.1e-05	2.06	directed
cg13572782	LOC283999	0.00029	1.70	directed
cg15391531	LOC283999	0.0047	1.32	directed
cg16409049	LOC283999	1e-05	2.07	directed
cg16459364	LOC283999	0.00027	1.70	directed
cg17266581	LOC283999	0.00024	-1.76	directed
cg19384775	LOC283999	0.0049	-1.36	directed
cg20801056	LOC283999	6.5e-05	1.87	directed
cg21245652	LOC283999	3.8e-07	2.39	directed
cg22213386	LOC283999	4.4e-05	1.92	directed
cg22266001	LOC283999	0.0001	1.82	directed
cg22839075	LOC283999	0.00014	-1.83	directed
cg23241335	LOC283999	0.0044	-1.37	directed
cg23483765	LOC283999	3.6e-07	-2.43	directed
cg25649038	LOC283999	0.0072	-1.30	directed
cg25887236	LOC283999	8.4e-07	-2.36	directed
cg26385222	LOC283999	3.1e-05	1.95	directed

cg02571816	LOC285780	0.00086	1.56	directed
cg20801056	LOC285780	0.0076	-1.29	directed
cg23483765	LOC285780	0.0036	1.36	directed
cg02571816	LRP2	0.00011	-1.85	directed
cg05851240	LRP2	0.0083	1.23	directed
cg02571816	MAL	0.0032	1.38	directed
cg02571816	MBP	0.0019	-1.50	directed
cg23483765	MBP	0.007	-1.30	directed
cg02127209	MCM10	0.0001	-1.86	directed
cg02571816	MCM10	3.9e-11	-3.15	directed
cg02700894	MCM10	0.0027	1.40	directed
cg05851240	MCM10	0.00029	1.70	directed
cg16409049	MCM10	2.4e-05	1.98	directed
cg17266581	MCM10	0.0022	-1.48	directed
cg20801056	MCM10	0.00087	1.56	directed
cg21245652	MCM10	0.00024	1.72	directed
cg22213386	MCM10	4.1e-06	2.16	directed
cg22266001	MCM10	3.3e-05	1.95	directed
cg22839075	MCM10	0.0011	-1.57	directed
cg23483765	MCM10	6.1e-05	-1.92	directed
cg25742540	MCM10	0.0095	-1.25	directed
cg25887236	MCM10	1.8e-05	-2.05	directed
cg00962781	MELK	0.0013	-1.54	directed
cg02127209	MELK	0.0011	1.53	directed
cg02571816	MELK	1.6e-08	2.66	directed
cg05851240	MELK	0.0002	-1.78	directed
cg16409049	MELK	0.00019	-1.79	directed
cg16459364	MELK	0.0047	-1.36	directed
cg17266581	MELK	0.0022	1.43	directed
cg19384775	MELK	0.0055	1.30	directed
cg20801056	MELK	0.00082	-1.61	directed
cg22213386	MELK	0.0028	-1.44	directed
cg22839075	MELK	0.00081	1.57	directed
cg23483765	MELK	7.2e-06	2.11	directed
cg25887236	MELK	8.3e-06	2.09	directed
cg26385222	MELK	0.00062	-1.64	directed
cg02571816	MKI67	0.0024	-1.46	directed
cg02127209	NCKAP1L	0.0066	1.27	directed
cg02571816	NCKAP1L	5.4e-10	2.92	directed
cg05851240	NCKAP1L	0.0014	-1.53	directed

cg16409049	NCKAP1L	0.00039	-1.70	directed
cg16459364	NCKAP1L	0.0055	-1.34	directed
cg19384775	NCKAP1L	0.0059	1.28	directed
cg20801056	NCKAP1L	0.0029	-1.43	directed
cg21245652	NCKAP1L	0.0029	-1.43	directed
cg22213386	NCKAP1L	2.3e-05	-2.03	directed
cg22266001	NCKAP1L	0.00064	-1.64	directed
cg22839075	NCKAP1L	0.0051	1.31	directed
cg23483765	NCKAP1L	1.6e-05	2.02	directed
cg25887236	NCKAP1L	0.00012	1.80	directed
cg00962781	NECAB1	0.00086	-1.60	directed
cg02127209	NECAB1	6.7e-07	2.34	directed
cg02571816	NECAB1	7.1e-13	3.38	directed
cg02700894	NECAB1	6.7e-05	-1.91	directed
cg05301866	NECAB1	0.0022	1.43	directed
cg05851240	NECAB1	8.4e-07	-2.36	directed
cg15391531	NECAB1	0.007	-1.30	directed
cg16409049	NECAB1	0.00015	-1.82	directed
cg16459364	NECAB1	0.0078	-1.28	directed
cg17266581	NECAB1	0.00021	1.74	directed
cg18898125	NECAB1	0.0099	1.20	directed
cg20801056	NECAB1	2.5e-06	-2.25	directed
cg21245652	NECAB1	0.00027	-1.75	directed
cg22213386	NECAB1	4.3e-08	-2.62	directed
cg22266001	NECAB1	4.2e-06	-2.20	directed
cg22839075	NECAB1	8.3e-05	1.84	directed
cg23241335	NECAB1	0.0058	1.29	directed
cg23483765	NECAB1	2.5e-08	2.62	directed
cg25887236	NECAB1	4.4e-08	2.57	directed
cg26385222	NECAB1	0.00052	-1.67	directed
cg02571816	NEFL	0.0064	1.27	directed
cg02127209	NINJ2	0.0026	1.41	directed
cg02571816	NINJ2	3e-05	1.96	directed
cg05851240	NINJ2	0.0001	-1.86	directed
cg20801056	NINJ2	0.0088	-1.26	directed
cg22213386	NINJ2	0.0025	-1.46	directed
cg23483765	NINJ2	0.0018	1.46	directed
cg25887236	NINJ2	5.9e-05	1.88	directed
cg02571816	NRGN	0.0011	-1.57	directed
cg16409049	NRGN	0.0096	1.21	directed

cg02571816	NUSAP1	0.00098	1.54	directed
cg22213386	NUSAP1	0.0095	-1.25	directed
cg02571816	PACSIN1	5.6e-05	1.89	directed
cg02571816	S1PR5	0.0093	-1.25	directed
cg02571816	SLC17A7	1.2e-05	2.05	directed
cg05851240	SLC17A7	0.0012	-1.55	directed
cg16409049	SLC17A7	0.0056	-1.33	directed
cg20801056	SLC17A7	0.0087	-1.26	directed
cg22213386	SLC17A7	0.0099	-1.24	directed
cg22266001	SLC17A7	0.0042	-1.38	directed
cg23483765	SLC17A7	0.0054	1.30	directed
cg02571816	SPOCK3	0.0085	-1.27	directed
cg25887236	SPOCK3	0.0068	-1.30	directed
cg00962781	STX1A	0	3.92	directed
cg01657744	STX1A	0.0013	-1.55	directed
cg02127209	STX1A	3.2e-12	-3.32	directed
cg02571816	STX1A	0	-5.44	directed
cg02589899	STX1A	4.3e-06	-2.20	directed
cg02700894	STX1A	2.5e-08	2.62	directed
cg03537215	STX1A	0.004	-1.39	directed
cg04920452	STX1A	3e-05	1.96	directed
cg05301866	STX1A	4.4e-09	-2.80	directed
cg05851240	STX1A	0	4.99	directed
cg08326075	STX1A	0.0046	1.32	directed
cg13572782	STX1A	9.1e-10	2.88	directed
cg15391531	STX1A	7e-06	2.11	directed
cg16409049	STX1A	3.2e-14	3.58	directed
cg16459364	STX1A	1.5e-06	2.26	directed
cg17266581	STX1A	4.9e-12	-3.30	directed
cg17860158	STX1A	0.00079	1.57	directed
cg18898125	STX1A	0.00031	-1.73	directed
cg19384775	STX1A	1.2e-06	-2.33	directed
cg20232102	STX1A	0.00061	1.60	directed
cg20801056	STX1A	5.1e-15	3.69	directed
cg21245652	STX1A	4.4e-15	3.70	directed
cg21322241	STX1A	0.0096	-1.25	directed
cg22213386	STX1A	0	4.27	directed
cg22266001	STX1A	2e-14	3.61	directed
cg22557662	STX1A	0.0051	-1.35	directed
cg22839075	STX1A	1e-07	-2.54	directed

cg23241335	STX1A	6.6e-06	-2.16	directed
cg23483765	STX1A	0	-4.08	directed
cg25604883	STX1A	4.7e-05	-1.95	directed
cg25649038	STX1A	0.0057	-1.33	directed
cg25742540	STX1A	0.0012	-1.55	directed
cg25887236	STX1A	0	-4.26	directed
cg26385222	STX1A	1.9e-12	3.32	directed
cg27180880	STX1A	0.00094	-1.59	directed
cg00962781	TAOK1	0.0088	1.22	directed
cg02127209	TAOK1	0.00034	-1.72	directed
cg02571816	TAOK1	4.7e-12	-3.30	directed
cg02700894	TAOK1	0.0088	1.22	directed
cg05851240	TAOK1	1.2e-05	2.05	directed
cg13572782	TAOK1	0.0078	1.24	directed
cg16409049	TAOK1	2.9e-05	1.96	directed
cg16459364	TAOK1	0.0061	1.28	directed
cg17266581	TAOK1	0.002	-1.48	directed
cg20801056	TAOK1	3e-05	1.96	directed
cg21245652	TAOK1	0.00027	1.71	directed
cg22213386	TAOK1	3.2e-07	2.40	directed
cg22266001	TAOK1	6.8e-05	1.87	directed
cg22839075	TAOK1	0.0015	-1.53	directed
cg23483765	TAOK1	1.9e-07	-2.49	directed
cg25887236	TAOK1	6.2e-05	-1.92	directed
cg26385222	TAOK1	0.0095	1.21	directed
cg02571816	TBR1	0.0054	1.30	directed
cg00962781	TMEM144	1.3e-11	-3.23	directed
cg01657744	TMEM144	8e-05	1.85	directed
cg02127209	TMEM144	2.5e-10	2.98	directed
cg02202193	TMEM144	0.0072	-1.29	directed
cg02571816	TMEM144	0	5.70	directed
cg02589899	TMEM144	3.8e-08	2.59	directed
cg02700894	TMEM144	6.8e-09	-2.77	directed
cg03303359	TMEM144	0.0098	-1.25	directed
cg03537215	TMEM144	0.00033	1.68	directed
cg03593189	TMEM144	0.0072	1.25	directed
cg04920452	TMEM144	0.00012	-1.85	directed
cg05301866	TMEM144	3.4e-06	2.18	directed
cg05851240	TMEM144	0	-4.32	directed
cg08326075	TMEM144	0.0017	-1.51	directed

cg09515953	TMEM144	0.0059	-1.33	directed
cg11902728	TMEM144	0.0034	-1.41	directed
cg13572782	TMEM144	5.9e-08	-2.59	directed
cg15391531	TMEM144	5.8e-05	-1.93	directed
cg16409049	TMEM144	2.2e-16	-3.91	directed
cg16459364	TMEM144	2.5e-09	-2.85	directed
cg17266581	TMEM144	7.3e-11	3.07	directed
cg17860158	TMEM144	0.0017	-1.51	directed
cg18898125	TMEM144	6e-05	1.88	directed
cg19384775	TMEM144	3.1e-07	2.41	directed
cg20232102	TMEM144	0.00019	-1.79	directed
cg20801056	TMEM144	5.5e-14	-3.59	directed
cg21245652	TMEM144	2.7e-11	-3.18	directed
cg21322241	TMEM144	0.0021	1.44	directed
cg22213386	TMEM144	1.9e-12	-3.36	directed
cg22266001	TMEM144	8.9e-16	-3.83	directed
cg22557662	TMEM144	0.00036	1.67	directed
cg22839075	TMEM144	1.9e-08	2.65	directed
cg23241335	TMEM144	1.4e-06	2.27	directed
cg23483765	TMEM144	0	4.39	directed
cg25604883	TMEM144	0.0002	1.74	directed
cg25649038	TMEM144	0.0013	1.50	directed
cg25742540	TMEM144	9.6e-05	1.83	directed
cg25887236	TMEM144	2.2e-16	3.87	directed
cg26385222	TMEM144	2.9e-11	-3.17	directed
cg27180880	TMEM144	0.00023	1.72	directed
cg02127209	TOP2A	0.0065	1.27	directed
cg02571816	TOP2A	8.1e-06	2.10	directed
cg16409049	TOP2A	0.0097	-1.25	directed
cg21245652	TOP2A	0.005	-1.35	directed
cg22213386	TOP2A	0.0037	-1.40	directed
cg23483765	TOP2A	0.00042	1.65	directed
cg25887236	TOP2A	0.0026	1.41	directed
cg02571816	TROAP	3e-08	-2.65	directed
cg21245652	TROAP	0.0049	1.31	directed
cg23483765	TROAP	0.0034	-1.41	directed
cg02127209	TTK	0.0012	1.52	directed
cg02571816	TTK	1.2e-06	2.28	directed
cg02700894	TTK	0.0024	-1.46	directed
cg05851240	TTK	5.1e-05	-1.94	directed

cg13572782	TTK	0.0053	-1.34	directed
cg16409049	TTK	0.00044	-1.69	directed
cg17266581	TTK	0.0073	1.25	directed
cg20801056	TTK	0.00011	-1.85	directed
cg21245652	TTK	0.0003	-1.73	directed
cg22213386	TTK	0.0012	-1.56	directed
cg22266001	TTK	0.002	-1.48	directed
cg22839075	TTK	0.00098	1.54	directed
cg23483765	TTK	3.1e-05	1.95	directed
cg25887236	TTK	0.00017	1.76	directed
cg26385222	TTK	0.0045	-1.37	directed
cg02571816	VSNL1	0.0051	1.31	directed
cg00754896	cg01657744	0.0011	3.27	undirected
cg00962781	cg01657744	0.00023	4.13	undirected
cg01657744	cg02127209	0.00028	4.03	undirected
cg01657744	cg02202193	0.0083	1.95	undirected
cg01657744	cg02571816	0.00034	3.92	undirected
cg01657744	cg02589899	0.0014	3.13	undirected
cg01657744	cg03148084	0.00091	3.36	undirected
cg00754896	cg03303359	0.0011	3.27	undirected
cg00962781	cg03303359	0.00021	4.17	undirected
cg02127209	cg03303359	0.00021	4.17	undirected
cg02202193	cg03303359	0.0077	2.00	undirected
cg02571816	cg03303359	0.00032	3.96	undirected
cg02589899	cg03303359	0.0015	3.07	undirected
cg03148084	cg03303359	0.0011	3.27	undirected
cg01657744	cg03537215	0.00042	3.80	undirected
cg03303359	cg03537215	0.0005	3.70	undirected
cg01657744	cg03593189	0.0002	4.20	undirected
cg03303359	cg03593189	0.00018	4.26	undirected
cg01657744	cg03732545	0.0007	3.52	undirected
cg03303359	cg03732545	0.00097	3.33	undirected
cg01657744	cg04920452	0.00025	4.07	undirected
cg03303359	cg04920452	0.00024	4.10	undirected
cg01657744	cg05017199	8.2e-05	4.66	undirected
cg03303359	cg05017199	9.4e-05	4.59	undirected
cg01657744	cg05142617	0.0023	2.80	undirected
cg03303359	cg05142617	0.0025	2.75	undirected
cg01657744	cg05301866	0.00022	4.14	undirected
cg03303359	cg05301866	0.00012	4.45	undirected

cg01657744	cg05851240	0.00065	3.55	undirected
cg03303359	cg05851240	0.00056	3.64	undirected
cg01657744	cg06570930	0.00091	3.36	undirected
cg03303359	cg06570930	0.00082	3.42	undirected
cg01657744	cg08052629	0.00023	4.13	undirected
cg03303359	cg08052629	0.00023	4.12	undirected
cg01657744	cg08326075	0.0091	1.88	undirected
cg03303359	cg08326075	0.0067	2.10	undirected
cg01657744	cg09515953	0.00025	4.08	undirected
cg03303359	cg09515953	0.0002	4.20	undirected
cg01657744	cg11598403	0.0015	3.06	undirected
cg03303359	cg11598403	0.0018	2.96	undirected
cg01657744	cg11902728	0.00082	3.43	undirected
cg03303359	cg11902728	0.00079	3.44	undirected
cg01657744	cg13442966	0.0012	3.18	undirected
cg03303359	cg13442966	0.0012	3.21	undirected
cg01657744	cg13564825	0.00026	4.05	undirected
cg03303359	cg13564825	0.00023	4.13	undirected
cg01657744	cg13572782	0.0028	2.67	undirected
cg03303359	cg13572782	0.0037	2.49	undirected
cg01657744	cg14991769	8.6e-05	4.63	undirected
cg03303359	cg14991769	8.2e-05	4.66	undirected
cg01657744	cg15391531	0.00075	3.47	undirected
cg03303359	cg15391531	0.00091	3.36	undirected
cg01657744	cg16409049	0.0058	2.20	undirected
cg03303359	cg16409049	0.0041	2.44	undirected
cg01657744	cg16459364	0.00029	4.01	undirected
cg03303359	cg16459364	0.00026	4.05	undirected
cg01657744	cg17199181	0.0028	2.67	undirected
cg03303359	cg17199181	0.0027	2.71	undirected
cg01657744	cg17266581	0.0013	3.15	undirected
cg03303359	cg17266581	0.0017	2.99	undirected
cg01657744	cg17860158	0.00064	3.56	undirected
cg03303359	cg17860158	0.00086	3.40	undirected
cg01657744	cg18366748	0.00017	4.30	undirected
cg03303359	cg18366748	0.00014	4.39	undirected
cg01657744	cg18898125	0.00045	3.76	undirected
cg03303359	cg18898125	0.00047	3.74	undirected
cg01657744	cg19384775	0.00068	3.53	undirected
cg03303359	cg19384775	0.00094	3.34	undirected

cg01657744	cg19694850	0.0014	3.12	undirected
cg03303359	cg19694850	0.0015	3.08	undirected
cg01657744	cg19843036	8.8e-05	4.62	undirected
cg03303359	cg19843036	8.8e-05	4.62	undirected
cg01657744	cg19917714	0.00042	3.80	undirected
cg03303359	cg19917714	0.00051	3.70	undirected
cg01657744	cg20232102	0.00016	4.31	undirected
cg03303359	cg20232102	0.00015	4.35	undirected
cg01657744	cg20801056	8.4e-05	4.65	undirected
cg03303359	cg20801056	4.7e-05	4.93	undirected
cg01657744	cg21245652	0.00099	3.31	undirected
cg03303359	cg21245652	0.001	3.30	undirected
cg01657744	cg21322241	0.00098	3.32	undirected
cg03303359	cg21322241	0.00086	3.39	undirected
cg01657744	cg22213386	0.00019	4.22	undirected
cg03303359	cg22213386	0.00017	4.30	undirected
cg01657744	cg22266001	0.00058	3.62	undirected
cg03303359	cg22266001	0.00056	3.64	undirected
cg01657744	cg22557662	0.00035	3.90	undirected
cg03303359	cg22557662	0.0003	3.99	undirected
cg03303359	cg23241335	0.0096	1.85	undirected
cg01657744	cg23483765	0.0002	4.20	undirected
cg03303359	cg23483765	0.00019	4.24	undirected
cg01657744	cg23761616	0.00027	4.05	undirected
cg03303359	cg23761616	0.00024	4.12	undirected
cg01657744	cg23975646	0.0012	3.19	undirected
cg03303359	cg23975646	0.001	3.31	undirected
cg03303359	cg25742540	0.0084	1.95	undirected
cg01657744	cg25887236	0.00017	4.30	undirected
cg03303359	cg25887236	9.9e-05	4.57	undirected
cg01657744	cg26327804	0.00011	4.49	undirected
cg03303359	cg26327804	8.9e-05	4.62	undirected
cg01657744	cg26330518	0.00016	4.33	undirected
cg03303359	cg26330518	0.00014	4.40	undirected
cg01657744	cg26385222	0.00012	4.46	undirected
cg03303359	cg26385222	0.00011	4.49	undirected
cg01657744	cg27180880	7.6e-05	4.70	undirected
cg03303359	cg27180880	6.9e-05	4.75	undirected
cg01657744	cg27240008	0.0014	3.10	undirected
cg03303359	cg27240008	0.0015	3.07	undirected

AK5	AURKB	3.3e-05	-1.99	undirected
AK5	BUB1	1.2e-08	2.68	undirected
AURKB	BUB1	0.0037	1.36	undirected
AK5	C1QB	0.00014	1.79	undirected
AURKB	C1QB	8.2e-06	2.09	undirected
BUB1	C1QB	0.0015	-1.52	undirected
AK5	C1QC	5.4e-05	-1.94	undirected
AURKB	C1QC	9.5e-05	-1.87	undirected
BUB1	C1QC	1.2e-06	2.28	undirected
AK5	CARNS1	0.00041	-1.69	undirected
C1QC	CARNS1	0.0016	-1.52	undirected
AK5	CDK1	5.3e-07	2.36	undirected
AURKB	CDK1	1.9e-05	2.01	undirected
BUB1	CDK1	5.3e-07	-2.40	undirected
C1QB	CDK1	0.0002	-1.78	undirected
C1QC	CDK1	5.3e-08	2.56	undirected
CARNS1	CDK1	0.0051	1.31	undirected
AK5	CENPF	2.7e-07	-2.46	undirected
AURKB	CENPF	5e-05	-1.94	undirected
BUB1	CENPF	1.9e-06	2.24	undirected
C1QB	CENPF	0.0001	1.82	undirected
C1QC	CENPF	5.1e-05	-1.94	undirected
CDK1	CENPF	5e-06	2.14	undirected
AK5	CKAP2L	0.007	1.26	undirected
BUB1	CKAP2L	0.0014	-1.54	undirected
C1QB	CKAP2L	6.9e-06	-2.15	undirected
C1QC	CKAP2L	0.0083	1.23	undirected
CDK1	CKAP2L	0.00062	-1.64	undirected
CENPF	CKAP2L	0.0048	1.32	undirected
AK5	CNDP1	0.0011	-1.57	undirected
BUB1	CNDP1	0.0011	1.52	undirected
C1QB	CNDP1	0.00061	1.60	undirected
C1QC	CNDP1	0.0056	-1.33	undirected
CDK1	CNDP1	0.00029	1.70	undirected
CENPF	CNDP1	0.0036	-1.40	undirected
AK5	CRYM	4.6e-05	1.91	undirected
AURKB	CRYM	0.0026	1.41	undirected
BUB1	CRYM	0.0012	-1.55	undirected
C1QB	CRYM	0.00017	-1.80	undirected
C1QC	CRYM	0.00047	1.64	undirected

CDK1	CRYM	0.0005	-1.67	undirected
CENPF	CRYM	3.4e-05	1.95	undirected
CKAP2L	CRYM	0.0099	-1.24	undirected
AK5	DTL	0.0024	-1.46	undirected
CENPF	DTL	0.0091	-1.26	undirected
AK5	FCGR1A	2.5e-08	2.62	undirected
AURKB	FCGR1A	0.00071	1.58	undirected
BUB1	FCGR1A	4.7e-05	-1.95	undirected
C1QB	FCGR1A	4.2e-05	-1.96	undirected
C1QC	FCGR1A	2.2e-06	2.23	undirected
CDK1	FCGR1A	1.5e-05	-2.07	undirected
CENPF	FCGR1A	0.00016	1.77	undirected
CKAP2L	FCGR1A	0.0024	-1.46	undirected
CNDP1	FCGR1A	0.002	1.44	undirected
CRYM	FCGR1A	0.00052	-1.67	undirected
AK5	FCGR1B	2.7e-07	-2.46	undirected
AURKB	FCGR1B	0.0027	-1.45	undirected
BUB1	FCGR1B	6.8e-09	2.73	undirected
C1QB	FCGR1B	3e-07	2.41	undirected
C1QC	FCGR1B	0.0001	-1.86	undirected
CARNS1	FCGR1B	0.0004	-1.70	undirected
CDK1	FCGR1B	3.3e-08	2.60	undirected
CENPF	FCGR1B	1.5e-06	-2.30	undirected
CKAP2L	FCGR1B	0.01	1.20	undirected
CNDP1	FCGR1B	0.0058	-1.33	undirected
CRYM	FCGR1B	0.0057	1.29	undirected
FCGR1A	FCGR1B	0.0041	1.34	undirected
AK5	FCGR1C	0.0014	-1.53	undirected
BUB1	FCGR1C	0.0014	1.50	undirected
CDK1	FCGR1C	0.007	1.26	undirected
FCGR1B	FCGR1C	1.6e-06	-2.30	undirected
AK5	FOXM1	0.00028	-1.74	undirected
BUB1	FOXM1	0.00012	1.81	undirected
C1QB	FOXM1	0.0056	1.29	undirected
C1QC	FOXM1	0.0029	-1.43	undirected
CENPF	FOXM1	0.002	-1.48	undirected
CKAP2L	FOXM1	0.0064	1.27	undirected
FCGR1A	FOXM1	0.00035	1.67	undirected
FCGR1B	FOXM1	0.0087	-1.26	undirected
AK5	GDA	4.8e-07	2.36	undirected

AURKB	GDA	6.2e-06	2.12	undirected
BUB1	GDA	8.7e-06	-2.13	undirected
C1QB	GDA	5e-05	-1.94	undirected
C1QC	GDA	1.7e-06	2.25	undirected
CARNS1	GDA	0.00011	1.81	undirected
CDK1	GDA	3.6e-09	-2.82	undirected
CENPF	GDA	0.00012	1.80	undirected
CKAP2L	GDA	5.6e-05	-1.93	undirected
CNDP1	GDA	0.0023	1.42	undirected
CRYM	GDA	0.00015	-1.82	undirected
FCGR1A	GDA	8.8e-10	-2.93	undirected
FCGR1B	GDA	5.9e-11	3.08	undirected
FCGR1C	GDA	0.00022	1.73	undirected
FOXM1	GDA	0.00086	1.56	undirected
AK5	GJC2	1.1e-05	-2.10	undirected
AURKB	GJC2	0.0036	-1.40	undirected
BUB1	GJC2	4.6e-07	2.37	undirected
C1QB	GJC2	1.2e-06	2.28	undirected
C1QC	GJC2	0.00041	-1.70	undirected
CARNS1	GJC2	0.0047	-1.36	undirected
CDK1	GJC2	1.1e-07	2.49	undirected
CENPF	GJC2	5.3e-06	-2.18	undirected
FCGR1A	GJC2	0.00014	1.79	undirected
FCGR1B	GJC2	0.0018	-1.50	undirected
GDA	GJC2	3.4e-06	2.18	undirected
AK5	HOOK1	9.4e-06	2.08	undirected
AURKB	HOOK1	0.0089	1.22	undirected
BUB1	HOOK1	0.00088	-1.60	undirected
C1QB	HOOK1	0.00087	-1.60	undirected
C1QC	HOOK1	0.00056	1.61	undirected
CDK1	HOOK1	5e-06	-2.18	undirected
FCGR1A	HOOK1	0.0009	-1.60	undirected
FCGR1B	HOOK1	3.3e-05	1.95	undirected
FOXM1	HOOK1	0.0094	1.21	undirected
GDA	HOOK1	7e-05	-1.91	undirected
AK5	KCNK9	0.00072	1.58	undirected
AURKB	KCNK9	0.0042	1.34	undirected
BUB1	KCNK9	0.00054	-1.66	undirected
C1QC	KCNK9	0.00043	1.65	undirected
CDK1	KCNK9	0.00061	-1.64	undirected

CKAP2L	KCNK9	0.0077	-1.28	undirected
FCGR1A	KCNK9	0.0002	-1.78	undirected
FCGR1B	KCNK9	0.0066	1.27	undirected
GDA	KCNK9	0.00061	-1.65	undirected
AK5	KIF20A	0.00065	1.60	undirected
BUB1	KIF20A	0.0015	-1.52	undirected
C1QB	KIF20A	0.0098	-1.24	undirected
CDK1	KIF20A	0.00092	-1.59	undirected
FCGR1A	KIF20A	0.0014	-1.53	undirected
AK5	KIFC1	5.8e-05	-1.93	undirected
AURKB	KIFC1	0.0053	-1.34	undirected
BUB1	KIFC1	0.0011	1.52	undirected
C1QB	KIFC1	0.0019	1.45	undirected
C1QC	KIFC1	0.00052	-1.67	undirected
CARNS1	KIFC1	0.0092	-1.25	undirected
CDK1	KIFC1	1.4e-05	2.04	undirected
CENPF	KIFC1	5.2e-05	-1.94	undirected
FCGR1A	KIFC1	4.8e-05	1.91	undirected
FCGR1B	KIFC1	0.00039	-1.70	undirected
GDA	KIFC1	2.1e-07	2.44	undirected
GJC2	KIFC1	0.0061	-1.32	undirected
HOOK1	KIFC1	0.0026	1.41	undirected
KCNK9	KIFC1	0.00097	1.54	undirected
KIF20A	KIFC1	0.0062	1.28	undirected
AK5	LOC283999	5.8e-10	-2.96	undirected
AURKB	LOC283999	4.8e-07	-2.41	undirected
BUB1	LOC283999	1.6e-07	2.47	undirected
C1QB	LOC283999	5.4e-06	2.14	undirected
C1QC	LOC283999	1.3e-08	-2.72	undirected
CARNS1	LOC283999	5.3e-05	-1.94	undirected
CDK1	LOC283999	3.9e-08	2.58	undirected
CENPF	LOC283999	1.2e-08	-2.72	undirected
CKAP2L	LOC283999	6.6e-09	2.73	undirected
CNDP1	LOC283999	2.6e-05	-2.02	undirected
CRYM	LOC283999	0.00015	1.78	undirected
FCGR1A	LOC283999	9e-09	2.70	undirected
FCGR1B	LOC283999	6.6e-09	-2.77	undirected
FCGR1C	LOC283999	0.0033	-1.41	undirected
FOXM1	LOC283999	0.0031	-1.42	undirected
GDA	LOC283999	3.9e-07	2.38	undirected

GJC2	LOC283999	4.2e-08	-2.62	undirected
HOOK1	LOC283999	3.3e-06	2.18	undirected
KCNK9	LOC283999	4.4e-05	1.92	undirected
KIF20A	LOC283999	1.4e-05	2.04	undirected
KIFC1	LOC283999	1.4e-07	-2.52	undirected
FCGR1B	LOC285780	0.0057	1.29	undirected
LOC283999	LOC285780	0.0014	1.49	undirected
BUB1	LRP2	0.0097	1.21	undirected
LOC283999	LRP2	0.0098	-1.24	undirected
GDA	MBP	0.00064	1.60	undirected
LOC283999	MBP	0.0016	-1.52	undirected
AK5	MCM10	1.8e-05	-2.05	undirected
AURKB	MCM10	1.2e-05	-2.09	undirected
BUB1	MCM10	0.00075	1.58	undirected
C1QB	MCM10	0.0064	1.27	undirected
C1QC	MCM10	0.0005	-1.67	undirected
CARNS1	MCM10	0.0054	-1.34	undirected
CDK1	MCM10	1.5e-05	2.03	undirected
CENPF	MCM10	4.4e-07	-2.42	undirected
CKAP2L	MCM10	0.00021	1.74	undirected
CNDP1	MCM10	0.0016	-1.51	undirected
CRYM	MCM10	0.00045	1.64	undirected
DTL	MCM10	0.00025	-1.76	undirected
FCGR1A	MCM10	3.3e-06	2.18	undirected
FCGR1B	MCM10	6.9e-05	-1.91	undirected
FCGR1C	MCM10	0.0092	-1.26	undirected
GDA	MCM10	1.2e-08	2.68	undirected
GJC2	MCM10	1.6e-06	-2.30	undirected
HOOK1	MCM10	6.4e-05	1.88	undirected
KIF20A	MCM10	0.0015	1.49	undirected
KIFC1	MCM10	0.00072	-1.62	undirected
LOC283999	MCM10	1.2e-06	-2.32	undirected
LRP2	MCM10	0.0085	-1.27	undirected
AK5	MELK	6.1e-06	2.12	undirected
AURKB	MELK	0.0012	1.52	undirected
BUB1	MELK	3.7e-06	-2.21	undirected
C1QB	MELK	0.00015	-1.82	undirected
C1QC	MELK	2.1e-06	2.23	undirected
CARNS1	MELK	7.7e-05	1.85	undirected
CDK1	MELK	2.9e-07	-2.45	undirected

CENPF	MELK	2.9e-11	3.13	undirected
CKAP2L	MELK	0.0012	-1.56	undirected
FCGR1A	MELK	1.8e-07	-2.49	undirected
FCGR1B	MELK	0.00033	1.68	undirected
FCGR1C	MELK	0.00086	1.56	undirected
GDA	MELK	0.00064	-1.64	undirected
GJC2	MELK	0.002	1.44	undirected
HOOK1	MELK	0.00017	-1.81	undirected
KIFC1	MELK	6e-05	1.88	undirected
LOC283999	MELK	2.1e-07	2.44	undirected
MBP	MELK	0.0043	1.33	undirected
MCM10	MELK	0.00034	1.68	undirected
FCGR1A	MKI67	0.006	1.28	undirected
FCGR1B	MKI67	0.0068	-1.30	undirected
GDA	MKI67	0.0032	1.38	undirected
LOC283999	MKI67	0.0084	-1.27	undirected
CDK1	MOBP	0.0051	1.31	undirected
CENPF	MOBP	0.0084	-1.27	undirected
GDA	MOBP	0.0066	1.27	undirected
LOC283999	MOBP	0.0045	-1.37	undirected
LOC283999	MOG	0.0083	1.23	undirected
AK5	NCKAP1L	5.7e-05	1.89	undirected
AURKB	NCKAP1L	0.0067	1.26	undirected
BUB1	NCKAP1L	1.7e-05	-2.06	undirected
C1QB	NCKAP1L	8.9e-06	-2.13	undirected
C1QC	NCKAP1L	0.0079	1.24	undirected
CARNS1	NCKAP1L	0.004	1.34	undirected
CDK1	NCKAP1L	1.3e-05	-2.09	undirected
CENPF	NCKAP1L	0.0059	1.28	undirected
CRYM	NCKAP1L	0.0011	-1.56	undirected
FCGR1A	NCKAP1L	0.00016	-1.81	undirected
FOXM1	NCKAP1L	0.0022	1.43	undirected
GDA	NCKAP1L	7e-07	-2.37	undirected
GJC2	NCKAP1L	0.0049	1.31	undirected
HOOK1	NCKAP1L	0.0099	-1.24	undirected
KIFC1	NCKAP1L	0.0011	1.52	undirected
LOC283999	NCKAP1L	2.3e-09	2.81	undirected
MCM10	NCKAP1L	0.00078	1.57	undirected
MELK	NCKAP1L	6.7e-07	-2.38	undirected
AK5	NECAB1	3.1e-06	2.19	undirected

AURKB	NECAB1	1.2e-06	2.28	undirected
BUB1	NECAB1	1.6e-07	-2.51	undirected
C1QB	NECAB1	0.00011	-1.85	undirected
C1QC	NECAB1	1.4e-06	2.27	undirected
CARNS1	NECAB1	0.00026	1.71	undirected
CDK1	NECAB1	1.9e-10	-3.04	undirected
CENPF	NECAB1	2.3e-07	2.43	undirected
CKAP2L	NECAB1	7.2e-05	-1.90	undirected
CNDP1	NECAB1	8.1e-05	1.85	undirected
CRYM	NECAB1	4.1e-06	-2.20	undirected
DTL	NECAB1	0.0032	1.37	undirected
FCGR1A	NECAB1	4.2e-08	-2.62	undirected
FCGR1B	NECAB1	3.8e-11	3.11	undirected
FCGR1C	NECAB1	0.0042	1.34	undirected
FOXM1	NECAB1	1.6e-05	2.02	undirected
GDA	NECAB1	1.3e-07	-2.52	undirected
GJC2	NECAB1	1.5e-06	2.26	undirected
HOOK1	NECAB1	0.00013	-1.84	undirected
KCNK9	NECAB1	0.00042	-1.69	undirected
KIF20A	NECAB1	0.00077	-1.61	undirected
KIFC1	NECAB1	1.9e-07	2.45	undirected
LOC283999	NECAB1	1.4e-14	3.63	undirected
LOC285780	NECAB1	0.0095	-1.25	undirected
MBP	NECAB1	0.0058	1.29	undirected
MCM10	NECAB1	1.4e-08	2.67	undirected
MELK	NECAB1	1.7e-08	-2.70	undirected
MKI67	NECAB1	0.0053	1.30	undirected
NCKAP1L	NECAB1	1.9e-11	-3.20	undirected
AK5	NINJ2	1.4e-05	2.04	undirected
AURKB	NINJ2	0.00033	1.68	undirected
BUB1	NINJ2	9e-05	-1.88	undirected
C1QC	NINJ2	0.0071	1.25	undirected
CDK1	NINJ2	0.00031	-1.73	undirected
CKAP2L	NINJ2	0.0039	-1.39	undirected
CRYM	NINJ2	0.0067	-1.31	undirected
FCGR1A	NINJ2	0.0066	-1.31	undirected
FCGR1B	NINJ2	8.2e-05	1.85	undirected
GDA	NINJ2	0.00056	-1.66	undirected
KIFC1	NINJ2	0.0014	1.49	undirected
LOC283999	NINJ2	4e-05	1.93	undirected

MCM10	NINJ2	0.00018	1.76	undirected
MELK	NINJ2	0.00014	-1.82	undirected
NCKAP1L	NINJ2	0.0028	-1.44	undirected
NECAB1	NINJ2	0.00041	-1.69	undirected
BUB1	NRGN	0.0012	1.51	undirected
C1QB	NRGN	0.0059	1.29	undirected
CDK1	NRGN	0.0094	1.21	undirected
FCGR1A	NRGN	0.0057	1.29	undirected
LOC283999	NRGN	0.0015	-1.52	undirected
NECAB1	NRGN	0.00062	1.60	undirected
CDK1	NUSAP1	0.0093	-1.25	undirected
FCGR1B	NUSAP1	0.0052	1.30	undirected
GDA	NUSAP1	0.0033	-1.41	undirected
LOC283999	NUSAP1	0.0014	1.50	undirected
NECAB1	NUSAP1	0.006	-1.32	undirected
AK5	PACSIN1	0.0085	1.23	undirected
BUB1	PACSIN1	0.0025	-1.46	undirected
CENPF	PACSIN1	0.001	1.54	undirected
GDA	PACSIN1	0.0055	-1.34	undirected
KIFC1	PACSIN1	0.0096	1.21	undirected
LOC283999	PACSIN1	0.0035	1.37	undirected
MCM10	PACSIN1	0.0099	1.20	undirected
NECAB1	PACSIN1	0.00055	-1.66	undirected
LOC283999	PPP1R14A	0.0074	-1.29	undirected
AK5	SLC17A7	0.0048	1.32	undirected
BUB1	SLC17A7	0.00047	-1.68	undirected
C1QB	SLC17A7	0.0004	-1.70	undirected
C1QC	SLC17A7	0.0066	1.27	undirected
CDK1	SLC17A7	8.8e-05	-1.88	undirected
CENPF	SLC17A7	7.3e-05	1.86	undirected
FCGR1B	SLC17A7	0.0038	1.35	undirected
GDA	SLC17A7	0.0011	-1.57	undirected
HOOK1	SLC17A7	0.0034	-1.41	undirected
KIFC1	SLC17A7	0.0066	1.27	undirected
LOC283999	SLC17A7	0.00014	1.78	undirected
MCM10	SLC17A7	0.0024	1.42	undirected
NECAB1	SLC17A7	0.00079	-1.61	undirected
NINJ2	SLC17A7	0.0052	-1.35	undirected
CENPF	SLC31A2	0.0037	1.36	undirected
LOC283999	SLC31A2	0.0059	1.29	undirected

BUB1	SPOCK3	0.0086	1.23	undirected
CDK1	SPOCK3	0.0031	1.38	undirected
FCGR1A	SPOCK3	0.0043	1.33	undirected
LOC283999	SPOCK3	0.0076	-1.29	undirected
AK5	STX1A	0	-4.06	undirected
AURKB	STX1A	0	-4.50	undirected
BIRC5	STX1A	0.0048	-1.36	undirected
BUB1	STX1A	0	4.64	undirected
C1QB	STX1A	1.9e-10	3.00	undirected
C1QC	STX1A	2.1e-13	-3.50	undirected
CARNS1	STX1A	4.1e-08	-2.62	undirected
CDK1	STX1A	0	4.40	undirected
CENPF	STX1A	1.2e-13	-3.54	undirected
CKAP2L	STX1A	9.9e-14	3.51	undirected
CNDP1	STX1A	9.5e-09	-2.74	undirected
CREG2	STX1A	0.00051	-1.67	undirected
CRYM	STX1A	9.5e-15	3.65	undirected
DOCK2	STX1A	0.0005	-1.67	undirected
DTL	STX1A	2.2e-05	-2.03	undirected
ERMN	STX1A	0.0081	1.23	undirected
FAM64A	STX1A	0.00012	-1.85	undirected
FCGR1A	STX1A	0	4.36	undirected
FCGR1B	STX1A	0	-5.62	undirected
FCGR1C	STX1A	7.2e-08	-2.57	undirected
FOLH1	STX1A	0.0018	-1.50	undirected
FOXM1	STX1A	2e-07	-2.49	undirected
GABRA5	STX1A	1.5e-06	-2.30	undirected
GDA	STX1A	6.2e-12	3.24	undirected
GJC2	STX1A	1.4e-12	-3.38	undirected
GNG3	STX1A	0.0013	-1.55	undirected
HJURP	STX1A	0.0032	1.38	undirected
HOOK1	STX1A	4.5e-13	3.41	undirected
KCNK9	STX1A	6.1e-13	3.39	undirected
KIF20A	STX1A	2.3e-06	2.22	undirected
KIFC1	STX1A	0	-4.38	undirected
LOC283999	STX1A	0	-4.49	undirected
LOC285780	STX1A	0.0073	1.25	undirected
LRP2	STX1A	0.00056	-1.66	undirected
MAL	STX1A	0.00011	1.82	undirected
MAP7D2	STX1A	0.0016	-1.51	undirected

MBP	STX1A	4e-07	-2.42	undirected
MCM10	STX1A	0	-4.94	undirected
MELK	STX1A	0	4.27	undirected
MKI67	STX1A	5.3e-05	-1.94	undirected
MOBP	STX1A	4.9e-05	-1.95	undirected
MOG	STX1A	0.00033	1.68	undirected
NCKAP1L	STX1A	0	4.91	undirected
NECAB1	STX1A	0	3.98	undirected
NEFL	STX1A	0.00041	1.66	undirected
NEFM	STX1A	0.00018	-1.80	undirected
NINJ2	STX1A	3.2e-08	2.60	undirected
NIPAL4	STX1A	0.0022	1.43	undirected
NRGN	STX1A	5.3e-14	-3.59	undirected
NUSAP1	STX1A	1.3e-05	2.05	undirected
PACSIN1	STX1A	1.6e-05	2.03	undirected
PPP1R14A	STX1A	0.0012	-1.56	undirected
RC3H2	STX1A	0.0055	1.29	undirected
RGS4	STX1A	0.0076	-1.29	undirected
S1PR5	STX1A	0.00063	-1.64	undirected
SLC17A7	STX1A	3.2e-06	2.19	undirected
SLC31A2	STX1A	4.4e-05	1.92	undirected
SPOCK3	STX1A	0.0037	-1.40	undirected
AK5	TAOK1	9.9e-06	-2.12	undirected
AURKB	TAOK1	1.4e-05	-2.08	undirected
BUB1	TAOK1	2.6e-08	2.62	undirected
C1QB	TAOK1	3.2e-08	2.60	undirected
C1QC	TAOK1	7.4e-08	-2.57	undirected
CARNS1	TAOK1	0.0024	-1.46	undirected
CDK1	TAOK1	2.9e-07	2.41	undirected
CENPF	TAOK1	1.2e-09	-2.90	undirected
CKAP2L	TAOK1	0.0095	1.21	undirected
CRYM	TAOK1	0.0052	1.30	undirected
FCGR1A	TAOK1	4e-08	2.58	undirected
FCGR1B	TAOK1	1e-08	-2.74	undirected
FCGR1C	TAOK1	0.0005	-1.67	undirected
GDA	TAOK1	1.5e-07	2.47	undirected
GJC2	TAOK1	0.0011	-1.57	undirected
HOOK1	TAOK1	6.3e-05	1.88	undirected
KIF20A	TAOK1	0.00075	1.58	undirected
KIFC1	TAOK1	4.5e-05	-1.95	undirected

LOC283999	TAOK1	8.7e-12	-3.26	undirected
MCM10	TAOK1	3.7e-06	-2.22	undirected
MELK	TAOK1	4.8e-05	1.91	undirected
NCKAP1L	TAOK1	0.00027	1.70	undirected
NECAB1	TAOK1	4e-11	3.11	undirected
NINJ2	TAOK1	4.4e-06	2.16	undirected
SLC17A7	TAOK1	0.0087	1.22	undirected
STX1A	TAOK1	0	-4.13	undirected
STX1A	TBR1	1.6e-05	2.02	undirected
AK5	TMEM144	0	4.80	undirected
AURKB	TMEM144	4.4e-16	3.84	undirected
BIRC5	TMEM144	0.0073	1.25	undirected
BUB1	TMEM144	0	-4.56	undirected
C1QB	TMEM144	0	-4.01	undirected
C1QC	TMEM144	0	4.26	undirected
CARNS1	TMEM144	1.1e-12	3.36	undirected
CDK1	TMEM144	0	-4.90	undirected
CENPF	TMEM144	0	4.48	undirected
CKAP2L	TMEM144	7.2e-07	-2.37	undirected
CLCA4	TMEM144	0.0022	-1.47	undirected
CNDP1	TMEM144	1.2e-08	2.68	undirected
CNTN2	TMEM144	0.0017	-1.51	undirected
CREG2	TMEM144	0.0033	1.37	undirected
CRYM	TMEM144	8.4e-13	-3.41	undirected
DOCK2	TMEM144	0.00018	1.76	undirected
DTL	TMEM144	2.3e-08	2.63	undirected
ELOVL1	TMEM144	0.0033	-1.41	undirected
ERMN	TMEM144	2.7e-06	-2.25	undirected
FAM64A	TMEM144	7.3e-05	1.86	undirected
FCGR1A	TMEM144	0	-4.69	undirected
FCGR1B	TMEM144	0	5.89	undirected
FCGR1C	TMEM144	3.3e-12	3.28	undirected
FOXM1	TMEM144	1.2e-09	2.86	undirected
GDA	TMEM144	0	-4.02	undirected
GJC2	TMEM144	1.5e-08	2.66	undirected
HJURP	TMEM144	0.0095	-1.25	undirected
HOOK1	TMEM144	1.1e-08	-2.73	undirected
KCNK9	TMEM144	2.8e-13	-3.48	undirected
KIF20A	TMEM144	3.3e-09	-2.82	undirected
KIFC1	TMEM144	0	3.93	undirected

LOC283999	TMEM144	0	5.58	undirected
LOC285780	TMEM144	8.3e-07	-2.36	undirected
LRP2	TMEM144	0.0028	1.40	undirected
MAL	TMEM144	3e-06	-2.23	undirected
MBP	TMEM144	1.1e-10	3.04	undirected
MCM10	TMEM144	0	5.36	undirected
MELK	TMEM144	0	-4.23	undirected
MKI67	TMEM144	1.1e-05	2.06	undirected
MOBP	TMEM144	0.00074	1.58	undirected
MOG	TMEM144	8.2e-06	-2.13	undirected
NCKAP1L	TMEM144	0	-4.30	undirected
NECAB1	TMEM144	0	-4.99	undirected
NEFL	TMEM144	0.00037	-1.71	undirected
NEFM	TMEM144	0.0014	1.50	undirected
NINJ2	TMEM144	2.2e-14	-3.64	undirected
NIPAL4	TMEM144	6.1e-05	-1.92	undirected
NRGN	TMEM144	9.5e-05	1.83	undirected
NUSAP1	TMEM144	3.7e-06	-2.21	undirected
PACSIN1	TMEM144	5e-07	-2.40	undirected
PPP1R14A	TMEM144	0.00019	1.75	undirected
RC3H2	TMEM144	0.0096	-1.25	undirected
RGS4	TMEM144	0.00075	1.58	undirected
S1PR5	TMEM144	1.6e-05	2.02	undirected
SLC17A7	TMEM144	4.2e-06	-2.20	undirected
SLC31A2	TMEM144	1.7e-07	-2.50	undirected
SPOCK3	TMEM144	0.00024	1.72	undirected
STX1A	TMEM144	0	6.54	undirected
TAOK1	TMEM144	0	4.04	undirected
TBR1	TMEM144	0.0027	-1.44	undirected
STX1A	TMEM176A	0.00077	1.57	undirected
TMEM144	TMEM176A	0.0021	-1.48	undirected
AK5	TOP2A	0.0016	1.47	undirected
AURKB	TOP2A	0.0081	1.24	undirected
BUB1	TOP2A	0.0004	-1.70	undirected
C1QB	TOP2A	0.0051	-1.35	undirected
C1QC	TOP2A	0.0023	1.42	undirected
CDK1	TOP2A	0.0017	-1.51	undirected
CENPF	TOP2A	0.0022	1.43	undirected
FCGR1A	TOP2A	0.0012	-1.56	undirected
FCGR1B	TOP2A	0.0016	1.48	undirected

GDA	TOP2A	0.00033	-1.72	undirected
GJC2	TOP2A	0.0087	1.22	undirected
HOOK1	TOP2A	0.008	-1.28	undirected
KIFC1	TOP2A	0.0079	1.24	undirected
LOC283999	TOP2A	0.0011	1.53	undirected
MCM10	TOP2A	0.004	1.34	undirected
MELK	TOP2A	0.0019	-1.50	undirected
NECAB1	TOP2A	3.1e-05	-2.00	undirected
STX1A	TOP2A	6.2e-11	3.08	undirected
TAOK1	TOP2A	0.0032	1.38	undirected
TMEM144	TOP2A	1.4e-09	-2.89	undirected
STX1A	TPX2	0.0045	-1.37	undirected
TMEM144	TPX2	0.0074	1.25	undirected
TMEM144	TRIM59	0.0015	-1.53	undirected
AK5	TROAP	0.0011	-1.56	undirected
AURKB	TROAP	0.0022	-1.47	undirected
BUB1	TROAP	0.0067	1.27	undirected
C1QB	TROAP	0.0033	1.37	undirected
C1QC	TROAP	0.0021	-1.48	undirected
CDK1	TROAP	0.0038	1.35	undirected
FCGR1A	TROAP	0.004	1.34	undirected
FCGR1B	TROAP	0.00048	-1.68	undirected
GDA	TROAP	0.00085	1.56	undirected
GJC2	TROAP	0.0073	-1.29	undirected
KIF20A	TROAP	0.0083	1.23	undirected
KIFC1	TROAP	0.0008	-1.61	undirected
MCM10	TROAP	0.0031	-1.42	undirected
MELK	TROAP	0.002	1.44	undirected
NCKAP1L	TROAP	0.0096	1.21	undirected
NECAB1	TROAP	0.00018	1.75	undirected
NINJ2	TROAP	0.0038	1.35	undirected
STX1A	TROAP	5e-11	-3.14	undirected
TAOK1	TROAP	0.0017	-1.51	undirected
TMEM144	TROAP	4.8e-11	3.10	undirected
AK5	TTK	1.3e-05	2.05	undirected
AURKB	TTK	3.5e-05	1.94	undirected
BUB1	TTK	2.7e-05	-2.01	undirected
C1QB	TTK	0.0017	-1.51	undirected
C1QC	TTK	0.0026	1.41	undirected
CARNS1	TTK	0.0054	1.30	undirected

CDK1	TTK	1e-06	-2.34	undirected
CENPF	TTK	0.0042	1.34	undirected
CKAP2L	TTK	0.00056	-1.66	undirected
CNDP1	TTK	0.0035	1.36	undirected
CRYM	TTK	0.0002	-1.78	undirected
FCGR1A	TTK	7.1e-06	-2.15	undirected
FCGR1B	TTK	0.00024	1.72	undirected
FCGR1C	TTK	0.0081	1.23	undirected
FOXM1	TTK	0.0068	1.26	undirected
GDA	TTK	0.00094	-1.59	undirected
GJC2	TTK	0.0023	1.42	undirected
HOOK1	TTK	0.0034	-1.41	undirected
KCNK9	TTK	0.0018	-1.50	undirected
KIFC1	TTK	0.00075	1.58	undirected
LOC283999	TTK	1.3e-07	2.48	undirected
MCM10	TTK	3.9e-05	1.93	undirected
MELK	TTK	1.9e-06	-2.28	undirected
NECAB1	TTK	4.6e-05	-1.95	undirected
PACSIN1	TTK	0.0094	-1.25	undirected
STX1A	TTK	1.2e-08	2.68	undirected
TAOK1	TTK	3.4e-06	2.18	undirected
TMEM144	TTK	3.5e-11	-3.16	undirected
TOP2A	TTK	0.00053	-1.66	undirected
TROAP	TTK	0.00099	1.54	undirected
BUB1	VSNL1	0.003	-1.43	undirected
STX1A	VSNL1	0.001	1.54	undirected
TMEM144	VSNL1	0.00023	-1.77	undirected

B.3 GBM vs LGG

Table B.3: Glioblastoma iDINGO summary denoting the differential edges with the respective p-value, differential score and edge type.

Source node	Target node	p-value	differential score	interaction type
cg00765737	ANXA2P2	0	-7.34	directed
cg01644592	ANXA2P2	0.0064	1.54	directed
cg02072495	ANXA2P2	0	5.42	directed
cg02306630	ANXA2P2	0.00017	2.14	directed
cg02658690	ANXA2P2	2.6e-06	-2.76	directed
cg02988755	ANXA2P2	1.3e-07	3.02	directed

cg03148084	ANXA2P2	3.3e-13	-4.26	directed
cg03296248	ANXA2P2	1e-05	-2.59	directed
cg04771838	ANXA2P2	0	5.41	directed
cg05017199	ANXA2P2	0	-4.89	directed
cg05081953	ANXA2P2	6.1e-08	-3.18	directed
cg07251711	ANXA2P2	0.0011	1.85	directed
cg08382774	ANXA2P2	0	-5.31	directed
cg08441806	ANXA2P2	4.6e-14	4.33	directed
cg08959039	ANXA2P2	3.3e-05	2.37	directed
cg10171448	ANXA2P2	4.5e-11	3.78	directed
cg10908116	ANXA2P2	0.00016	-2.23	directed
cg11068946	ANXA2P2	1.5e-11	-3.95	directed
cg11096515	ANXA2P2	0.0011	-1.92	directed
cg11342452	ANXA2P2	0	5.97	directed
cg11598403	ANXA2P2	1.3e-15	-4.67	directed
cg12712747	ANXA2P2	8.9e-16	4.62	directed
cg12811033	ANXA2P2	4.6e-11	3.78	directed
cg13141061	ANXA2P2	1.5e-10	3.67	directed
cg13572782	ANXA2P2	0	5.81	directed
cg14397690	ANXA2P2	0	5.54	directed
cg14991769	ANXA2P2	1.8e-08	3.22	directed
cg15304699	ANXA2P2	8.4e-12	-4.00	directed
cg15379757	ANXA2P2	4.7e-06	2.61	directed
cg15391531	ANXA2P2	5e-05	2.31	directed
cg15441535	ANXA2P2	9.6e-07	2.80	directed
cg16872841	ANXA2P2	1.1e-08	3.27	directed
cg16891318	ANXA2P2	1.8e-06	-2.80	directed
cg19701540	ANXA2P2	2.7e-07	2.94	directed
cg20818806	ANXA2P2	2.2e-16	4.76	directed
cg20917241	ANXA2P2	8.4e-07	2.82	directed
cg21322241	ANXA2P2	1.2e-06	-2.85	directed
cg22232704	ANXA2P2	9.5e-10	-3.58	directed
cg22382805	ANXA2P2	0.00012	2.19	directed
cg23008352	ANXA2P2	1.8e-08	-3.30	directed
cg23125970	ANXA2P2	0	-4.93	directed
cg25887236	ANXA2P2	6.9e-10	3.54	directed
cg27180880	ANXA2P2	2e-09	-3.51	directed
cg27240008	ANXA2P2	1.7e-14	4.41	directed
cg27583690	ANXA2P2	0.001	-1.94	directed
cg00765737	ANXA2	0	6.70	directed

cg01303723	ANXA2	0.0047	-1.68	directed
cg02072495	ANXA2	0	-5.29	directed
cg02306630	ANXA2	0.0039	-1.71	directed
cg02988755	ANXA2	0.004	-1.70	directed
cg03148084	ANXA2	7.2e-05	2.26	directed
cg04771838	ANXA2	1.6e-14	-4.49	directed
cg05017199	ANXA2	0.00044	2.00	directed
cg05081953	ANXA2	5.3e-05	2.30	directed
cg05851240	ANXA2	0.0038	1.64	directed
cg07251711	ANXA2	3.9e-05	-2.42	directed
cg08382774	ANXA2	2.2e-16	4.72	directed
cg08959039	ANXA2	1.3e-05	-2.56	directed
cg10171448	ANXA2	0.0079	-1.58	directed
cg11068946	ANXA2	0.0061	1.55	directed
cg11096515	ANXA2	6.1e-06	2.58	directed
cg11342452	ANXA2	0	-5.25	directed
cg11598403	ANXA2	1.8e-14	4.40	directed
cg12712747	ANXA2	8.8e-12	-3.99	directed
cg13141061	ANXA2	2.1e-06	-2.78	directed
cg13572782	ANXA2	0	-5.36	directed
cg14397690	ANXA2	0	-5.53	directed
cg14991769	ANXA2	2.4e-05	-2.48	directed
cg15304699	ANXA2	0.00034	2.04	directed
cg15441535	ANXA2	0.0017	-1.85	directed
cg16891318	ANXA2	0.00024	2.09	directed
cg20818806	ANXA2	0.0089	-1.55	directed
cg20917241	ANXA2	6.1e-05	-2.36	directed
cg22232704	ANXA2	1.3e-10	3.68	directed
cg22382805	ANXA2	0.0079	-1.58	directed
cg23125970	ANXA2	1.1e-11	3.90	directed
cg25887236	ANXA2	0.002	-1.83	directed
cg27180880	ANXA2	1.5e-07	3.00	directed
cg02072495	ASPM	1.2e-11	3.89	directed
cg14397690	ASPM	2e-13	4.22	directed
cg23125970	ASPM	0.0069	-1.60	directed
cg14397690	BUB1	0.0095	-1.54	directed
cg00765737	C1QB	0.0037	1.65	directed
cg02072495	C1QB	0.0003	-2.13	directed
cg05851240	C1QB	0.00015	2.16	directed
cg07251711	C1QB	0.003	-1.76	directed

cg08382774	C1QB	0.0075	1.51	directed
cg08959039	C1QB	0.0013	-1.91	directed
cg11096515	C1QB	0.00065	1.94	directed
cg11598403	C1QB	0.00098	1.87	directed
cg13572782	C1QB	0.0046	-1.68	directed
cg14397690	C1QB	0	-5.45	directed
cg19701540	C1QB	0.0095	1.47	directed
cg00706570	C1QC	2.9e-06	-2.75	directed
cg00765737	C1QC	0.00026	2.08	directed
cg01303723	C1QC	0.0089	-1.55	directed
cg02072495	C1QC	0	-4.93	directed
cg05017199	C1QC	0.0014	-1.88	directed
cg05851240	C1QC	2.3e-07	2.96	directed
cg06548292	C1QC	2.8e-05	2.39	directed
cg07251711	C1QC	1.5e-06	-2.82	directed
cg08382774	C1QC	3.6e-06	2.65	directed
cg08959039	C1QC	1.6e-06	-2.81	directed
cg11096515	C1QC	1.4e-06	2.75	directed
cg11342452	C1QC	0.0091	-1.55	directed
cg11598403	C1QC	0.00024	2.09	directed
cg12712747	C1QC	0.005	-1.66	directed
cg12811033	C1QC	0.0001	-2.29	directed
cg13442966	C1QC	8.6e-06	-2.62	directed
cg13572782	C1QC	9.6e-09	-3.36	directed
cg14397690	C1QC	0	-5.05	directed
cg15304699	C1QC	0.00062	1.94	directed
cg15495463	C1QC	0.0008	-1.98	directed
cg16872841	C1QC	0.0021	-1.82	directed
cg17061340	C1QC	0.00034	2.04	directed
cg19701540	C1QC	2.1e-05	2.43	directed
cg23125970	C1QC	1.5e-14	4.42	directed
cg25682171	C1QC	4.4e-06	2.62	directed
cg27240008	C1QC	1.4e-07	3.01	directed
cg27583690	C1QC	0.0012	-1.91	directed
cg02072495	CCNA2	0	-4.93	directed
cg14397690	CCNA2	0	-4.88	directed
cg23125970	CCNA2	3.2e-08	3.17	directed
cg00706570	COL4A1	2e-10	-3.72	directed
cg00765737	COL4A1	5.4e-14	4.32	directed
cg01303723	COL4A1	0.00056	-2.04	directed

cg02072495	COL4A1	0	-5.25	directed
cg03296248	COL4A1	0.0017	-1.85	directed
cg04771838	COL4A1	0.00011	-2.28	directed
cg05851240	COL4A1	3e-11	3.81	directed
cg06548292	COL4A1	1.3e-07	3.02	directed
cg07251711	COL4A1	1.1e-07	-3.12	directed
cg08382774	COL4A1	8e-09	3.30	directed
cg08959039	COL4A1	5.4e-08	-3.19	directed
cg10171448	COL4A1	0.0048	1.60	directed
cg11096515	COL4A1	2.1e-08	3.21	directed
cg11342452	COL4A1	1.4e-05	-2.55	directed
cg11598403	COL4A1	3e-08	3.17	directed
cg12712747	COL4A1	0	-5.38	directed
cg12811033	COL4A1	1.5e-06	-2.82	directed
cg13141061	COL4A1	0.0046	-1.68	directed
cg13442966	COL4A1	1.2e-07	-3.10	directed
cg13572782	COL4A1	0	-4.93	directed
cg14397690	COL4A1	0	-5.42	directed
cg15304699	COL4A1	2.2e-05	2.42	directed
cg15495463	COL4A1	4.4e-05	-2.40	directed
cg16872841	COL4A1	0.00042	-2.08	directed
cg16891318	COL4A1	0.0028	1.69	directed
cg17061340	COL4A1	2.2e-08	3.20	directed
cg19701540	COL4A1	9e-09	3.29	directed
cg20801056	COL4A1	4.9e-05	-2.39	directed
cg20818806	COL4A1	2.8e-07	-3.02	directed
cg23125970	COL4A1	0	4.91	directed
cg25682171	COL4A1	1.6e-06	2.74	directed
cg27180880	COL4A1	0.0016	1.79	directed
cg27240008	COL4A1	3.5e-06	2.65	directed
cg27583690	COL4A1	3.8e-07	-2.98	directed
cg02072495	COL4A2	0	4.80	directed
cg14397690	COL4A2	0.00089	1.89	directed
cg00765737	FAM115C	0.00037	2.03	directed
cg01303723	FAM115C	0.003	-1.76	directed
cg02072495	FAM115C	0	-5.24	directed
cg04771838	FAM115C	0.0057	-1.64	directed
cg05851240	FAM115C	0.00047	1.99	directed
cg07251711	FAM115C	0.0032	-1.75	directed
cg08382774	FAM115C	4e-06	2.63	directed

cg08959039	FAM115C	7.8e-06	-2.63	directed
cg11096515	FAM115C	5.4e-06	2.60	directed
cg11342452	FAM115C	0.0062	-1.62	directed
cg11598403	FAM115C	0.00022	2.10	directed
cg13572782	FAM115C	6.8e-08	-3.16	directed
cg14397690	FAM115C	0	-5.41	directed
cg20801056	FAM115C	0.0017	-1.86	directed
cg23125970	FAM115C	0.0018	1.77	directed
cg25682171	FAM115C	0.003	1.68	directed
cg27240008	FAM115C	1.2e-07	3.03	directed
cg27583690	FAM115C	0.0012	-1.91	directed
cg02072495	KLK6	0.0017	1.78	directed
cg00765737	LOC154761	8.2e-05	2.24	directed
cg01303723	LOC154761	0.00029	-2.14	directed
cg02072495	LOC154761	0	-5.03	directed
cg04771838	LOC154761	0.0012	-1.91	directed
cg05851240	LOC154761	8.4e-06	2.54	directed
cg07251711	LOC154761	0.00085	-1.97	directed
cg08382774	LOC154761	2.4e-08	3.19	directed
cg08959039	LOC154761	3e-07	-3.00	directed
cg11096515	LOC154761	2e-07	2.98	directed
cg11342452	LOC154761	0.00076	-1.99	directed
cg11598403	LOC154761	9.7e-06	2.52	directed
cg12712747	LOC154761	0.0031	-1.75	directed
cg12778065	LOC154761	0.0015	-1.88	directed
cg12811033	LOC154761	0.0021	-1.82	directed
cg13572782	LOC154761	3.2e-10	-3.68	directed
cg14397690	LOC154761	0	-5.21	directed
cg15495463	LOC154761	0.0025	-1.79	directed
cg20801056	LOC154761	0.001	-1.94	directed
cg20917241	LOC154761	1.7e-05	-2.53	directed
cg23125970	LOC154761	0.00091	1.88	directed
cg25682171	LOC154761	0.00016	2.15	directed
cg26385222	LOC154761	0.0093	1.47	directed
cg27180880	LOC154761	0.0054	1.57	directed
cg27240008	LOC154761	3.2e-10	3.61	directed
cg27583690	LOC154761	7.2e-05	-2.34	directed
cg02072495	MOBP	9.1e-07	2.81	directed
cg02072495	PLP1	0.0061	1.55	directed
cg00706570	S100A8	5.2e-11	-3.84	directed

cg00765737	S100A8	6.8e-10	3.54	directed
cg01303723	S100A8	0.00011	-2.28	directed
cg02072495	S100A8	0	-5.09	directed
cg03296248	S100A8	0.003	-1.75	directed
cg04771838	S100A8	5.7e-05	-2.37	directed
cg05017199	S100A8	7.8e-05	-2.33	directed
cg05851240	S100A8	8.1e-09	3.30	directed
cg06548292	S100A8	3e-08	3.17	directed
cg07251711	S100A8	2.5e-07	-3.03	directed
cg08382774	S100A8	3.5e-07	2.91	directed
cg08441806	S100A8	0.0078	-1.58	directed
cg08959039	S100A8	2.1e-08	-3.28	directed
cg10171448	S100A8	0.0092	1.47	directed
cg11096515	S100A8	3.4e-09	3.39	directed
cg11342452	S100A8	1.3e-06	-2.84	directed
cg11598403	S100A8	1.9e-09	3.44	directed
cg12712747	S100A8	0.00014	-2.25	directed
cg12778065	S100A8	0.0019	-1.84	directed
cg12811033	S100A8	5.1e-05	-2.38	directed
cg13442966	S100A8	1.4e-06	-2.83	directed
cg13572782	S100A8	0	-4.89	directed
cg13852093	S100A8	0.0047	1.60	directed
cg14397690	S100A8	0	-5.23	directed
cg15304699	S100A8	2e-05	2.43	directed
cg15495463	S100A8	2.7e-05	-2.47	directed
cg16872841	S100A8	0.00065	-2.01	directed
cg17061340	S100A8	1.6e-11	3.87	directed
cg19005390	S100A8	0.0082	1.49	directed
cg19701540	S100A8	2.1e-06	2.71	directed
cg20801056	S100A8	2.2e-05	-2.50	directed
cg20818806	S100A8	3.7e-12	-4.06	directed
cg20917241	S100A8	0.0058	-1.64	directed
cg21655444	S100A8	0.0018	-1.85	directed
cg23125970	S100A8	2.2e-16	4.75	directed
cg25682171	S100A8	4.1e-07	2.90	directed
cg26385222	S100A8	0.0056	1.57	directed
cg27180880	S100A8	0.00095	1.88	directed
cg27240008	S100A8	2.3e-13	4.21	directed
cg27583690	S100A8	0.00049	-2.06	directed
cg00706570	S100A9	2.5e-10	-3.70	directed

cg00765737	S100A9	5.4e-14	4.32	directed
cg01303723	S100A9	0.001	-1.94	directed
cg02072495	S100A9	0	-5.32	directed
cg04771838	S100A9	5.2e-07	-2.95	directed
cg05017199	S100A9	0.002	-1.83	directed
cg05851240	S100A9	1.6e-07	3.00	directed
cg06548292	S100A9	1.7e-07	2.99	directed
cg07251711	S100A9	2.4e-07	-3.03	directed
cg08382774	S100A9	6.5e-13	4.13	directed
cg08441806	S100A9	0.0034	-1.73	directed
cg08959039	S100A9	5.9e-08	-3.18	directed
cg11068946	S100A9	0.0022	1.74	directed
cg11096515	S100A9	1.1e-07	3.04	directed
cg11342452	S100A9	2.2e-08	-3.28	directed
cg11598403	S100A9	3.3e-13	4.18	directed
cg12712747	S100A9	4.8e-08	-3.20	directed
cg12778065	S100A9	0.0036	-1.72	directed
cg12811033	S100A9	6.2e-05	-2.36	directed
cg13442966	S100A9	4.7e-07	-2.96	directed
cg13572782	S100A9	4.4e-16	-4.75	directed
cg14397690	S100A9	0	-5.39	directed
cg15304699	S100A9	2.1e-05	2.43	directed
cg15495463	S100A9	0.0001	-2.29	directed
cg16872841	S100A9	3.6e-05	-2.43	directed
cg17061340	S100A9	8.5e-08	3.07	directed
cg19701540	S100A9	2.2e-05	2.42	directed
cg20801056	S100A9	1.2e-05	-2.58	directed
cg20818806	S100A9	3.3e-08	-3.24	directed
cg20917241	S100A9	0.0027	-1.78	directed
cg22232704	S100A9	0.0068	1.53	directed
cg23125970	S100A9	0	4.86	directed
cg25682171	S100A9	1.2e-07	3.03	directed
cg26385222	S100A9	0.006	1.55	directed
cg27180880	S100A9	8.4e-06	2.54	directed
cg27240008	S100A9	2.4e-14	4.38	directed
cg27583690	S100A9	4.3e-05	-2.41	directed
cg02072495	TMEM125	1.9e-08	-3.29	directed
cg14397690	TMEM125	0.00091	-1.96	directed
cg03148084	cg06548292	0.0051	3.04	undirected
cg03148084	cg13141061	0.0045	3.07	undirected

cg03148084	cg15495463	0.0089	2.87	undirected
cg03148084	cg16891318	0.0097	2.85	undirected
ANXA2P2	ANXA2	0	-5.43	undirected
ANXA2	ASPM	1.2e-05	-2.57	undirected
ANXA2P2	BUB1	5.5e-08	-3.19	undirected
ANXA2	BUB1	5.2e-05	2.31	undirected
ANXA2P2	C1QA	2.2e-13	-4.29	undirected
ANXA2	C1QA	1.7e-05	2.45	undirected
ANXA2P2	C1QB	0	-5.36	undirected
ANXA2	C1QB	0	4.81	undirected
ANXA2P2	C1QC	0	-5.19	undirected
ANXA2	C1QC	0	5.28	undirected
ASPM	C1QC	1.2e-07	-3.11	undirected
C1QB	C1QC	1.4e-14	4.42	undirected
ANXA2P2	CARNS1	2e-11	-3.92	undirected
ANXA2	CARNS1	2.5e-09	3.42	undirected
ANXA2P2	CCNA2	1.2e-14	-4.51	undirected
ANXA2	CCNA2	0	4.97	undirected
C1QB	CCNA2	1.4e-13	4.25	undirected
C1QC	CCNA2	1.1e-15	4.60	undirected
ANXA2P2	CKAP2L	1.2e-10	3.70	undirected
ANXA2	CKAP2L	3.7e-07	-2.98	undirected
ANXA2P2	COL4A1	0	-5.71	undirected
ANXA2	COL4A1	0	5.66	undirected
ASPM	COL4A1	1.3e-05	-2.57	undirected
C1QB	COL4A1	0	5.35	undirected
C1QC	COL4A1	0	5.03	undirected
CCNA2	COL4A1	1.6e-14	4.41	undirected
ANXA2P2	COL4A2	0.001	-1.94	undirected
ANXA2	COL4A2	2.3e-06	-2.77	undirected
C1QB	COL4A2	0.0003	-2.13	undirected
C1QC	COL4A2	5.2e-11	-3.84	undirected
COL4A1	COL4A2	0	5.34	undirected
ANXA2P2	ERMN	3.6e-05	2.36	undirected
ANXA2P2	FAM115C	0	-5.01	undirected
ANXA2	FAM115C	0	5.05	undirected
ASPM	FAM115C	1.2e-07	-3.11	undirected
C1QB	FAM115C	3.7e-14	4.35	undirected
C1QC	FAM115C	0	4.99	undirected
CCNA2	FAM115C	1.7e-14	4.41	undirected

COL4A1	FAM115C	0	5.42	undirected
ANXA2	GABRG2	0.0001	-2.29	undirected
ANXA2P2	KIF14	1.7e-11	-3.94	undirected
ANXA2	KIF14	2.4e-09	3.42	undirected
ANXA2P2	KLK6	4.5e-13	4.16	undirected
ANXA2	KLK6	7.3e-10	-3.61	undirected
FAM115C	KLK6	0.0075	-1.59	undirected
ANXA2P2	LOC154761	0	-5.80	undirected
ANXA2	LOC154761	0	5.42	undirected
ASPM	LOC154761	2.2e-10	-3.71	undirected
C1QB	LOC154761	5.7e-14	4.32	undirected
C1QC	LOC154761	2.2e-16	4.76	undirected
CARNS1	LOC154761	0.0052	1.58	undirected
CCNA2	LOC154761	0	4.89	undirected
COL4A1	LOC154761	0	5.14	undirected
COL4A2	LOC154761	0	-5.63	undirected
FAM115C	LOC154761	3.3e-13	4.18	undirected
KLK6	LOC154761	0.0013	-1.89	undirected
ANXA2P2	MAG	0.0049	-1.67	undirected
ANXA2P2	MBP	5.3e-06	-2.68	undirected
ANXA2	MBP	0.0023	1.73	undirected
ANXA2	MOBP	0.0031	-1.75	undirected
C1QC	MOBP	0.0016	-1.87	undirected
COL4A1	MOBP	5.7e-05	-2.37	undirected
FAM115C	MOBP	1.3e-06	-2.84	undirected
LOC154761	MOBP	2.4e-09	-3.50	undirected
ANXA2P2	MOG	0	5.26	undirected
ANXA2	MOG	6.7e-12	-4.01	undirected
ANXA2P2	NKX6-2	2.5e-05	-2.48	undirected
ANXA2	NKX6-2	0.0005	1.98	undirected
ANXA2P2	PLP1	1.6e-09	3.46	undirected
ANXA2	PLP1	1.5e-06	-2.83	undirected
LOC154761	PLP1	0.0041	-1.70	undirected
ANXA2P2	S100A8	0	-5.59	undirected
ANXA2	S100A8	0	5.45	undirected
ASPM	S100A8	3.1e-11	-3.88	undirected
C1QB	S100A8	0	5.00	undirected
C1QC	S100A8	0	4.78	undirected
CARNS1	S100A8	0.0097	1.46	undirected
CCNA2	S100A8	0	5.05	undirected

COL4A1	S100A8	0	5.20	undirected
COL4A2	S100A8	0	-5.63	undirected
FAM115C	S100A8	0	5.26	undirected
KLK6	S100A8	3.3e-05	-2.44	undirected
LOC154761	S100A8	0	4.96	undirected
MOBP	S100A8	4.4e-15	-4.58	undirected
PLP1	S100A8	0.0067	-1.61	undirected
ANXA2P2	S100A9	0	-5.45	undirected
ANXA2	S100A9	0	5.48	undirected
ASPM	S100A9	2.1e-10	-3.72	undirected
C1QB	S100A9	0	5.51	undirected
C1QC	S100A9	0	5.13	undirected
CARNS1	S100A9	1.1e-05	2.51	undirected
CCNA2	S100A9	0	5.24	undirected
CKAP2L	S100A9	0.0077	-1.58	undirected
COL4A1	S100A9	0	5.41	undirected
COL4A2	S100A9	0	-6.46	undirected
FAM115C	S100A9	0	5.38	undirected
KIF14	S100A9	0.0048	1.60	undirected
KLK6	S100A9	2.6e-08	-3.26	undirected
LOC154761	S100A9	0	5.24	undirected
MOBP	S100A9	1.1e-15	-4.68	undirected
MOG	S100A9	0.0057	-1.64	undirected
PLP1	S100A9	0.00021	-2.19	undirected
S100A8	S100A9	0	5.55	undirected
ANXA2P2	TMEM125	3.8e-05	-2.42	undirected
ANXA2	TMEM125	4.6e-08	3.13	undirected
C1QC	TMEM125	0.00043	2.00	undirected
COL4A1	TMEM125	2.7e-07	2.94	undirected
FAM115C	TMEM125	1.5e-08	3.25	undirected
LOC154761	TMEM125	2.2e-11	3.84	undirected
S100A8	TMEM125	2.5e-14	4.38	undirected
S100A9	TMEM125	2.7e-15	4.54	undirected
ANXA2P2	TMEM176A	8.3e-14	-4.36	undirected
ANXA2	TMEM176A	6.4e-12	3.94	undirected
COL4A1	TMEM176A	0.0096	1.46	undirected
S100A9	TMEM176A	0.00082	1.90	undirected
ANXA2P2	TMEM176B	8.5e-11	3.72	undirected
ANXA2	TMEM176B	1.1e-08	-3.35	undirected

C

Causal algorithms evaluation metrics summary

C.1 Astrocytoma RNA

C.2 Astrocytoma CpG

C.3 Oligodendroglioma RNA

C.4 Oligodendroglioma CpG

C.5 GBM RNA

C.6 GBM CpG

Table C.1: Summary of the robustness and stability evaluation results regarding the application of the different causal discovery algorithms to astrocytoma RNA dataset

	PC	PC LiNGAM	FCI	GES	Direct LiNGAM	ANM
SHD	132.66 ± 8.4354	123.01 ± 8.2701	78.93 ± 9.9628	1079.94 ± 61.7625	2703.06 ± 92.061	85.45 ± 7.603
Precision	0.4854 ± 0.0379	0.425 ± 0.0445	0.6208 ± 0.0383	0.3075 ± 0.043	0.1722 ± 0.0265	0.5313 ± 0.0637
F1-score	0.4847 ± 0.0427	0.3943 ± 0.0418	0.591 ± 0.0407	0.2871 ± 0.037	0.1805 ± 0.0279	0.43 ± 0.0637
FDR	0.5146 ± 0.0379	0.575 ± 0.0445	0.3792 ± 0.0383	0.6925 ± 0.043	0.8278 ± 0.0265	0.4687 ± 0.0637
Recall	0.4854 ± 0.0539	0.3681 ± 0.0415	0.5652 ± 0.0497	0.2696 ± 0.0334	0.19 ± 0.0305	0.3647 ± 0.0689

Table C.2: Summary of the robustness and stability evaluation results regarding the application of the different causal discovery algorithms to astrocytoma CpG dataset

	PC	PC LiNGAM	FCI	GES	Direct LiNGAM	ANM
SHD	64.71 ± 5.294	63.72 ± 5.7929	68.28 ± 5.2591	448.85 ± 13.2325	808.73 ± 22.3502	54.3 ± 5.8075
Precision	0.543 ± 0.0365	0.4495 ± 0.0643	0.7356 ± 0.0417	0.0722 ± 0.0143	0.1854 ± 0.0238	0.4038 ± 0.0628
F-Score	0.5487 ± 0.0463	0.4143 ± 0.0562	0.6624 ± 0.0462	0.0673 ± 0.0134	0.1718 ± 0.0222	0.3801 ± 0.0691
FDR	0.457 ± 0.0365	0.5505 ± 0.0643	0.2644 ± 0.0417	0.9278 ± 0.0143	0.8146 ± 0.0238	0.5962 ± 0.0628
Recall	0.5585 ± 0.0735	0.3849 ± 0.052	0.6054 ± 0.0629	0.063 ± 0.0129	0.1604 ± 0.0219	0.3664 ± 0.0867

Table C.3: Summary of the robustness and stability evaluation results regarding the application of the different causal discovery algorithms to oligodendroglioma RNA dataset

	PC	PC LiNGAM	FCI	GES	Direct LiNGAM	ANM
SHD	134.25 ± 9.0525	131.54 ± 7.3408	110.05 ± 14.3121	1548.26 ± 84.464	4658.37 ± 196.3351	80.09 ± 7.31
Precision	0.3886 ± 0.0367	0.3223 ± 0.0486	0.5462 ± 0.0522	0.059 ± 0.0066	0.152 ± 0.0387	0.3401 ± 0.0676
F-score	0.4024 ± 0.0404	0.2969 ± 0.0449	0.5013 ± 0.048	0.0756 ± 0.0094	0.2138 ± 0.0545	0.3159 ± 0.0696
FDR	0.6114 ± 0.0367	0.6777 ± 0.0486	0.4538 ± 0.0522	0.941 ± 0.0066	0.848 ± 0.0387	0.6599 ± 0.0676
Recall	0.4193 ± 0.0535	0.2756 ± 0.0433	0.4644 ± 0.0504	0.1058 ± 0.0172	0.3602 ± 0.092	0.2983 ± 0.078

Table C.4: Summary of the robustness and stability evaluation results regarding the application of the different causal discovery algorithms to oligodendroglioma CpG dataset

	PC	PC LiNGAM	FCI	GES	Direct LiNGAM	ANM
SHD	68.8 ± 4.9686	68.63 ± 5.7624	67.15 ± 4.717	513.39 ± 13.4839	942.02 ± 33.3863	53.38 ± 5.4989
Precision	0.524 ± 0.0366	0.3785 ± 0.0537	0.5698 ± 0.04	0.0724 ± 0.0145	0.1697 ± 0.0224	0.4065 ± 0.0721
F-score	0.5358 ± 0.0425	0.3635 ± 0.05	0.5779 ± 0.0376	0.0657 ± 0.013	0.1599 ± 0.0201	0.3382 ± 0.0714
FDR	0.476 ± 0.0366	0.6215 ± 0.0537	0.4302 ± 0.04	0.9276 ± 0.0145	0.8303 ± 0.0224	0.5935 ± 0.0721
Recall	0.5508 ± 0.062	0.3504 ± 0.0493	0.5884 ± 0.0505	0.0603 ± 0.012	0.1516 ± 0.0199	0.2955 ± 0.0822

Table C.5: Summary of the robustness and stability evaluation results regarding the application of the different causal discovery algorithms to GBM RNA dataset

	PC	PC LiNGAM	FCI	GES	Direct LiNGAM	ANM
SHD	22.95 ± 3.0497	23.69 ± 3.5096	14.6 ± 2.937	178.01 ± 9.3176	221.16 ± 12.22	17.34 ± 2.8788
Precision	0.6188 ± 0.0566	0.5248 ± 0.0926	0.7605 ± 0.0638	0.11 ± 0.024	0.1988 ± 0.0375	0.4408 ± 0.1679
F-score	0.626 ± 0.0505	0.482 ± 0.0748	0.7044 ± 0.0528	0.1114 ± 0.0249	0.1962 ± 0.0359	0.3184 ± 0.1506
FDR	0.3812 ± 0.0566	0.4752 ± 0.0926	0.2395 ± 0.0638	0.89 ± 0.024	0.8012 ± 0.0375	0.5492 ± 0.1711
Recall	0.6364 ± 0.062	0.4488 ± 0.071	0.6589 ± 0.0621	0.1136 ± 0.0277	0.1948 ± 0.0373	0.2634 ± 0.1478

Table C.6: Summary of the robustness and stability evaluation results regarding the application of the different causal discovery algorithms to GBM CpG dataset

	PC	PC LiNGAM	FCI	GES	Direct LiNGAM	ANM
SHD	57.69 ± 4.2538	53.6 ± 5.162	32 ± 6.241	467.68 ± 15.2513	1812.56 ± 47.4928	41.92 ± 4.0544
Precision	0.7525 ± 0.0464	0.4692 ± 0.0645	0.8056 ± 0.0474	0.0786 ± 0.0164	0.1101 ± 0.0145	0.3089 ± 0.0718
F-score	0.6664 ± 0.0393	0.4116 ± 0.0549	0.7002 ± 0.0397	0.0659 ± 0.0136	0.1684 ± 0.0221	0.27 ± 0.0742
FDR	0.2475 ± 0.0464	0.5308 ± 0.0645	0.1944 ± 0.0474	0.9214 ± 0.0164	0.8899 ± 0.0145	0.6911 ± 0.0718
Recall	0.5991 ± 0.0423	0.3676 ± 0.0513	0.6201 ± 0.0417	0.0569 ± 0.012	0.3586 ± 0.0471	0.2455 ± 0.0828

D

Mapping of CpG sites to genes

D.1 Astrocytoma

Table D.1: Table providing the information regarding the mapping of CpG sites selected for astrocytoma, to the respective genes

Probe id	Gene
cg00017271	BIRC5
cg00017271	EPR1
cg00048832	GABRA1
cg00051179	GABRA1
cg00386007	WDFY4
cg00462525	TBR1
cg00765737	COL4A2
cg01132696	HLA-DPB1
cg02534163	ENPP2
cg02589899	MOG
cg03537215	CNTN2
cg03732545	CNTN2
cg04452959	ENPP2
cg04771838	COL4A2
cg04879832	CYBA
cg05022688	CLDN11
cg05096756	WDFY4
cg06023994	CLDN11
cg06954660	HLA-DOA
cg07366188	BIRC5
cg07366188	EPR1
cg07675031	SPI1

Continued on next page

Probe id	Gene
cg08843517	CYBA
cg09025501	CYBA
cg10070788	BIRC5
cg10070788	EPR1
cg10792225	KIFC1
cg11145160	CLDN11
cg11598403	MBP
cg12741994	CLDN11
cg13442966	MBP
cg13572782	MBP
cg13785221	ANXA2
cg13879776	CLDN11
cg14490686	HLA-DOA
cg15391531	MBP
cg15471953	SULT4A1
cg15982099	SPI1
cg16472060	ANXA2
cg16872841	COL4A2
cg17199181	MBP
cg17266581	MBP
cg17403520	KIFC1
cg17606183	HLA-DRA
cg17860158	CNTN2
cg19272238	BIRC5
cg19272238	EPR1
cg19384775	CNTN2
cg19843036	TMEM144
cg19917714	CNTN2
cg19990651	HLA-DPB1
cg20794824	CLDN11
cg20801056	MBP
cg20960405	SNCB
cg21322241	MBP
cg21925310	SULT4A1
cg21949512	TBR1
cg23732629	HLA-DRA
cg23761616	TMEM144
cg24476363	HLA-DOA
cg25228510	CDK1

Continued on next page

Probe id	Gene
cg25228510	CDC2
cg25887236	MBP
cg27180880	MBP
cg27240008	MBP

D.2 Oligodendroglioma

Table D.2: Table providing the information regarding the mapping of the CpG sites selected for oligodendroglioma to the respective genes

Probe id	Gene
cg00059015	HOOK1
cg00754896	LOC283999
cg00962781	TMEM144
cg01657744	LDB3
cg02127209	MAG
cg02202193	LOC285780
cg02571816	PPP1R14A
cg02589899	MOG
cg02700894	SYN2
cg03148084	CNDP1
cg03303359	LDB3
cg03537215	CNTN2
cg03593189	STX1A
cg03732545	CNTN2
cg04118306	NEFM
cg04920452	LOC283999
cg05017199	TMEM176B
cg05017199	TMEM176A
cg05142617	MAL
cg05301866	TBR1
cg05851240	CNDP1
cg06570930	PPP1R14A
cg08052629	PPP1R14A
cg08326075	SV2B
cg09515953	PPP1R14A
cg11598403	MBP
cg11902728	MAG

Continued on next page

Probe id	Gene
cg13442966	MBP
cg13564825	PPP1R14A
cg13572782	MBP
cg14991769	TMEM176B
cg14991769	TMEM176A
cg15391531	MBP
cg15873301	SYN2
cg16409049	AK5
cg16459364	NEFM
cg17078116	NEFM
cg17199181	MBP
cg17266581	MBP
cg17860158	CNTN2
cg18366748	TMEM144
cg18898125	NEFM
cg19384775	CNTN2
cg19694850	MAG
cg19843036	TMEM144
cg19917714	CNTN2
cg20232102	PPP1R14A
cg20585869	NEFM
cg20801056	MBP
cg21177558	PBK
cg21245652	MAL
cg21322241	MBP
cg22213386	NIPAL4
cg22266001	MAG
cg22557662	PPP1R14A
cg22839075	SYN2
cg23241335	AK5
cg23483765	NIPAL4
cg23761616	TMEM144
cg23975646	MBP
cg25604883	AK5
cg25649038	LOC285780
cg25742540	SV2B
cg25887236	MBP
cg26327804	KCNK9
cg26330518	NEFM

Continued on next page

Probe id	Gene
cg26385222	TMEM176B
cg26385222	TMEM176A
cg27180880	MBP
cg27240008	MBP

D.3 GBM

Table D.3: Table providing the information regarding the mapping of the CpG sites selected for GBM to the respective genes

Probe id	Gene
cg00706570	MBP
cg00765737	COL4A2
cg01303723	NKX6-2
cg01644592	MOG
cg02072495	ANXA2
cg02306630	NKX6-2
cg02658690	COL4A1
cg02988755	COL4A1
cg03147503	MOG
cg03148084	CNDP1
cg03296248	COL4A1
cg04771838	COL4A2
cg05017199	TMEM176B
cg05017199	TMEM176A
cg05081953	ERMN
cg05851240	CNDP1
cg06548292	MBP
cg06648277	NKX6-2
cg07251711	NKX6-2
cg08382774	NKX6-2
cg08441806	NKX6-2
cg08959039	COL4A2
cg10171448	NKX6-2
cg10908116	COL4A1
cg11068946	NKX6-2
cg11096515	COL4A2
cg11342452	NKX6-2

Continued on next page

Probe id	Gene
cg11598403	MBP
cg12712747	COL4A1
cg12778065	COL4A1
cg12811033	CNDP1
cg13141061	MBP
cg13442966	MBP
cg13572782	MBP
cg13852093	MBP
cg14397690	ANXA2
cg14991769	TMEM176B
cg14991769	TMEM176A
cg15304699	NKX6-2
cg15379757	COL4A1
cg15391531	MBP
cg15441535	CNDP1
cg15495463	MBP
cg16872841	COL4A2
cg16891318	CNDP1
cg17061340	MBP
cg17266581	MBP
cg19005390	COL4A1
cg19701540	NKX6-2
cg20801056	MBP
cg20818806	COL4A1
cg20917241	COL4A1
cg21322241	MBP
cg21563471	ERMN
cg21655444	MBP
cg22232704	COL4A2
cg22382805	COL4A2
cg23008352	COL4A1
cg23125970	COL4A1
cg25682171	COL4A1
cg25887236	MBP
cg26385222	TMEM176B
cg26385222	TMEM176A
cg27180880	MBP
cg27240008	MBP

Continued on next page

Probe id	Gene
cg27583690	NKX6-2

E

Enrichment analysis

E.1 Astrocytoma CpG sites GO

Table E.1: Gene Ontology Enrichment analysis results concerning the application of missMethyl functionality to Astrocytoma CpG sites

Term	P-value	CpG sites
mitotic metaphase plate congression	9.42e-05	cg07366188(BIRC5), cg10792225(KIFC1)
mitotic spindle assembly	1.58e-04	cg07366188(BIRC5), cg10792225(KIFC1)
metaphase plate congression	2.03e-04	cg07366188(BIRC5), cg10792225(KIFC1)
response to activity	2.40e-04	cg04771838(COL4A2), cg08843517(CYBA), cg16872841(COL4A2)
establishment of chromosome localization	2.81e-04	cg07366188(BIRC5), cg10792225(KIFC1)
positive regulation of interleukin-6 production	3.45e-04	cg08843517(CYBA), cg27180880(MBP)
chromosome localization	3.95e-04	cg07366188(BIRC5), cg10792225(KIFC1)
survivin complex	5.00e-04	cg07366188(BIRC5)
negative regulation of renal output by angiotensin	5.62e-04	cg08843517(CYBA)

Continued on next page

Term	P-value	CpG sites
negative regulation of glomerular filtration by angiotensin	5.62e-04	cg08843517(CYBA)
smooth muscle hypertrophy	5.62e-04	cg08843517(CYBA)
spindle assembly	6.70e-04	cg07366188(BIRC5), cg10792225(KIFC1)
mitotic spindle organization	6.82e-04	cg07366188(BIRC5), cg10792225(KIFC1)
interleukin-6 production	8.23e-04	cg08843517(CYBA), cg27180880(MBP)
regulation of interleukin-6 production	8.23e-04	cg08843517(CYBA), cg27180880(MBP)
protein-containing complex	8.38e-04	cg00048832(COL4A2), cg04771838(COL4A2), cg07366188(BIRC5), cg08843517(CYBA), cg10792225(KIFC1), cg16872841(COL4A2), cg27180880(MBP)
chromosome passenger complex	1.10e-03	cg07366188(BIRC5)
regulation of renal output by angiotensin	1.12e-03	cg08843517(CYBA)
positive regulation of mitotic cytokinesis	1.18e-03	cg07366188(BIRC5)
microtubule cytoskeleton organization involved in mitosis	1.21e-03	cg07366188(BIRC5), cg10792225(KIFC1)

E.2 GBM CpG sites GO

Table E.2: Gene Ontology Enrichment analysis results concerning the application of missMethyl functionality to GBM CpG sites

Term	P-value	CpG sites
myelin sheath	4.85e-06	cg00706570(ANXA2), cg02072495(MBP), cg05081953(ERMN), cg06548292(ANXA2), cg11598403(MBP), cg13141061(ERMN), cg13442966(ANXA2), cg13572782(MBP), cg13852093(ERMN), cg14397690(ANXA2), cg15391531(MBP), cg15495463(ERMN), cg17061340(ANXA2), cg20801056(MBP), cg21322241(ERMN), cg21655444(ANXA2), cg25887236(MBP), cg27180880(ERMN), cg27240008(ANXA2)
internode region of axon	5.42e-06	cg00706570(ANXA2), cg05081953(ERMN), cg06548292(ANXA2), cg11598403(MBP), cg13141061(ERMN), cg13442966(ANXA2), cg13572782(MBP), cg13852093(ERMN), cg15391531(MBP), cg15495463(ERMN), cg17061340(ANXA2), cg20801056(MBP), cg21322241(ERMN), cg21655444(ANXA2), cg25887236(MBP), cg27180880(ERMN), cg27240008(ANXA2)

Continued on next page

Term	P-value	CpG sites
collagen type IV trimer	1.20e-05	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term		P-value	CpG sites
network-forming trimer	collagen	1.46e-05	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
collagen network	1.46e-05	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
basement membrane collagen trimer	1.62e-05	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
basement membrane	3.87e-05	cg00765737(COL4A1), cg02072495(MBP), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg14397690(ANXA2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)
compact myelin	4.58e-05	cg00706570(ANXA2), cg02072495(MBP), cg06548292(ANXA2), cg11598403(MBP), cg13141061(ERMN), cg13442966(ANXA2), cg13572782(MBP), cg13852093(ERMN), cg14397690(ANXA2), cg15391531(MBP), cg15495463(ERMN), cg17061340(ANXA2), cg20801056(MBP), cg21322241(ERMN), cg21655444(ANXA2), cg25887236(MBP), cg27180880(ERMN), cg27240008(ANXA2)

Continued on next page

Term	P-value	CpG sites
collagen-activated tyrosine kinase receptor signaling pathway	4.70e-05	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
collagen-activated signaling pathway	7.51e-05	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
complex of collagen trimers	1.07e-04	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
negative regulation of multicellular organismal process	2.82e-04	cg00706570(ANXA2), cg00765737(COL4A1), cg01303723(ANXA2), cg02072495(MBP), cg02306630(NKX6-2), cg04771838(COL4A1), cg05017199(TMEM176B), cg06548292(ANXA2), cg07251711(MBP), cg08382774(NKX6-2), cg08441806(COL4A2), cg08959039(COL4A2), cg10171448(ANXA2), cg11068946(MBP), cg11096515(COL4A2), cg11342452(COL4A2), cg11598403(MBP), cg13141061(ERMN), cg13442966(ANXA2), cg13572782(MBP), cg13852093(ERMN), cg14397690(ANXA2), cg14991769(ANXA2), cg15304699(MBP), cg15391531(MBP), cg15495463(ERMN), cg16872841(COL4A2), cg17061340(ANXA2), cg19701540(MBP), cg20801056(MBP), cg21322241(ERMN), cg21655444(ANXA2), cg22232704(COL4A2), cg22382805(COL4A1), cg25887236(MBP), cg26385222(COL4A2), cg27180880(ERMN), cg27240008(ANXA2), cg27583690(MBP)

Continued on next page

Term	P-value	CpG sites
extracellular matrix structural constituent conferring tensile strength	5.96e-04	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
negative regulation of cell development	8.72e-04	cg00706570(ANXA2), cg01303723(ANXA2), cg02306630(NKX6-2), cg05017199(TMEM176B), cg06548292(ANXA2), cg07251711(MBP), cg08382774(NKX6-2), cg08441806(COL4A2), cg10171448(ANXA2), cg11068946(MBP), cg11342452(COL4A2), cg11598403(MBP), cg13141061(ERMN), cg13442966(ANXA2), cg13572782(MBP), cg13852093(ERMN), cg14991769(ANXA2), cg15304699(MBP), cg15391531(MBP), cg15495463(ERMN), cg17061340(ANXA2), cg19701540(MBP), cg20801056(MBP), cg21322241(ERMN), cg21655444(ANXA2), cg25887236(MBP), cg26385222(COL4A2), cg27180880(ERMN), cg27240008(ANXA2), cg27583690(MBP)
positive regulation of low-density lipoprotein particle receptor binding	9.57e-04	cg02072495(MBP), cg14397690(ANXA2)

Continued on next page

Term	P-value	CpG sites
extracellular matrix organization	1.18e-03	cg00765737(COL4A1), cg02072495(MBP), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg14397690(ANXA2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
external encapsulating structure organization	1.20e-03	cg00765737(COL4A1), cg02072495(MBP), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg14397690(ANXA2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
extracellular structure organization	1.20e-03	cg00765737(COL4A1), cg02072495(MBP), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg14397690(ANXA2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
regulation of multicellular organismal development	1.25e-03	cg00706570(ANXA2), cg00765737(COL4A1), cg01303723(ANXA2), cg02072495(MBP), cg02306630(NKX6-2), cg04771838(COL4A1), cg05017199(TMEM176B), cg06548292(ANXA2), cg07251711(MBP), cg08382774(NKX6-2), cg08441806(COL4A2), cg08959039(COL4A2), cg10171448(ANXA2), cg11068946(MBP), cg11096515(COL4A2), cg11342452(COL4A2), cg11598403(MBP), cg13141061(ERMN), cg13442966(ANXA2), cg13572782(MBP), cg13852093(ERMN), cg14397690(ANXA2), cg14991769(ANXA2), cg15304699(MBP), cg15391531(MBP), cg15495463(ERMN), cg16872841(COL4A2), cg17061340(ANXA2), cg19701540(MBP), cg20801056(MBP), cg21322241(ERMN), cg21655444(ANXA2), cg22232704(COL4A2), cg22382805(COL4A1), cg25887236(MBP), cg26385222(COL4A2), cg27180880(ERMN), cg27240008(ANXA2), cg27583690(MBP)

Continued on next page

Term	P-value	CpG sites
main axon	1.48e-03	cg00706570(ANXA2), cg05081953(ERMN), cg06548292(ANXA2), cg11598403(MBP), cg13141061(ERMN), cg13442966(ANXA2), cg13572782(MBP), cg13852093(ERMN), cg15391531(MBP), cg15495463(ERMN), cg17061340(ANXA2), cg20801056(MBP), cg21322241(ERMN), cg21655444(ANXA2), cg25887236(MBP), cg27180880(ERMN), cg27240008(ANXA2)

E.3 Oligodendroglioma CpG sites GO

Table E.3: Gene Ontology Enrichment analysis results concerning the application of missMethyl functionality to Oligodendroglioma CpG sites

Term	P-value	CpG sites
myelin sheath	4.70e-07	cg02127209(MAG), cg03537215(MAL), cg11902728(MBP), cg13572782(CNTN2), cg15391531(MAG), cg17266581(MAL), cg17860158(MBP), cg19384775(CNTN2), cg20801056(MAG), cg21245652(MAL), cg21322241(MBP), cg22266001(CNTN2), cg25887236(MAG), cg27180880(MAL)
compact myelin	8.71e-07	cg02127209(MAG), cg11902728(MBP), cg13572782(CNTN2), cg15391531(MAG), cg17266581(MAL), cg20801056(MAG), cg21245652(MAL), cg21322241(MBP), cg22266001(CNTN2), cg25887236(MAG), cg27180880(MAL)
central nervous system myelination	5.14e-06	cg02127209(MAG), cg03537215(MAL), cg11902728(MBP), cg17860158(MBP), cg19384775(CNTN2), cg21245652(MAL), cg22266001(CNTN2)

Continued on next page

Term	P-value	CpG sites
axon ensheathment in central nervous system	5.14e-06	cg02127209(MAG), cg03537215(MAL), cg11902728(MBP), cg17860158(MBP), cg19384775(CNTN2), cg21245652(MAL), cg22266001(CNTN2)
myelination	2.69e-05	cg02127209(MAG), cg03537215(MAL), cg11902728(MBP), cg13572782(CNTN2), cg15391531(MAG), cg17266581(MAL), cg17860158(MBP), cg19384775(CNTN2), cg20801056(MAG), cg21245652(MAL), cg21322241(MBP), cg22266001(CNTN2), cg25887236(MAG), cg27180880(MAL)
ensheathment of neurons	2.89e-05	cg02127209(MAG), cg03537215(MAL), cg11902728(MBP), cg13572782(CNTN2), cg15391531(MAG), cg17266581(MAL), cg17860158(MBP), cg19384775(CNTN2), cg20801056(MAG), cg21245652(MAL), cg21322241(MBP), cg22266001(CNTN2), cg25887236(MAG), cg27180880(MAL)
axon ensheathment	2.89e-05	cg02127209(MAG), cg03537215(MAL), cg11902728(MBP), cg13572782(CNTN2), cg15391531(MAG), cg17266581(MAL), cg17860158(MBP), cg19384775(CNTN2), cg20801056(MAG), cg21245652(MAL), cg21322241(MBP), cg22266001(CNTN2), cg25887236(MAG), cg27180880(MAL)
oligodendrocyte development	2.97e-05	cg02127209(MAG), cg03537215(MAL), cg11902728(MBP), cg17860158(MBP), cg19384775(CNTN2), cg21245652(MAL), cg22266001(CNTN2)
structural constituent of myelin sheath	5.88e-05	cg13572782(CNTN2), cg15391531(MAG), cg17266581(MAL), cg20801056(MAG), cg21245652(MAL), cg21322241(MBP), cg25887236(MAG), cg27180880(MAL)

Continued on next page

Term	P-value	CpG sites
protein localization to axon	7.60e-05	cg03537215(MAL), cg17860158(MBP), cg19384775(CNTN2), cg21245652(MAL)
Schmidt-Lanterman incisure	8.36e-05	cg02127209(MAG), cg11902728(MBP), cg21245652(MAL), cg22266001(CNTN2)
main axon	1.23e-04	cg02127209(MAG), cg03537215(MAL), cg11902728(MBP), cg13572782(CNTN2), cg15391531(MAG), cg17266581(MAL), cg17860158(MBP), cg19384775(CNTN2), cg20801056(MAG), cg21322241(MBP), cg22266001(CNTN2), cg25887236(MAG), cg27180880(MAL)
oligodendrocyte differentiation	4.08e-04	cg02127209(MAG), cg03537215(MAL), cg11902728(MBP), cg17860158(MBP), cg19384775(CNTN2), cg21245652(MAL), cg22266001(CNTN2)
glial cell development	4.53e-04	cg02127209(MAG), cg03537215(MAL), cg11902728(MBP), cg17860158(MBP), cg19384775(CNTN2), cg21245652(MAL), cg22266001(CNTN2)
synaptic vesicle membrane	5.63e-04	cg02700894(STX1A), cg03593189(SYN2), cg08326075(SV2B), cg22839075(STX1A), cg25742540(SYN2)
exocytic vesicle membrane	5.63e-04	cg02700894(STX1A), cg03593189(SYN2), cg08326075(SV2B), cg22839075(STX1A), cg25742540(SYN2)
neurotransmitter secretion	8.57e-04	cg02700894(STX1A), cg03593189(SYN2), cg08326075(SV2B), cg22839075(STX1A), cg25742540(SYN2)
signal release from synapse	8.57e-04	cg02700894(STX1A), cg03593189(SYN2), cg08326075(SV2B), cg22839075(STX1A), cg25742540(SYN2)
protein insertion into plasma membrane	8.67e-04	cg21245652(MAL)

Continued on next page

Term	P-value	CpG sites
hinge region between urothelial plaques of apical plasma membrane	8.67e-04	cg21245652(MAL)

E.4 Astrocytoma CpG sites KEGG pathways

Table E.4: KEGG pathways Enrichment analysis results concerning the application of missMethyl functionality to Astrocytoma CpG sites

Term	P-value	CpG sites
Apoptosis - multiple species	1.27e-02	cg07366188(BIRC5)
Nicotine addiction	2.11e-02	cg00048832(GABRA1)
Pathways in cancer	2.42e-02	cg04771838(COL4A2), cg07366188(BIRC5), cg16872841(COL4A2)
Taste transduction	2.50e-02	cg00048832(GABRA1)
Platinum drug resistance	2.70e-02	cg07366188(BIRC5)
Leishmaniasis	2.99e-02	cg08843517(CYBA)
Amoebiasis	4.08e-02	cg04771838(COL4A2), cg16872841(COL4A2)
Protein digestion and absorption	4.12e-02	cg04771838(COL4A2), cg16872841(COL4A2)
Colorectal cancer	4.37e-02	cg07366188(BIRC5)
Neutrophil extracellular trap formation	4.42e-02	cg08843517(CYBA)
Small cell lung cancer	4.48e-02	cg04771838(COL4A2), cg16872841(COL4A2)
NOD-like receptor signaling pathway	4.56e-02	cg08843517(CYBA)

Continued on next page

Term	P-value	CpG sites
Leukocyte transendothelial migration	4.62e-02	cg08843517(CYBA)
AGE-RAGE signaling pathway in diabetic complications	4.67e-02	cg04771838(COL4A2), cg16872841(COL4A2)
ECM-receptor interaction	4.71e-02	cg04771838(COL4A2), cg16872841(COL4A2)
GABAergic synapse	4.72e-02	cg00048832(GABRA1)

E.5 GBM CpG sites KEGG pathways

Table E.5: KEGG pathways Enrichment analysis results concerning the application of missMethyl functionality to GBM CpG sites

Term	P-value	CpG sites
Amoebiasis	1.57e-03	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
Small cell lung cancer	1.79e-03	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
AGE-RAGE signaling pathway in diabetic complications	1.92e-03	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
Protein digestion and absorption	1.94e-03	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
ECM-receptor interaction	1.95e-03	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
Relaxin signaling pathway	3.47e-03	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)
Histidine metabolism	9.15e-03	cg03148084(CNDP1), cg05851240(CNDP1), cg12811033(CNDP1), cg15441535(CNDP1), cg16891318(CNDP1)

Continued on next page

Term	P-value	CpG sites
Focal adhesion	1.01e-02	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
Cytoskeleton in muscle cells	1.20e-02	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)
beta-Alanine metabolism	1.71e-02	cg03148084(CNDP1), cg05851240(CNDP1), cg12811033(CNDP1), cg15441535(CNDP1), cg16891318(CNDP1)
Arginine and proline metabolism	1.89e-02	cg03148084(CNDP1), cg05851240(CNDP1), cg12811033(CNDP1), cg15441535(CNDP1), cg16891318(CNDP1)

Continued on next page

Term	P-value	CpG sites
Human papillomavirus infection	2.22e-02	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
PI3K-Akt signaling pathway	2.31e-02	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

Continued on next page

Term	P-value	CpG sites
Pathways in cancer	4.74e-02	cg00765737(COL4A1), cg02658690(COL4A2), cg02988755(COL4A1), cg03296248(COL4A2), cg04771838(COL4A1), cg08959039(COL4A2), cg10908116(COL4A1), cg11096515(COL4A2), cg12712747(COL4A1), cg12778065(COL4A2), cg15379757(COL4A1), cg16872841(COL4A2), cg19005390(COL4A1), cg20818806(COL4A2), cg20917241(COL4A1), cg22232704(COL4A2), cg22382805(COL4A1), cg23008352(COL4A2), cg23125970(COL4A1), cg25682171(COL4A2)

E.6 Oligodendroglioma CpG sites KEGG pathways

Table E.6: KEGG pathways Enrichment analysis results concerning the application of missMethyl functionality to Oligodendroglioma CpG sites

Term	P-value	CpG sites
Cell adhesion molecules	1.70e-02	cg02127209(MAG), cg03537215(CNTN2), cg11902728(MAG), cg17860158(CNTN2), cg19384775(MAG), cg22266001(CNTN2)
Histidine metabolism	1.90e-02	cg03148084(CNDP1), cg05851240(CNDP1)
Thiamine metabolism	1.94e-02	cg16409049(AK5), cg23241335(AK5), cg25604883(AK5)

Continued on next page

Term	P-value	CpG sites
SNARE interactions in vesicular transport	2.94e-02	cg03593189(STX1A)
beta-Alanine metabolism	3.42e-02	cg03148084(CNDP1), cg05851240(CNDP1)
Arginine and proline metabolism	4.35e-02	cg03148084(CNDP1), cg05851240(CNDP1)

E.7 Astrocytoma genes GO

Table E.7: Gene Ontology Enrichment analysis results concerning the application of GOATOOLS functionality to Astrocytoma genes

Term	P-value	CpG sites
antigen processing and presentation of exogenous peptide antigen via MHC class II	2.85e-03	HLA-DOA, HLA-DPA1, HLA-DPB1, CD74, HLA-DRA
positive regulation of calcium-mediated signaling	2.85e-03	SYK, GPR62, CCL3, CCL4, NEUROD2
peptide antigen assembly with MHC class II protein complex	3.39e-03	HLA-DPA1, HLA-DRA, HLA-DPB1, HLA-DOA
response to toxic substance	5.44e-03	NEFL, CCL3, GJC2, CCL4, CDK1, MBP
synapse pruning	8.75e-03	C1QA, C1QB, C1QC
mitotic cytokinesis	2.15e-02	KIF20A, NUSAP1, BIRC5, KIF23, AURKB
mitotic spindle midzone assembly	2.15e-02	AURKB, BIRC5, KIF23
collagen-activated tyrosine kinase receptor signaling pathway	2.15e-02	SYK, COL4A1, COL4A2
protein localization to kinetochore	2.15e-02	TTK, AURKB, CDK1

Continued on next page

Term	P-value	CpG sites
positive regulation of T cell activation	2.15e-02	HLA-DPA1, HLA-DRA, HLA-DPB1, HLA-DOA
positive regulation of immune response	2.48e-02	HLA-DPA1, HLA-DRA, HLA-DPB1, HLA-DOA
positive regulation of B cell differentiation	3.70e-02	SPI1, SYK, NCKAP1L
cell division	4.23e-02	TPX2, NCAPH, CENPF, CCNB2, CDK1, NCAPG, KIFC1, BUB1B, UBE2C
mitotic spindle assembly	4.90e-02	KIFC1, BIRC5, AURKB, TPX2
MHC class II protein complex	2.34e-05	HLA-DOA, HLA-DPA1, HLA-DPB1, CD74, HLA-DRA
complement component C1 complex	6.28e-05	C1QA, C1QB, C1QC
complement component C1q complex	6.28e-05	C1QA, C1QB, C1QC
myelin sheath	7.36e-05	MAG, GJC2, CNTN2, MBP, ERMN
spindle	1.01e-04	KIF20A, CENPF, TTK, BIRC5, KIF23, AURKB, BUB1B, TPX2
clathrin-coated endocytic vesicle membrane	2.81e-04	FCGR1A, VAMP8, HLA-DPA1, HLA-DPB1, CD74, HLA-DRA

E.8 GBM genes GO

Table E.8: Gene Ontology Enrichment analysis results concerning the application of GOATOOLS functionality to GBM genes

Term	P-value	CpG sites
synapse pruning	1.06e-03	C1QA, C1QB, C1QC
astrocyte development	5.81e-03	S100A9, PLP1, S100A8
neutrophil aggregation	5.81e-03	S100A9, 'S100A8

Continued on next page

Term	P-value	CpG sites
sequestering of zinc ion	5.81e-03	S100A9, S100A8
complement activation, classical pathway	2.32e-02	C1QA, C1QB, C1QC
peptidyl-cysteine S-nitrosylation	2.32e-02	S100A9, S100A8
negative regulation of dendritic cell differentiation	4.34e-02	TMEM176A, TMEM176B
axon ensheathment	4.34e-02	PLP1, MBP
complement component C1 complex	2.26e-06	C1QA, C1QB, C1QC
complement component C1q complex	2.26e-06	C1QA, C1QB, C1QC
collagen-containing extracellular matrix	2.29e-05	COL4A1, COL4A2, S100A8, S100A9, C1QA, C1QB, C1QC, ANXA2
myelin sheath	2.45e-05	MBP, PLP1, MAG, ERMN
postsynapse	2.89e-04	GABRG2, C1QA, C1QB, C1QC, GABRA1
calprotectin complex	5.76e-04	S100A9', S100A8
synapse	7.57e-04	GABRG2, C1QA, C1QB, C1QC, PLP1, GABRA1, MBP
compact myelin	2.30e-03	MBP, MAG
internode region of axon	2.30e-03	MBP, ERMN
myelin sheath adaxonal region	4.70e-03	MAG, ANXA2
collagen type IV trimer	4.70e-03	COL4A1, COL4A2
collagen trimer	1.16e-02	C1QA, C1QB, C1QC

E.9 Oligodendroglioma genes GO

Table E.9: Gene Ontology Enrichment analysis results concerning the application of GOATOOLS functionality to Oligodendroglioma genes

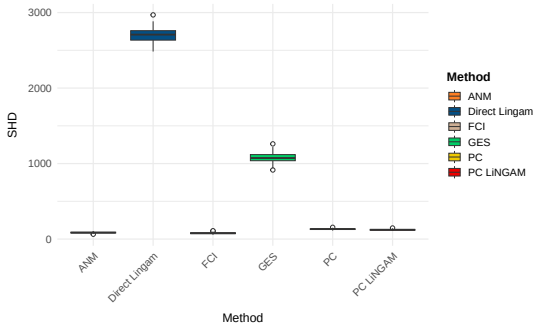
Term	P-value	CpG sites
myelin sheath	2.95e-04	MAG, GJC2, CNTN2, MBP, ERMN
spindle	3.54e-03	KIF20A, CENPF, TTK, BIRC5, AURKB, BUB1B, TPX2
kinetochore	5.35e-03	CENPF, TTK, BIRC5, BUB1, HJURP, AURKB, BUB1B
outer kinetochore	7.56e-03	BUB1, BUB1B, CENPF
paranode region of axon	8.04e-03	MAG, GJC2, ERMN
microtubule cytoskeleton	9.09e-03	CKAP2L, MAP7D2, BIRC5, TAOK1, GTSE1, AURKB, TPX2
complement component C1 complex	1.51e-02	C1QB, C1QC
complement component C1q complex	1.51e-02	C1QB, C1QC
synapse	1.51e-02	GABRB2, SYN2, SLC6A7, C1QB, C1QC, PACSIN1, CNTN2, GABRA1, MBP, GABRA5
GABA-A receptor complex	1.72e-02	GABRB2, GABRA1, GABRA5
chloride channel complex	1.72e-02	GABRB2, GABRA1, SLC17A7, GABRA5
postsynaptic intermediate filament cytoskeleton	1.72e-02	NEFL, NEFM
compact myelin	1.72e-02	MBP, MAG
internode region of axon	1.72e-02	MBP, ERMN
chromosome, centromeric region	1.89e-02	TOP2A, HJURP, BIRC5, CENPF
condensed chromosome	2.83e-02	MKI67, TOP2A, NCAPG

Continued on next page

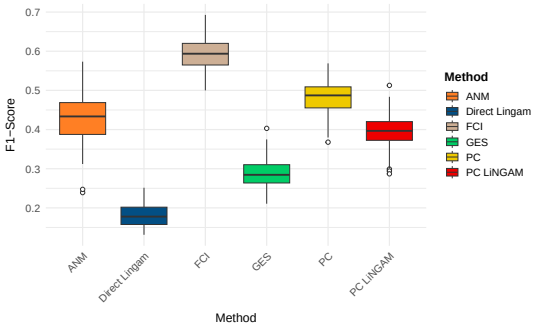
Term	P-value	CpG sites
chromosome passenger complex	3.52e-02	BIRC5, AURKB
structural constituent of myelin sheath	3.32e-02	MOBP, MBP, MAL
chloride channel activity	3.32e-02	GABRB2, SLC17A7, CLCA4, GABRA1, GABRA5
GABA receptor activity	3.32e-02	GABRB2, GABRA1

I

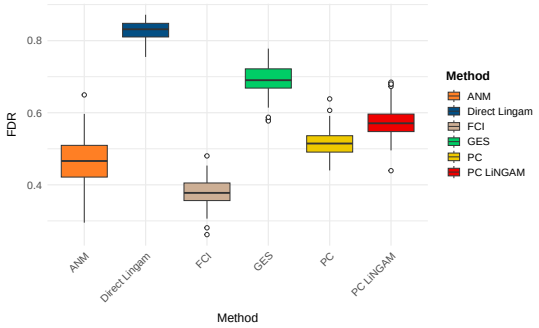
Causal algorithms robustness and stability evaluation



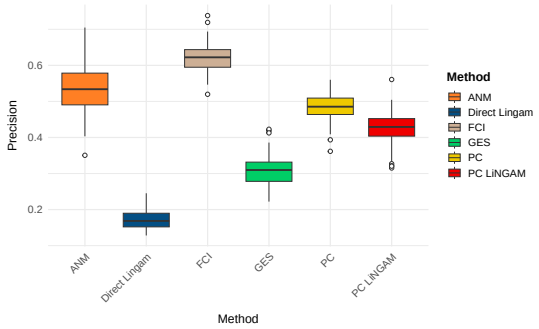
(a) SHD values from the causal discovery algorithms applied to astro RNA



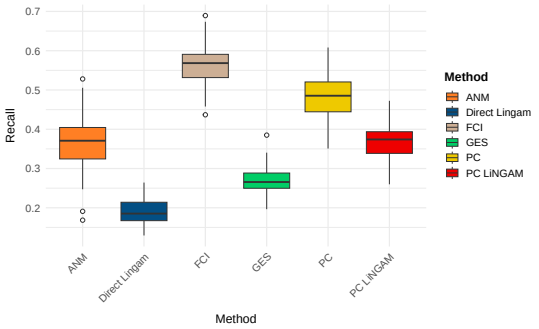
(b) F-score values from the causal discovery algorithms applied to astro RNA



(c) FDR values from the causal discovery algorithms applied to astro RNA

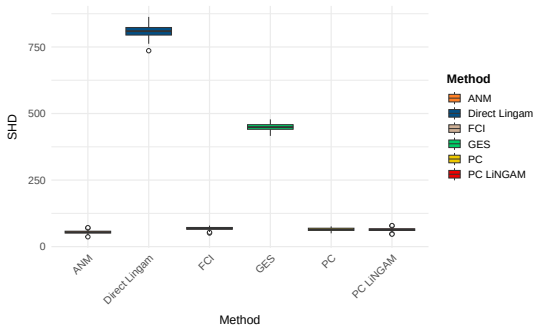


(d) Precision values from the causal discovery algorithms applied to astro RNA

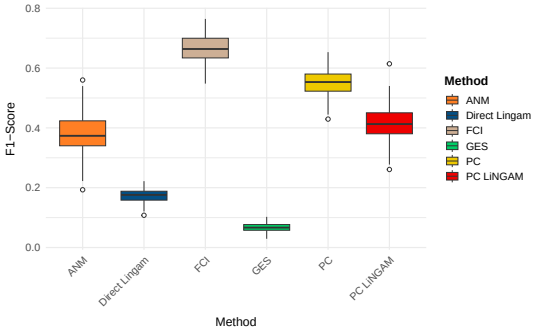


(e) Recall values from the causal discovery algorithms applied to astro RNA

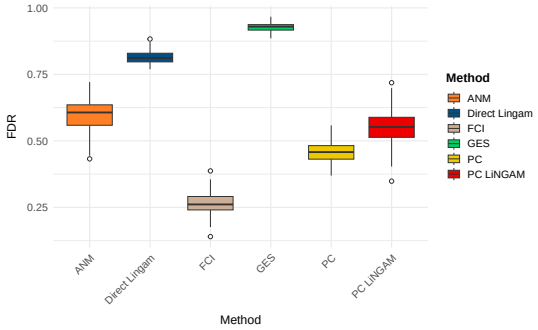
Figure I.1: Bar plots describing the results of the evaluation metrics, namely SHD, F-score, FDR, precision and recall to assess the robustness of the different causal discovery algorithms applied to the astrocytoma RNA dataset.



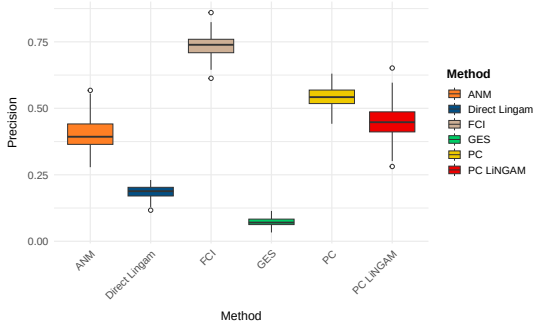
(a) SHD values from the causal discovery algorithms applied to astro CpG



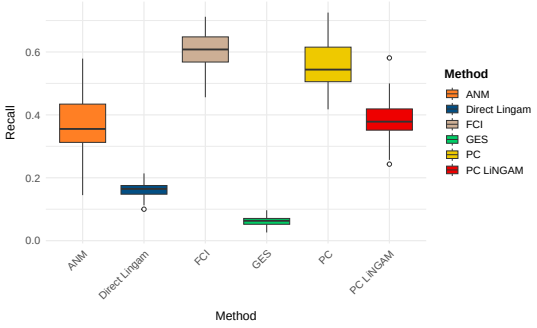
(b) F-score values from the causal discovery algorithms applied to astro CpG



(c) FDR values from the causal discovery algorithms applied to astro CpG

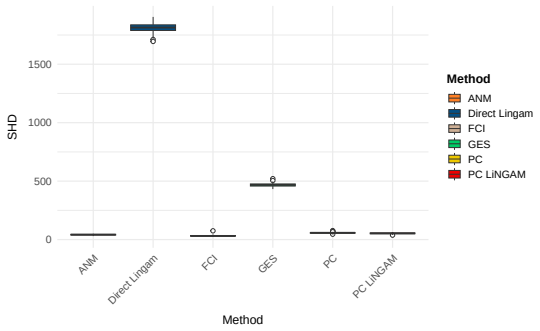


(d) Precision values from the causal discovery algorithms applied to astro CpG

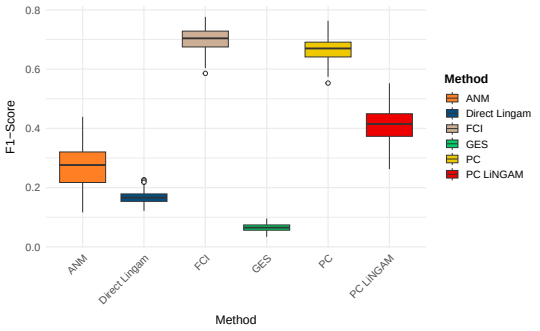


(e) Recall values from the causal discovery algorithms applied to astro CpG

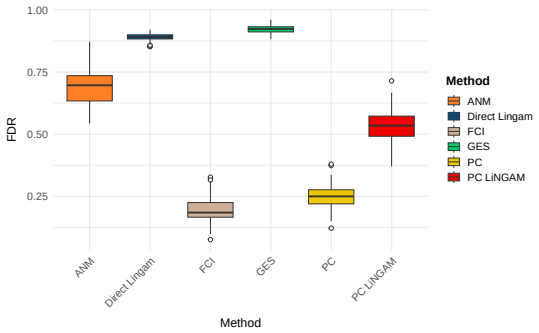
Figure I.2: Bar plots describing the results of the evaluation metrics, namely SHD, F-score, FDR, precision and recall to assess the robustness of the different causal discovery algorithms applied to the astrocytoma CpG dataset.



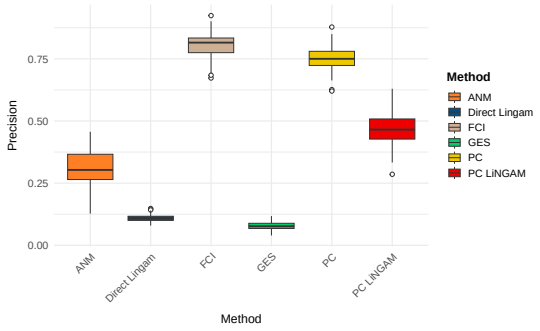
(a) SHD values from the causal discovery algorithms applied to GBM CpG



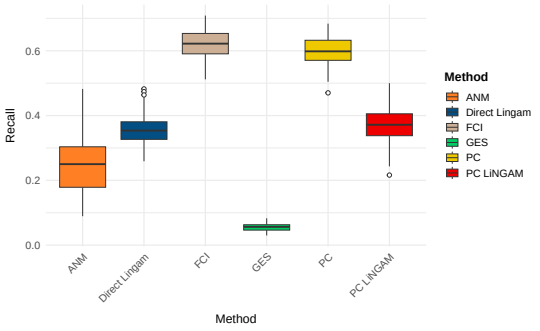
(b) F-score values from the causal discovery algorithms applied to GBM CpG



(c) FDR values from the causal discovery algorithms applied to GBM CpG

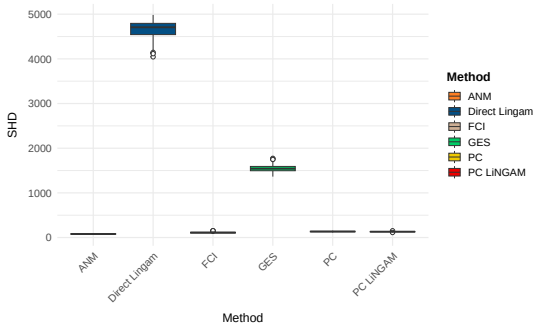


(d) Precision values from the causal discovery algorithms applied to GBM CpG

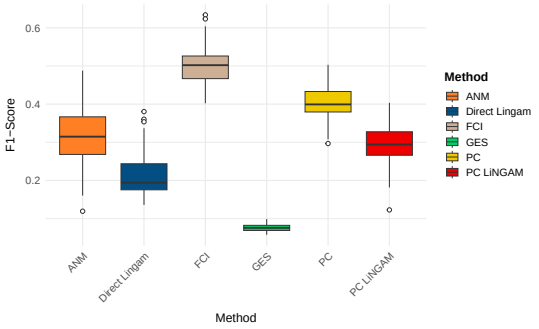


(e) Recall values from the causal discovery algorithms applied to GBM CpG

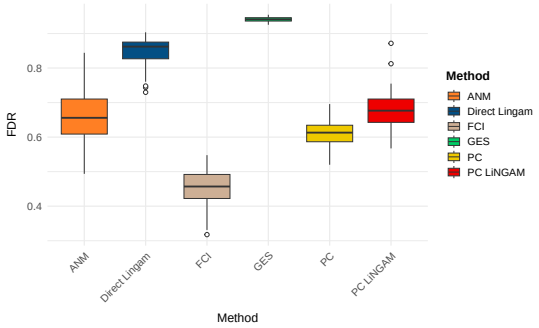
Figure I.3: Bar plots describing the results of the evaluation metrics, namely SHD, F-score, FDR, precision and recall to assess the robustness of the different causal discovery algorithms applied to the GBM CpG dataset.



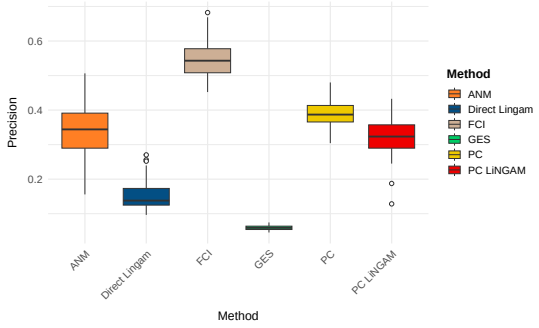
(a) SHD values from the causal discovery algorithms applied to oligo RNA



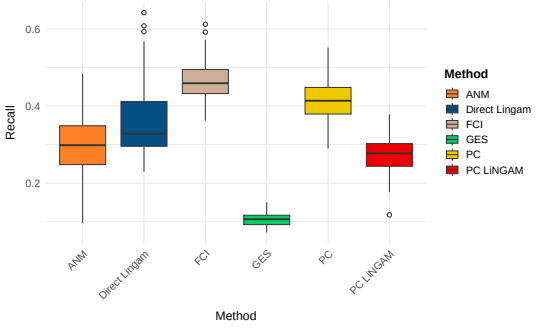
(b) F-score values from the causal discovery algorithms applied to oligo RNA



(c) FDR values from the causal discovery algorithms applied to oligo RNA

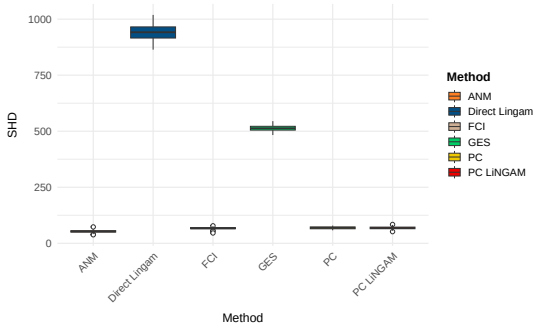


(d) Precision values from the causal discovery algorithms applied to oligo RNA

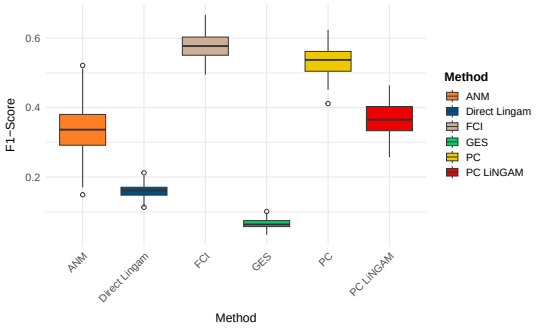


(e) Recall values from the causal discovery algorithms applied to oligo RNA

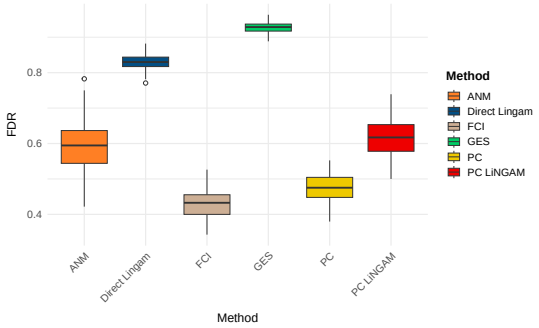
Figure I.4: Bar plots describing the results of the evaluation metrics, namely SHD, F-score, FDR, precision and recall to assess the robustness of the different causal discovery algorithms applied to the oligodendroglioma RNA dataset.



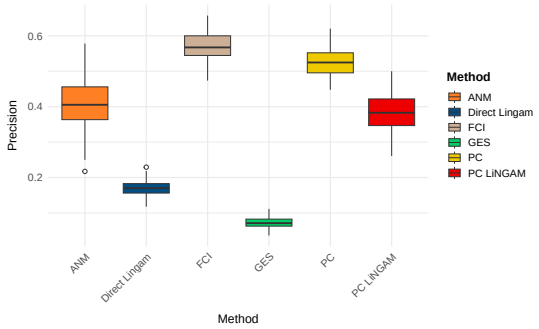
(a) SHD values from the causal discovery algorithms applied to oligo CpG



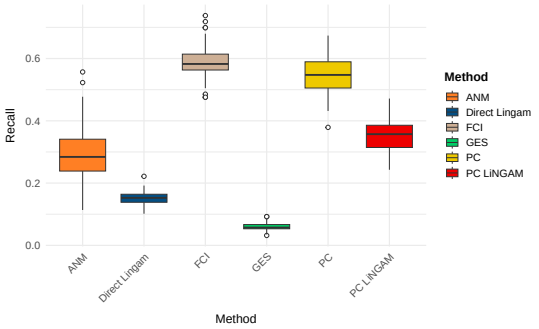
(b) F-score values from the causal discovery algorithms applied to oligo CpG



(c) FDR values from the causal discovery algorithms applied to oligo CpG



(d) Precision values from the causal discovery algorithms applied to oligo CpG



(e) Recall values from the causal discovery algorithms applied to oligo CpG

Figure I.5: Bar plots describing the results of the evaluation metrics, namely SHD, F-score, FDR, precision and recall to assess the robustness of the different causal discovery algorithms applied to the oligodendroglioma CpG dataset.

UNIVERSIDADE NOVA
DE LISBOA

2024

Causal insights from multi-omics data: uncovering gene networks and potential therapeutic targets in glioma

Pedro Filipe Albino Duarte

MASTER IN **COMPUTATIONAL BIOLOGY & BIOINFORMATICS**

SPECIALIZATION Multi-Omics for Life and Health Sciences

NOVA
UNIVERSIDADE NOVA DE LISBOA