Márcia Sofia Calado Grou

Licenciada em Biologia

A molecular and pathophysiological approach to tissue regeneration in the planarian model

Dissertação para a obtenção do Grau de Mestre em

Genética Molecular e Biomedicina

Orientador: Pedro Manuel Brôa Costa, Professor Auxiliar, FCT/UNL

Co-orientador: Carolina Madeira, Investigadora, FCT/UNL

Júri:

Presidente: Margarida Casal Ribeiro Castro Caldas Braga, Professora Auxiliar, FCT/UNL

Arguente: Ana Rita Fialho Grosso, Investigadora, FCT/UNL

Vogal: Pedro Manuel Brôa Costa, Professor Auxiliar, FCT/UNL

Novembro 2020

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Márcia Sofia Calado Grou

Licenciada em Biologia

A molecular and pathophysiological approach to tissue regeneration in the planarian model

Dissertação para a obtenção do Grau de Mestre em

Genética Molecular e Biomedicina

Orientador: Pedro Manuel Brôa Costa, Professor Auxiliar, FCT/UNL

Co-orientador: Carolina Madeira, Investigadora, FCT/UNL

Novembro 2020

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

A molecular and pathophysiological approach to tissue regeneration in the planarian model

Márcia Sofia Calado Grou, FCT/UNL

Agradecimentos

Aos meus orientadores, Pedro Costa e Carolina Madeira, começo por agradecer pela confiança que depositaram em mim, por todos os conselhos e ensinamentos que ambos me transmitiram e pelo constante incentivo em fazer sempre mais e melhor. Obrigada por todo o apoio, disponibilidade e orientação durante a realização desta tese, principalmente tendo este sido um ano tão difícil e com tantos constrangimentos.

A toda a equipa do SeaTox Lab, nomeadamente a Cátia, a Ana, a Mariaelena e a Carla por toda a ajuda e apoio que me deram e por me terem recebido e integrado tão bem num ambiente que tem tanto de tranquilo como de profissional. Sem dúvida que nunca esquecerei todos os momentos divertidos vividos convosco, principalmente nas saídas de campo e nos labmeetings.

À Leonor, a minha principal companheira durante esta etapa da minha vida académica, por toda a companhia e ajuda no laboratório e fora dele, por todas as conversas e desabafos e por todas as aventuras que já vivemos juntas e por aquelas que ainda estão para vir.

Aos meus amigos de infância, Mariana, Rita, Bia e Zé por me terem acompanhado desde sempre e por, mesmo após praticamente 20 anos, estarem sempre presentes quer nos bons quer nos maus momentos.

A todos os amigos que a faculdade me trouxe e com os quais vivi momentos fantásticos. Obrigada principalmente à Mafalda, ao André, às Carolinas e a tantos outros que de uma forma ou de outra estiveram presentes durante estes anos.

A todas as pessoas que conheci no mundo da ginástica, obrigada por me ajudarem a crescer e por me terem transmitido valores tão bons que vou guardar para sempre.

Ao meu namorado Duarte por ser uma pessoa fantástica que sempre me apoiou, acalmou e deu motivação. Obrigada também por todas as brincadeiras e picardias e principalmente por todo o amor e carinho que me dás a cada dia.

Por fim agradeço a toda a minha família, principalmente aos meus pais por serem um exemplo de trabalho e dedicação e por me terem tornado na pessoa que sou hoje. Obrigada do fundo do coração por todos os sacrifícios que fizeram por mim e por me terem ajudado a atingir todos os meus objetivos.

À minha irmã Catarina por ser capaz de me aturar e por estar sempre presente para alegrar os meus dias com as nossas brincadeiras e parvoíces. Continua com essa tua alegria e energia contagiantes que eu irei estar sempre ao teu lado para te acompanhar.

Abstract

Since early on, regeneration has been seen as a fascinating biological process of many living organisms. Planarians are especially suitable models to study the factors involved in this process given the extraordinary proliferative capacity of their adult pluripotent stem cells, the neoblasts. The critical understanding of how these cells work and the molecular networks and pathways underlying regeneration in these organisms are crucial to the development of potential therapeutic applications in the field of biomedical research. Regenerative medicine, for instance, is currently focused on studying the potential of stem cells to heal or replace missing tissues and organs. Thus, in order to ascertain the regenerative capacity of *Leptoplana* sp. and to evaluate the potential of neoblasts to replace lost and/or damaged cells, histological analyses were performed and allowed to conclude that regeneration is more restricted in this species and not always successful. However, neoblasts have proven their cell replacement ability and damage resistance, which demonstrates their importance and potential applications. Then, a second approach was carried out to perform gene expression analysis on *Schmidtea mediterranea* to identify the main genes and pathways involved in regeneration. Most of the genes were up-regulated in regenerating planarians when compared with the intact ones and within these, many were found to be involved in cell differentiation, and other biological processes, such as DNA replication and cell proliferation, as well as DNA damage detection and repair. This latter process has proven to be very important for the success of regeneration and demonstrated the potential use of planarians to study DNA damage. Taken together, these findings provide insights that this in silico approach using R-based bioinformatics enables a better perception of the processes behind regeneration which allow to explain and complement the anatomical observations of planarians and provide a global assessment of the entire regeneration process.

Keywords: planarians, regeneration, neoblasts, differentially-expressed genes (DEGs), cell proliferation, DNA repair.

Resumo

Há muito tempo que a regeneração é vista como um processo biológico fascinante em muitos organismos. As planárias, em particular, são consideradas um modelo bastante adequado para estudar os fatores envolvidos neste processo devido à extraordinária capacidade de proliferação das suas células estaminais pluripotentes adultas, os neoblastos. A compreensão do funcionamento destas células e das vias moleculares subjacentes à regeneração nestes organismos é crucial para o desenvolvimento de potenciais aplicações terapêuticas na área da investigação biomédica. A medicina regenerativa, por exemplo, encontra-se atualmente focada em estudar o potencial das células estaminais recuperarem ou substituírem tecidos e órgãos. Assim, de modo a determinar a capacidade regenerativa de *Leptoplana* sp. e avaliar o potencial dos neoblastos em substituírem células perdidas e/ou danificadas, foram realizadas análises histológicas que permitiram concluir que nesta espécie a regeneração é mais restrita e nem sempre bem-sucedida. Contudo, os neoblastos mostraram ter capacidade de substituição celular e de resistência a danos, o que demonstra a sua importância e potenciais aplicações. Neste trabalho, foi ainda realizada uma análise de expressão génica em *Schmidtea mediterranea* de modo a identificar os principais genes e vias envolvidos na regeneração. A maioria dos genes demonstrou ter níveis de expressão mais elevados em planárias em regeneração do que em planárias intactas, sendo que muitos deles se encontravam envolvidos em diferenciação celular e noutros processos, como a replicação de DNA e proliferação celular, bem como na deteção e reparação de danos no DNA. Este último processo provou ser bastante importante para o sucesso da regeneração e revelou um potencial uso das planárias como modelo para estudar e manipular danos no DNA. A conjugação destes resultados permitiu ter uma melhor perceção dos processos por detrás da regeneração que permitem complementar as observações anatómicas das planárias e obter uma avaliação global de todo o processo de regeneração.

Palavras-chave: planárias, regeneração, neoblastos, genes diferencialmente expressos, proliferação celular, reparação de DNA.

Index

Figure Index

Table Index

Abbreviation List

BER	Base-excision repair
BLAST	Basic Local Alignment Search Tool
bp	Base pair
COI-5P	Cytochrome c oxidase subunit I
CPDs	Cyclobutane pyrimidine dimers
DDR	DNA damage response
DEGs	Differentially-expressed genes
DNA	Deoxyribonucleic acid
dNTPs	Phosphated deoxynucleotides
DPX	Dibutylphthalate polystyrene xylene
DSBs	Double-strand breaks
FC	Fold change
FDR	False discovery rate
GEO	Gene Expression Omnibus
MCM	Mini chromosome maintenance
MEGA	Molecular Evolutionary Genetics Analysis
MQ	Milli-Q grade water (> 15 MΩ)
NCBI	National Center for Biotechnology Information
NER	Nucleotide-excision repair
nts	Nucleotides
o/n	Overnight
PCR	Polymerase chain reaction
ROS	Reactive oxygen species
SPIDR	Scaffolding protein involved in DNA repair
TAE	Tris-Acetate-EDTA
TEM	Transmission electron microscopy
UV	Ultraviolet
WIHC	Whole-mount immunohistochemistry
WISH	Whole-mount in situ hybridization
6-4PPs	Pyrimidine pyrimidone photoproducts

1. Introduction

Certain eumetazoans, such as *Hydra*, starfish and planarians, can regenerate their entire bodies from small fragments whereas more complex organisms lost this ability entirely. Such ability has been an extensively studied subject in the field of biomedical research (Cebrià et al. 2018; Mehta & Singh, 2019), due to its potential applications in treating severe injuries and degenerative diseases. Whereas the human capacity to regenerate is very limited, as in most complex animals, some vertebrates such as amphibians (e.g. salamanders and newts) and some fish are capable of efficiently regenerate full limbs (Stoick-Cooper et al. 2007). Yet, vertebrates are not able to regenerate complete body sections. As such, some invertebrate organisms, especially planarians, are considered the most suitable models. Even though there are huge structural, functional, and evolutionary differences between planarians and humans, they are both eumetazoans with well-defined body axis symmetries, three tissue layers (endoderm, ectoderm, and mesoderm), and multiple distinct organ systems arranged in specific patterns (Fields & Levin, 2020), conferring them many similarities.

1.1. Planarians as a model organism in regeneration research

Planarians are a group of aquatic flatworms (most of which free-living) of the Phylum Platyhelminthes that became an important biological model in diverse fields of biology. They may regenerate a whole new individual from a small fragment within a few days, an ability that has been explored for more than a century. In 1774, Peter Simon Pallas made the first observations of planarians and in the 1890s the interest in the regeneration of these metazoans emerged with the works of Thomas Hunt Morgan and Harriet Randolph (Randolph, 1897; Morgan, 1898; see also the review by Reddien & Alvarado, 2004). Recently, with the advance of science and technology and the development of new molecular tools, planarians strengthened their role as a model organism for studying regeneration, tissue renewal, stem cell biology, and DNA damage (Oviedo et al. 2008; Gentile et al. 2011). Many of these biological processes may have direct implications in human health and disease, so, by studying the biological pathways that control these processes in planarians, significant contributions can be made in different fields, for instance in regenerative medicine (Alvarado, 2004).

Regenerative medicine is focused on the development of therapies to heal or replace tissues and organs damaged or lost through disease or injury, as well as the treatment of some congenital defects. There are many therapies and approaches that are currently being explored, such as those involving stem cells and their differentiation capacity (see the recent review by Mao & Mooney, 2015). Stem cells are cells with the ability to divide continuously and to give rise to a large progeny of differentiated cells. Even though the mechanisms are not yet fully understood, stem cells are paramount to eumetazoans, including humans, since they enable the replacement of aged or damaged cells (Reddien et al. 2005). Besides this, these cells are extremely malleable and present a high degree of plasticity, which makes them an attractive therapeutic tool for self-cell repair and clinical applications by ex vivo manipulation (Caplan, 1991; Minguell et al. 2001). Altogether, its easy handling and maintenance in lab cultures and the presence of pluripotent stem cells called neoblasts, make these organisms an acknowledged in vivo

model to study regeneration processes and the underlying gene networks, such as the ones involved in DNA damage repair, which despite not being very studied is a key factor for regeneration (Gentile et al. 2011; Barghouth et al. 2019).

The regeneration process is molecularly and physiologically complex and it has been postulated that the high regeneration ability of planarians is due to the presence of neoblasts, which are the only known mitotically active cells in adult planarians that can differentiate into various cell types, including the germline (Orii et al. 2005; Oviedo et al. 2008). This proliferative capacity is essential for the initiation of the regeneration process since it allows the development of a specialized structure called the blastema, which consists of a mass of undifferentiated cells covered by epidermal cells (Reddien et al. 2005; Knakievicz & Ferreira, 2008). Besides this, the preservation of DNA is essential not only for regeneration, but also for the maintenance of homeostasis and well-being of the organism. However, there are several intrinsic and extrinsic factors that can damage DNA and cause mutations that can compromise cellular integrity and function if the damage is not repaired (Chatterjee & Walker, 2017; Barghouth et al. 2019).

1.2. The cascading impacts of environmental damaging agents on DNA

An example of an environmental agent that affects genome integrity is ultraviolet (UV) radiation, which is a type of electromagnetic radiation whose wavelength lies in between X-rays and visible light (10-400 nm). This short wavelength radiation is present in sunlight and can be subdivided into three main components, based on electrophysical properties: UV-A (315-400 nm), UV-B (280-315 nm), and UV-C (100-280 nm), the latter being the most energetic and therefore the most dangerous portion of the solar ultraviolet spectrum (Koronakis et al. 2002; Jackson & Bartek, 2009). Most UV-C radiation is absorbed and scattered by the Earth's atmosphere due to the presence of the stratospheric ozone layer. However, there is a portion of UV-A and UV-B radiation that is still capable of crossing (Koronakis et al. 2002), which brings many concerns, considering that in addition to being a major environmental concern, UV radiation is considered a "complete carcinogen" since it acts as a non-specific mutagenic agent, having characteristics of tumor initiator and promoter (D'Orazio et al. 2013). Therefore, one of the major targets of solar UV radiation is DNA, since nucleic acids easily absorb this type of radiation (particularly UV-B), which leads to the introduction of random mutations in the DNA due to the formation of various types of mutagenic and cytotoxic lesions (Sinha et al. 1996; Häder & Sinha, 2005). It is estimated that the UV component of sunlight can cause approximately 10^5 DNA lesions per cell per day such as cyclobutane pyrimidine dimers (CPDs), pyrimidine pyrimidone photoproducts (6-4PPs), and Dewar isomers (Häder & Sinha, 2005; Pattison & Davies, 2006; Hoeijmakers, 2009). If left unrepaired, these lesions will lead to the emergence of cytosine-thymine transitions (e.g. CC→TT and C→T) that can potentially compromise genome integrity and be transmitted to cellular progeny (Pattison & Davies, 2006; Lord & Ashworth, 2012). Besides these UV signature mutations caused by the direct UV absorption, there are also indirect mechanisms by which UV radiation can damage DNA due to the production of reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2), singlet oxygen ($^1O_2$),

and superoxide anion ($O_2^{\bullet-}$), which are toxic to cells due to its ability to undergo further reactions with cellular components (Häder & Sinha, 2005; Pattison & Davies, 2006; Ravanat & Douki, 2016).

In order to minimize or repair phototoxic DNA damage and to preserve genetic information and normal cellular functions, all living organisms affected by UV have developed highly conserved defense mechanisms, such as DNA damage response (DDR) and repair pathways, damage tolerance mechanisms, cell cycle checkpoints, and programmed cell death pathways (Häder & Sinha, 2005; Chatterjee & Walker, 2017; Barghouth et al. 2019). However, despite this wide range of mechanisms, some lesions are not repaired or endure erroneous repair. Thus, the impact of UV light on DNA and in several cellular processes that are dependent on its proper functioning influences the arising of a variety of health problems, including in humans, such as cancer and other "environmentally-influenced skin disorders", and age-related degenerative diseases (Koronakis et al. 2002; D'Orazio et al. 2013).

1.3. Aims and perspectives of this study

Owing to the relevance of regeneration in tissue repair and in other aspects associated with human health and biomedicine, the main focus of this thesis is to evaluate and describe the normal regeneration process in planarians, not only at a morphological level, but also at a molecular point of view. Therefore, this work will involve two components: the first will be focused on the observation of the structural changes that occur during regeneration, and the second will consist on an in silico transcriptomics approach to evaluate changes in expression levels and identify the main genes and pathways affected during this process. Then, I will combine the results obtained from these two methodologies to make a global assessment of the entire regeneration process. Moreover, I also intend to follow the differentiation stage of neoblasts over time in order to evaluate their regenerative capacity and their ability to resist and repair cellular damages since these stem cells have various DNA damage response and repair mechanisms. Finally, I also expect to identify possible correspondences between genes and processes described in planarians that also occur in humans that may have some relevance in future research, mainly those that concern DNA damage detect and repair.

Thus, the main research questions to be answered with this project can be summarized as follows:

- What are the main molecular and physiological changes that occur during regeneration in planarians? I hypothesize that during this process there is a massive cell proliferation followed by a cell differentiation step in order to re-establish the cells from the lost structures.
- Which genes are involved in the regeneration process and what is their role? Considering the cellular processes mentioned above, I expect that the genes expressed throughout regeneration are mainly involved in cell proliferation and cell differentiation, as well as in all the processes that precede and control them, with special focus on DNA damage detect and repair, given its importance for the success of regeneration.
- Considering the immensity of processes involved and all the control that is necessary during regeneration, which factors can affect this process and how? In addition to the environmental stressors that animals are subjected, such as temperature and/or salinity variations, etc., which

can modify their homeostasis and therefore affect a wide variety of cellular processes, there are other factors that can cause severe injuries, for instance in DNA, which, being the molecule of life, can compromise not only the success of regeneration but also the life of these organisms.

Thus, taking into account this last topic, another aim of this work was to understand and assess the impact of UV radiation in tissue regeneration, mainly at the level of DNA integrity, however, due to the COVID-19 pandemic situation this objective could not be fulfilled. Nonetheless, considering all the environmental and health concerns related to UV exposure, in the future, an exploration to evaluate the effects of UV radiation on the regeneration process should be done.

2. Materials and Methods

2.1. Animal collection, care, and maintenance

Planarian specimens (55 in total) were hand-collected at the end of October 2019 (n=18) and at the end of January 2020 (n=37) in an intertidal rocky shore environment, during low tide at the Avencas beach in Cascais, Portugal (38°41'17.1"N 9°21'36.5"W). Animals were carefully detached from rocks with a brush and tweezers and transferred to a plastic box containing local seawater with a salinity of 35, measured with a portable refractometer. Animals were then transported to SeaTox's lab facilities at Life Sciences Department, Nova School of Science and Technology.

After arrival at lab facilities, animals were firstly acclimated for 12h by dripping a mixture of artificial seawater with decreasing salinity from 35 to 30, to avoid osmotic shock while avoiding proliferation of marine microorganisms. Planarians were kept in an aquarium filled with approximately 7 liters of artificial seawater with constant aeration and recirculation, in shaded daylight conditions (10 light:14 dark photoperiod). Since planarians are mostly nocturnal organisms, the tank walls were covered with black plastic to prevent direct contact with light. The tank was enriched with natural pebbles from the Avencas beach to provide shelter and live prey (e.g. copepods, amphipods, and rotifers) for the planarians. Additionally, live rocks also help to stabilize water chemistry through the presence of high densities of nitrifying and denitrifying bacteria, increasing water quality. Artificial seawater was prepared using distilled water and purified sea salt. Water salinity was kept at 30, temperature at 18°C, by using a water refrigeration system (TK150 Chiller, from Teco), and pH at 8.1. Ammonia and nitrite levels were monitored using commercial colorimetric kits to restrain below 0.1 $mg.L^{-1}$ and 0.3 $mg.L^{-1}$, respectively. Planarian food was prepared by homogenizing fresh chicken liver with a Fisherbrand 150 Handheld Homogenizer Motor (Thermo Fisher Scientific) on ice-cold conditions, after removing all the visible veins and fat tissue with a scalpel. The liver homogenate obtained was then divided into 2 mL aliquots and frozen at -20°C. Planarians were fed once a week by adding approximately 1 mL of frozen chicken liver homogenate in the aquarium for about 30 minutes without water recirculation.

2.2. Species identification

The organisms were observed under an optical stereoscope to examine their anatomy and morphology and were identified according to Gibson & Knight-Jones (2017, pp. 135-141). *Discocelis tigrina* and *Leptoplana* sp. (*Leptoplana mediterranea* or *Leptoplana tremellaris*) were the species identified, wherein a total of 4 individuals of *D. tigrina* and 14 individuals of *Leptoplana* sp. were obtained in the first sampling time (October) and 2 individuals of *D. tigrina* and 34 individuals of *Leptoplana* sp. were obtained in the second sampling time (January), as well as one individual of an unidentified planarian species.

To confirm the identity of the morphospecies, a molecular analysis was employed by sequencing the *COI-5P* (cytochrome c oxidase subunit I) gene of *Leptoplana* sp. (n=2 individuals) and the unidentified species (n=1 individuals), according to the guidelines of the Barcode of Life described in

Lobo et al. (2013). No analyses were performed with *D. tigrina*, as all individuals were assigned to the regeneration assays described below.

Individuals were transferred to ≈100% ethanol to preserve DNA. Extraction was accomplished using the spin-column E.Z.N.A. Mollusc DNA Kit (Omega Bio-tek), following manufacturer's instructions with minor modifications. In brief, after the addition of BL Buffer and RNase, the samples were incubated at 65°C for 15 minutes and in the DNA elution step, the volume of Elution Buffer used was 50 µL which was left to incubate with the sample for 5 minutes at 65°C. After that, the total DNA yield, the 260/280 nm and 260/230 nm ratios were measured using a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific). The *COI-5P* sequence was then amplified by polymerase chain reaction (PCR). Each reaction mixture contained 5.96 µL H_2O MQ; 1 µL of PCR Buffer; 0.3 µL $MgCl_2$; 0.25 µL of each primer (retrieved from Lobo et al. 2013); 0.2 µL dNTPs and 0.04 µL Taq DNA Polymerase (Invitrogen). Each reaction contained 2 µL of DNA. Negative controls were included in the experiment. The thermal cycler (Biometra TOne, Analytik Jena AG) was programmed as shown in Table 2.1.

Table 2.1. Summary of PCR conditions for the amplification of the *COI-5P* gene, using primers Lobo F1 (5' KBTCHACAAAYCAYAARGAYATHGG 3') and Lobo R1 (5' TAAACYTCWGGRTGWCCRAARAAYCA 3').

Temperature	Time	Number of cycles
94°C	5 min	-
94°C	30 sec	5
45°C	1.5 min	
72°C	1 min	
94°C	30 sec	45
54°C	1.5 min	
72°C	1 min	
72°C	5 min	-

The PCR products were visualized on an 1.2 % m/v agarose gel prepared in 0.5% TAE (Tris-Acetate-EDTA) buffer, to which GelRed (Biotium) was added. The 100 bp DNA Ladder (Invitrogen) was used for reference. Electrophoresis was performed at 80 V and 400 mA for 60 minutes. The gel was imaged with trans-UV light on a GelDoc 2000 equipment (Bio-Rad). After confirmation of the presence of the *COI-5P* gene amplicon (≈650 bp), PCR products were purified and bidirectionally sequenced by Sanger sequencing using the primers Lobo F1 and Lobo R1 (Lobo et al. 2013). The chromatograms of all sequences obtained were visually inspected for ambiguous bases, and low-quality ends and primers were trimmed using MEGA (Kumar et al. 2018). Forward and reverse sequences of each sample were aligned using ClustalW and the consensus sequence were retrieved for each individual (Annex 7.1.). Species identification was achieved by homology matching using BLAST (Altschul et al. 1990) through GenBank (Benson et al. 2013) and Bold Systems (Ratnasingham & Hebert, 2007).

2.3. Regeneration experiments

After 1-week acclimatization, 6 individuals of *Leptoplana* sp. were randomly assigned to initiate the regeneration experiments. 4 individuals were cut in two portions and 2 individuals were cut in three portions, which were allowed to regenerate in Petri dishes or in beakers, both filled with artificial seawater and constant aeration as it is illustrated in Fig. 2.1. In both conditions, the planarians were protected from direct light and only during the feeding, cleaning and other husbandry processes, there was light exposure. These trials were also performed with 2 individuals of *D. tigrina* and repeated with 4 more individuals of *Leptoplana* sp. In all cases the survival time of the fragments were registered, and some statistical analyses were performed to assess differences in survival times. Beyond these regeneration experiments, some fragments of *Leptoplana* sp. were prepared for histological analyses after 4 or 32 days post-injury.

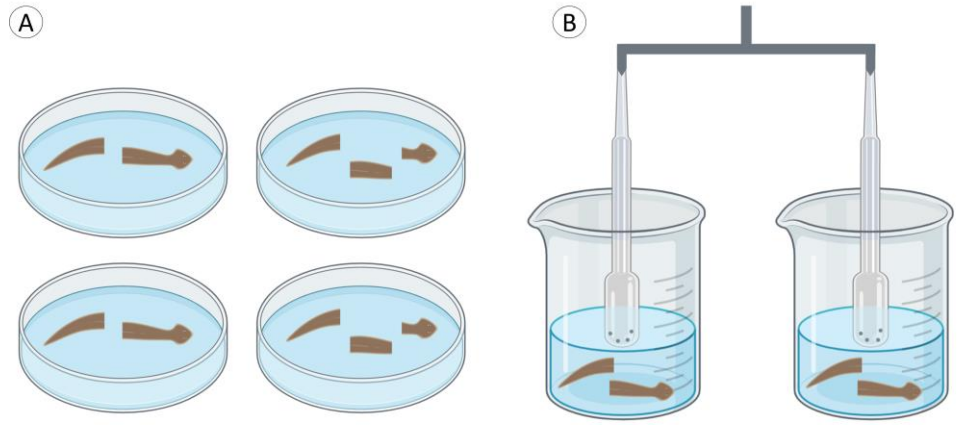

Figure 2.1. Experimental design of the preliminary regeneration trials carried out in *Leptoplana* sp. A) Petri dishes containing the regenerating planarians cut in two or three portions. The system also contained an aeration tube responsible for circulating the surrounding air. **B)** Beakers containing the regenerating planarians cut in two pieces and a Pasteur's pipette in direct contact with the water and connected to the aeration system. Illustration created with BioRender.com.

2.3.1. Histology

Samples were fixed in 2.5% (v/v) glutaraldehyde fixative (in 0.2 M cacodylate buffer, pH 7.4) for 2 hours. After this period the samples were washed with 0.2 M cacodylate buffer (pH 7.4) 3× for 15 min each. This was followed by a 2 days step of post-fixation in 1% osmium tetroxide (OsO_4) prepared in 0.2 M cacodylate buffer (pH 7.4). After that, the samples were washed with H_2O MQ 3× for 15 min each to remove all the osmium tetroxide sediments and then dehydrated in an increasing series of acetone percentages (30%, 50%, 70%, 90%, 100%) for 10 min each. To make a progressive infiltration, the samples were transferred through various solutions with different proportions of propylene oxide:Epon resin (2:1, 1:1, and 1:2) for 30 min each, ending with a final infiltration step in 100% Epon resin at low-pressure vacuum for 30 min. The molds were filled with Epon resin and the samples were placed at the bottom of the molds and left in the incubator at 65°C o/n to allow polymerization. The resin blocks were

shaped onto a trapezoid cutting surface to produce sections of 2-3 µm thickness with a RM2125 RTS manual rotatory microtome and a RM2245 semi-motorized rotatory microtome (Leica Microsystems) fitted with a tungsten carbide disposable blade. Sections were collected and placed onto a microscopy glass slide with distilled water and left to dry on a hot plate with a temperature between 60-70°C. Once dried, the slides were stained with Toluidine Blue by immersion for about 2 minutes, washed with H_2O MQ and left to dry on a hot plate. Slides were mounted with Dibutylphthalate Polystyrene Xylene (DPX) resin and observed under a DM2500 optical microscope equipped with a DFC 480 digital camera (Leica Microsystems). The Leica Application Suite (Leica Microsystems) was used to obtain the images used in the histological analysis. At least 2 or 3 slides corresponding to each sample were always analyzed, each containing about 20 semithin sections. Then, the zones of potential interest were photographed and examined, and the image processing was done using ImageJ (Schneider et al. 2012) and GIMP (2.10.4).

2.4. A transcriptomic approach to the regeneration process in planarians: focus on DNA repair

2.4.1. Data collection and statistical analyses

The molecular approach to regeneration in planarians was carried out using the raw data of an experiment conducted by Li et al. 2019 (available in NCBI/GEO – accession GSE121045). These data consisted of expression profiles of a total of 36 035 contigs, corresponding to individual genes obtained by next-gen high throughput sequencing (RNA-Seq) performed on specific body regions of planarian *Schmidtea mediterranea*. The regions under study were the pole piece of the anterior end and their flanking regions in uninjured animals and regenerating animals 72-hours after transverse amputation.

To annotate the 36 035 contigs, the planarian transcriptome was assembled against the reference transcriptome smed_20140614 defined in GEO submission: GSE72389 and the expression levels were quantified using Kallisto v0.44.0 (Bray et al. 2016).

R (4.0.0) software (Ihaka & Gentleman, 1996) was used to analyze the data corresponding to the number of read counts per gene per sample (3 per treatment/experimental condition). Through packages edgeR v3.30.0 (Robinson et al. 2010; McCarthy et al. 2012) and limma v3.44.1 (Ritchie et al. 2015) these data were grouped by treatment, normalized and converted to logFC (Fold Change) values using linear models to make paired comparisons between treatments and to estimate relative expression levels. The planned comparisons are represented in Fig. 2.2. and were designed to determine if the biggest differences in terms of gene expression were between the body region or between the condition to which planarians were subjected, that is, if the animals were intact or regenerating.

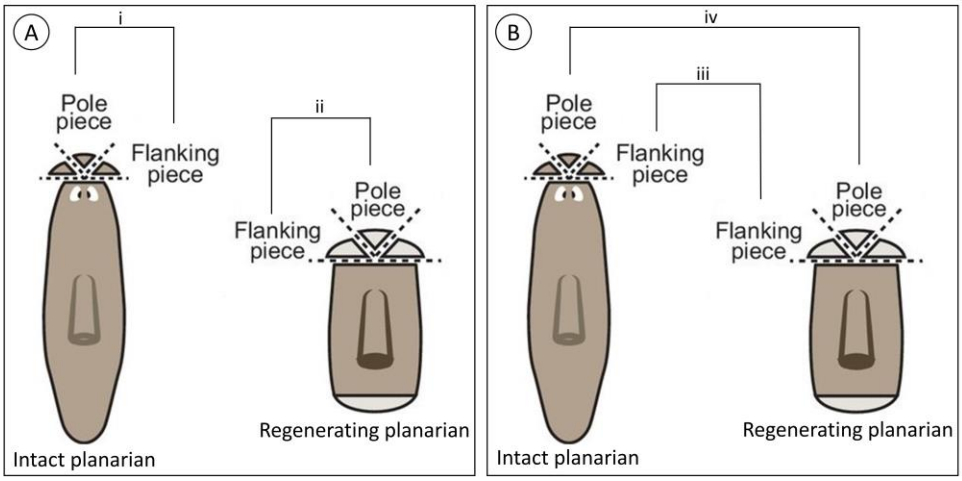

Figure 2.2. Illustration of the comparisons made in this study. A) First set of comparisons, which consisted of comparing the gene expression levels of the pole piece of the head with the flanking piece in intact planarians (i) and comparing the expression levels of the same regions in regenerating planarians (ii). In both cases the expression values corresponding to the flanking piece were used as control. **B)** The second set of comparisons consisted of comparing the expression levels of the flanking piece of regenerating planarians with the corresponding region in intact planarians (iii) and comparing the expression levels of the pole piece of regenerating planarians with the same region in intact planarians (iv). In both cases the expression values of the intact planarians were used as control. Figure adapted from Li et al. 2019 and RNA sequencing data obtained from GSE121045.

The results obtained from the comparisons based on fold change variations (logFC value greater than or equal to |1.5| and a significant FDR (false discovery rate) adjusted p-value set at 0.05) allowed post-processing of the data to identify differentially-expressed genes (DEGs). Further details on bioinformatic analysis can be found in Annex 7.11. which contains the script created for these analyzes.

2.4.2. Gene enrichment and pathway analysis

Contigs, corresponding to individual 36 035 genes were contrasted against several customized databases assembled from the Nucleotide Database of NCBI containing sequences of Eumetazoan genes involved in the several processes known to be associated with regeneration such as cell proliferation and differentiation, cell cycle control, and DNA repair and detection. The contrast was made using BLAST (Altschul et al. 1990) and afterwards the R software allowed to obtain the DEGs involved in these processes. Then, a search was carried out using the STRING database (Szklarczyk et al. 2019) to identify protein networks based on canonical interactions between these protein-coding genes, applied to processes associated with regeneration. The search was made based on the homologous proteins in *Homo sapiens*, as planarians are absent from STRING and still do not benefit from high levels of genomic annotation.

3. Results

3.1. Planarian species identification

Individuals representing the three captured and identified specimens are represented in Fig. 3.1.: *Discocelis tigrina* (Fig. 3.1.A), *Leptoplana* sp. (*L. mediterranea,* or *L. tremellaris*) (Fig. 3.1.B) and an unidentified planarian species (Fig. 3.1.C).

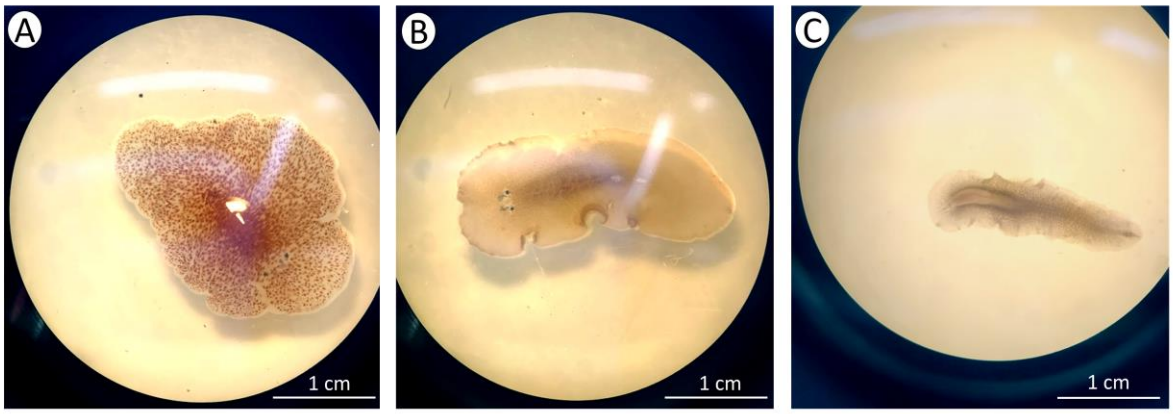

Figure 3.1. Planarian species observed under an optical stereoscope. A) *D. tigrina*. **B)** *Leptoplana* sp. **C)** Unidentified species.

D. tigrina was characterized by a pigmented and patterned body with several brown spots in the dorsal surface, with a higher pigment density in the center of the body, which corresponded to the pharynx region. These specimens ranged between ≈2-3 cm in length and ≈1.5-2 cm in width. *Leptoplana* sp. had a light brown elongated body with a darker color in the pharynx area, ranging between ≈1-3 cm in length and ≈0.5-1 cm in width. The “eyes” of both species corresponded of two clusters of small scattered eyes. The unidentified planarian species had ≈2 cm in length and ≈0.7 cm in width and a light brown elongated body with two dark pigmented lateral lines in the anterior portion that merged in the posterior portion as its pigmentation faded. The “eyes” (ocelli) were not conspicuous.

Results from the molecular validation to confirm the species taxonomy (Fig. 3.2.) showed that in all DNA samples corresponding to the two species under study there was a band with a size around 600 bp, corresponding to the *COI-5P* gene since its amplicon size was ≈650 bp. In both *Leptoplana* sp. samples this band was very dense, while in the sample corresponding to the unidentified planarian the density was lower.

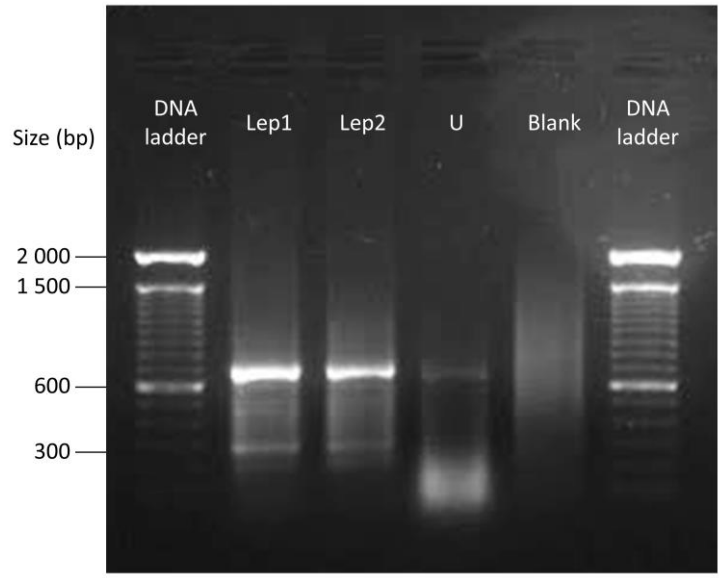

Figure 3.2. Agarose gel (1.2%) correspondent to the PCR products obtained from the sequencing of *COI-5P* gene of different planarian species. From left to right: DNA marker, *Leptoplana* sp., *Leptoplana* sp., Unidentified planarian species, Blank, DNA marker.

The consensus sequences for each sample (presented in Annex 7.1.) were used for homology-based matching, which allowed to conclude that all planarian species in study belonged to Order Polycladida. The planarians identified as *Leptoplana* sp. (Lep1 and Lep2) were included in Suborder Acotylea and Family Leptoplanidae. Conversely, the unidentified planarian (U) belonged to Suborder Cotylea, Family Prosthiostomidae, without a conclusive match. The taxonomic results obtained from both GenBank and Bold Systems are represented in more detail in Annexes 7.2. - 7.10. The Bold Systems results provided the percentage of similarity between the *COI-5P* sequences of the samples in study and the ones registered in the database and allowed a species-level identification when it was possible. The GenBank results included the description of the sequences producing significant alignments according to several parameters, such as maximum score, total score, query coverage, expected value, percentage of identity and the accession number of the most similar sequence in the public database.

3.2. Regeneration experiments

3.2.1. Histological analyses

The normal structure of the epidermis of *Leptoplana* sp. (represented in the upper panel of Fig. 3.3.) consisted of a single layer of epidermal cells juxtaposed to each other and attached to the basal membrane, separating it from the connective tissue. In Fig. 3.3.C it is possible to observe the nuclei of each one of the epidermal cells in more detail and the fibrocytes and other cell types present in the mesenchymal tissue.

As for the regenerating tissue analysis, the anterior fragment of a specimen of *Leptoplana* sp. 4 days post-injury was observed in two different areas, one immediately after the sectioning site and other in a more anterior location, as it is schematized in Fig. 3.3.D and G, respectively. Thus, near the wound site (as represented in the middle panel) it was possible to observe the junction of two damaged epidermis ends (indicated by the arrows) and the presence of undifferentiated cells, namely neoblasts in the wound region (Fig. 3.3.E). Neoblasts are pluripotent stem cells capable of forming the blastema, the structure from which regeneration takes place. These cells were located under the basal membrane, dispersed in mesenchymal tissue, and are represented in more detail in Fig. 3.3.F. They could be easily distinguished by their spherical shape, round and big nuclei and reduced cytoplasmatic content. Finally, in the bottom panel, it was possible to notice that the wound was a little more closed and seemed to be healing (Fig. 3.3.H), since it corresponded to a more anterior region, near to the uninjured tissue of the head (as it is schematized in Fig. 3.3.G). Besides this, there were also neoblasts present in the mesenchymal tissue under the basal membrane (Fig. 3.3.I), however, these were not as undifferentiated as the ones present in Fig. 3.3.F. and their nuclei were not clearly visible. Altogether these observations indicated an active regeneration of the damaged structures in this organism.

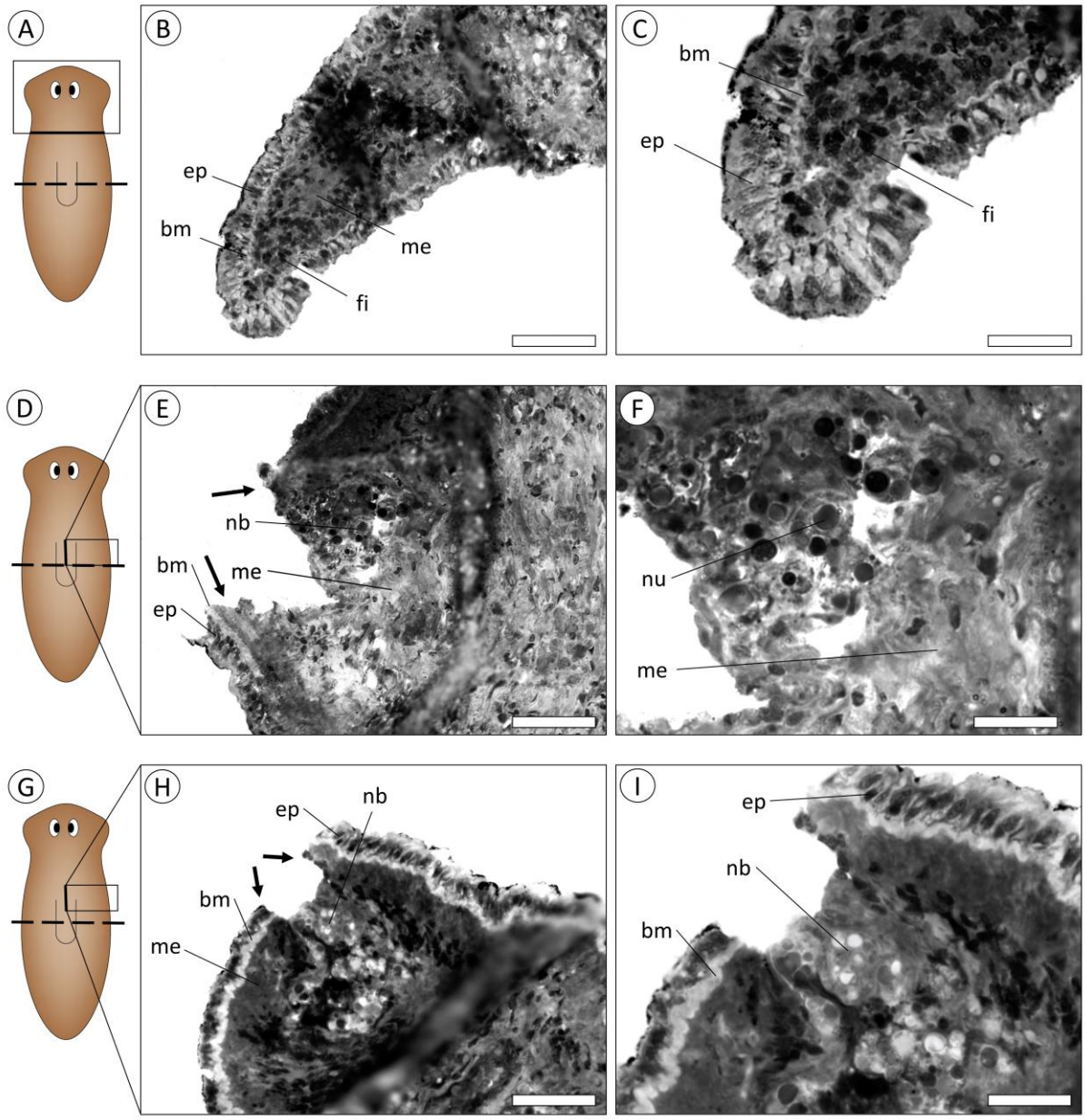

Figure 3.3. Histological sections from different body parts of planarian *Leptoplana* sp. 4 days post-injury. A) Representative scheme of the cross sections made above the blastema region of the head fragment. **B)** Normal structure of the planarian epidermis which consists of a columnar epithelium with a single layer of epidermal cells (ep) attached to the basal membrane (bm) and below this there is mesenchymal tissue (me) and fibrocytes (fi). Magnification of 200x. Scale bar: 100 µm. **C)** Epidermal cells represented at a magnification of 400x. Scale bar: 50 µm. **D)** Representative scheme of the sagittal sections made in the blastema region at the wound site of the head. **E)** Dense aggregation of neoblasts (nb) scattered in the mesenchymal tissue below the basal membrane which indicates the formation of the blastema at the wound site. The arrows indicate two epidermis ends that are joining. Magnification of 400x. Scale bar: 50 µm. **F)** Neoblasts observed at a magnification of 1000x, easily distinguished by their spherical shape, round nuclei (nu), and low cytoplasmatic content. Scale bar: 20 µm. **G)** Representative scheme of the sagittal sections made in a more anterior region of the blastema formed at the wound site of the head. **H)** Neoblasts in the mesenchymal tissue and arrows indicating the wound closure. Magnification of 400x. Scale bar: 50 µm. **I)** Neoblasts and epidermis represented in more detail, with a magnification of 1000x. Scale bar: 20 µm. The planarians illustrations (A, D, and G) are adapted from Newmark & Alvarado, 2002. All sections (B, C, E, F, H, and I) have 2-3 µm thickness and are stained with Toluidine Blue and represented in grayscale.

Beyond these results, there were also histological samples of 1-month regenerating planarians, which are represented in Fig. 3.4. The sections corresponded to the posterior portion of the body and were represented in a sagittal view, as it is schematized in Fig. 3.4.A. In this case, it was possible to observe an epidermis cell layer practically continuous, and some cellular content dispersed along the connective tissue, such as fibroblasts as it is represented in Fig. 3.4.B and C.

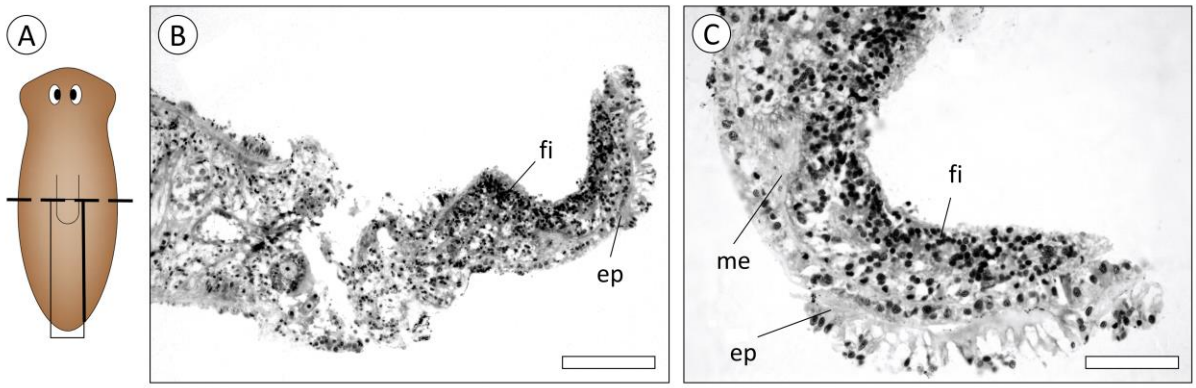

Figure 3.4. Histological sections from the posterior fragment of the body of *Leptoplana* sp. 1-month post-injury. A) Representative scheme of the sagittal sections made in this sample. **B)** Epidermis structure practically restored besides some artifacts in the epidermal cell layer (ep). There are also some fibroblasts (fi) scattered in the mesenchymal tissue (me). Magnification of 100x. Scale bar: 200 µm. **C)** Epidermis structure and mesenchymal tissue with cellular content represented in a magnification of 200x. Scale bar: 100 µm. The planaria illustration (A) is adapted from Newmark & Alvarado, 2002. All sections (B, and C) have 2-3 µm thickness and are stained with Toluidine Blue and represented in grayscale.

3.2.2. Survival times

The results from the preliminary trials of the regeneration experiments mentioned in section 2.3. are presented in Table 3.1.A and correspond of a total of n=6 individuals of *Leptoplana* sp. For each of these individuals the two or three fragments in which they were cut are represented as well as their respective survival time (indicated in days). To complement these results, a second set of regeneration experiments was made 10 days after the first experiments. For that, 4 additional individuals of *Leptoplana* sp. were cut in two pieces and allowed to regenerate in Petri dishes. The survival times were also analyzed, as represented in Table 3.1.B. The findings from the experiments with *D. tigrina* are presented in Table 3.1.C.

Table 3.1.A. Survival time of different fragments from *Leptoplana* sp. when subjected to the first regeneration experiments initiated 1 week after acclimatization.

Species	Experiment	Fragment	Survival time (days)
***Leptoplana* sp.**	Petri dishes	Anterior	4
		Intermediate	3
		Posterior	8
		Anterior	7
		Intermediate	7
		Posterior	3
		Anterior	9
		Posterior	10
		Anterior	19
		Posterior	32
	Beakers	Anterior	2
		Posterior	2
		Anterior	4
		Posterior	4

Table 3.1.B. Survival time of different fragments from *Leptoplana* sp. when subjected to the second set of regeneration experiments.

Species	Experiment	Fragment	Survival time (days)
***Leptoplana* sp.**	Petri dishes	Anterior	4
		Posterior	4
		Anterior	6
		Posterior	6
		Anterior	14
		Posterior	14
		Anterior	14
		Posterior	14

Table 3.1.C. Survival time of different fragments from *D. tigrina* when subjected to the regeneration experiments initiated 1 week after acclimatization.

Species	Experiment	Fragment	Survival time (days)
Discocelis tigrina	Petri dishes	Anterior	1
		Intermediate	3
		Posterior	1
		Anterior	1
		Posterior	14

Non-parametric Dunn's test (using Benjamini-Hochberg adjustment) was performed for multiple pairwise comparisons between the fragments prevenient from *Leptoplana* sp. individuals cut in two and in three portions. No significant differences were observed (adjusted p-value ≈0.2), however there was a tendency that seemed to indicate that cutting the animals in two fragments was more efficient than in three. To complement these results, an analysis of variance was also performed which further evidenced the tendency of the number of fragments to have an effect on survival time, although not significant (p-value 0.072). The mean and standard deviation of the survival time of each fragment of *Leptoplana* sp. and *D. tigrina* were calculated and are represented in the plot of Fig. 3.5. The results presented seemed to indicate the same tendency mentioned above and it was also possible to verify that there were cases where the posterior fragment appeared to have a longer survival time than the anterior fragment.

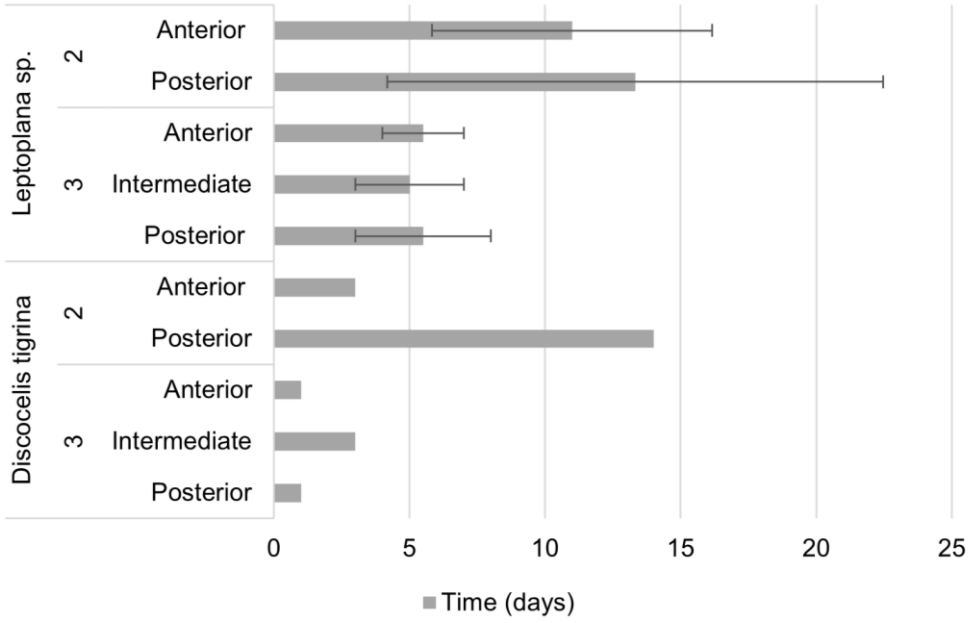

Figure 3.5. Bar plot of the survival time of each fragment from the planarian species under study. Each bar represents the average value with the respective standard deviation. For the *Leptoplana* sp. individuals cut in two fragments, n=6 individuals were analyzed, and for the ones cut in three fragments, n=2 individuals were analyzed. In the case of *D. tigrina*, only one individual was analyzed for each condition.

As for the planarians that were kept in beakers, the results were not taken into account for this analysis because these planarians were fixed for histology when they were still regenerating, so their survival time was not real.

3.3. Gene expression analysis

3.3.1. Differentially-expressed genes determination

The results regarding the gene expression comparison of different body parts of *Schmidtea mediterranea* made in this study showed that in the first set of comparisons (Fig. 2.2.A), a total of 402 of the 36 035 genes were differentially-expressed in the pole pieces (according to the criteria mentioned in section 2.4.1.). From these 402 DEGs, 41 were common to both comparisons, 77 were exclusively

differentially-expressed in the pole piece of intact planarians (comparison i) and 284 were exclusive of the pole piece of regenerating planarians (comparison ii). It was also verified that most of the genes were down-regulated. These results are represented in more detail in the Venn diagram of Fig. 3.6. which reveals the differentially up- (red) and down-regulated (blue) genes obtained in each comparison.

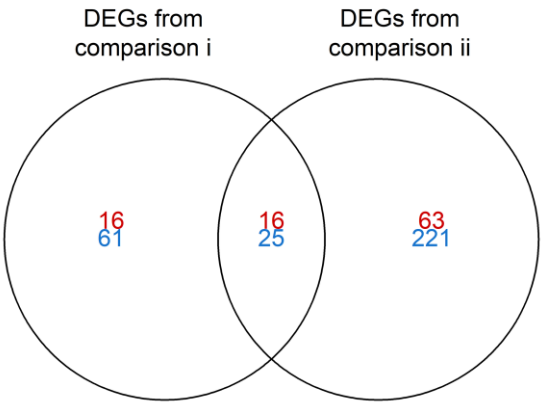

Figure 3.6. Venn diagram illustrating differentially-expressed genes (DEGs) between the pole piece of intact planarians when compared with the flanking piece (comparison i) and between the pole piece of regenerating planarians when compared with the flanking piece (comparison ii). Red and blue represent up- and down-regulated genes, respectively, based on the criteria logFC ≥ |1.5| and FDR-adjusted p-value ≤ 0.05.

As for the second set of comparisons (Fig. 2.2.B), a total of 5 561 of the 36 035 genes were found to be differentially-expressed in the regenerating planarians, according to the same criteria. Within these genes, 2 415 were differentially-expressed in the two comparisons, 1 454 were exclusively observed between the comparison of the two flanking pieces (comparison iii) and 1 692 were exclusive of the comparison between the two pole pieces (comparison iv). As in the previous case, most of the genes were down-regulated (blue). The results obtained are presented in Fig. 3.7.

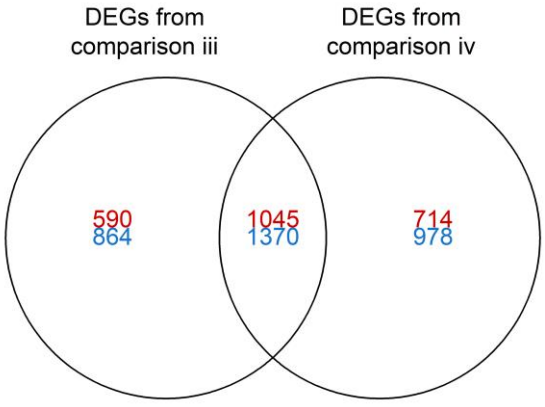

Figure 3.7. Venn diagram illustrating differentially-expressed genes (DEGs) between the flanking piece of the regenerating planarians when compared with the corresponding region in the intact planarians (comparison iii) and between the pole piece of the planarians in regeneration when compared with the corresponding region in the intact planarians (comparison iv). Red and blue represent up- and down-regulated genes, respectively, based on the criteria logFC ≥ |1.5| and FDR-adjusted p-value ≤ 0.05.

3.3.2. Homology matching and functional annotation

The customized cell proliferation database assembled from NCBI contained 24 975 genes, of which 49 were found in the gene list of the 36 035 genes identified in the planarian transcriptome. The DNA replication database included 23 660 genes, of which 21 matches were found in the planarian transcriptome. The cell differentiation database contained the largest number of genes, with 258 945 genes, of which 84 were found in the gene list in study. The DNA repair gene database contained 121 695 genes, of which 76 correspondences were found in the planarian transcriptome. As for the cell cycle arrest database, 2 112 genes were analyzed, and 5 matches were found. The cell death database contained 35 419 genes, of which 23 were present in the gene list in study, and finally, a database containing 141 tumor suppressor genes was also analyzed, however, no correspondences were found in the gene list in study. Therefore, a total of 258 genes out of 36 035 were found to be related with regeneration. These genes, as well as the information related to them are listed in Annex 7.12. The presence of these genes was evaluated in the list of DEGs obtained for each comparison and the results regarding the first set of comparisons revealed the presence of two DEGs in the pole piece of the intact planarians when compared against the flanking piece (comparison i), both being down-regulated, as it is possible to observe in the Venn diagram of Fig. 3.8. Of these two genes, one was related with DNA repair (*RAD51*) and the other, *TPH*, has found to be involved in cell proliferation and differentiation processes. The relative expression levels (presented as logFC) of these two genes are represented in Table 3.2. and in the heatmap of Fig. 3.9.

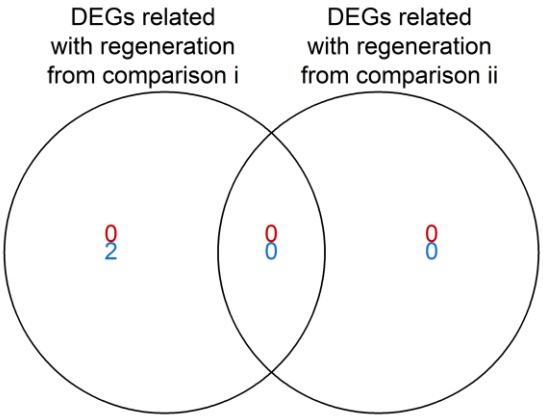

Figure 3.8. Venn diagram representing differentially-expressed genes (DEGs) associated with regeneration found for the comparison between the pole and flanking pieces in intact planarians (comparison i) and between the same regions in regenerating planarians (comparison ii). Red and blue represent up- and down-regulated genes, respectively, based on the criteria logFC ≥ |1.5| and FDR-adjusted p-value ≤ 0.05.

Table 3.2. Relative expression levels represented in logFC of two genes related with regeneration found for the comparison between the pole and flanking pieces in intact and in regenerating planarians 72-hours after transverse amputation (comparisons i and ii, respectively). The FDR-adjusted p-value associated with each logFC value is also represented.

Process	Gene	Comparison i		Comparison ii	
		logFC	FDR p-value	logFC	FDR p-value
DNA repair	*RAD51*	-1.6317	$9.8067e^{-7}$	-0.0674	0.7696
Cell proliferation / **Cell differentiation**	*TPH* - Tryptophan hydroxylase	-2.3432	$1.7624e^{-18}$	0.5284	0.0093

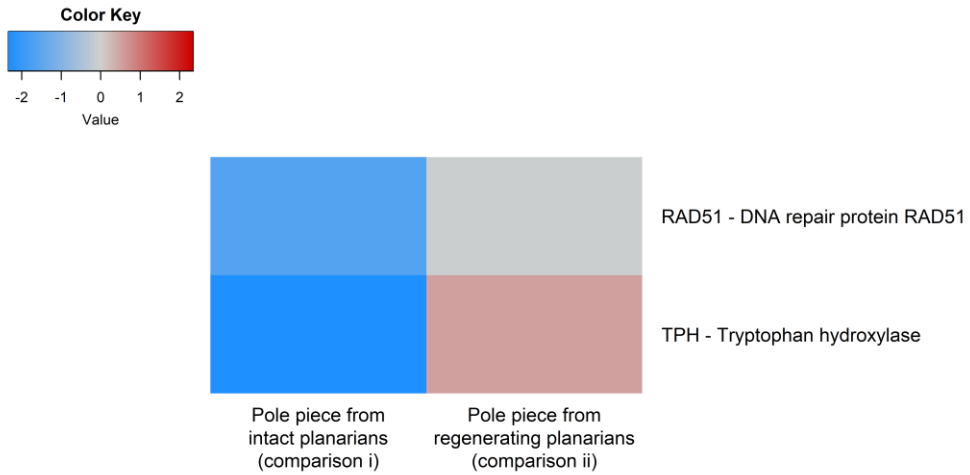

Figure 3.9. Heatmap showing differentially-expressed genes (DEGs) between the pole and flanking pieces in intact (comparison i) and in regenerating planarians (comparison ii).

The heatmap results showed that in intact planarians (comparison i) the gene expression levels of both genes were lower in the pole piece when compared to the flanking piece, which may indicate that they are down-regulated. On the other hand, in the regenerating planarians (comparison ii) the expression level of *TPH* was a little higher in the pole piece regarding the flanking piece, although not statistically significant (logFC < |1.5|, as it is possible to verify through the values presented in Table 3.2.) and there was practically no expression variation of *RAD51* (logFC = -0.0674).

As for the results concerning the comparisons iii and iv, according to the Venn diagram represented in Fig. 3.10., a total of 32 DEGs related with regeneration were found to be in common in the flanking piece and in the pole piece of regenerating planarians. Besides this, one gene (*PDCD5*) was exclusively down-regulated in the flanking piece (comparison iii) and 8 genes were found to be exclusively differentially-expressed in the pole piece (comparison iv). All these genes and their relative expression levels were represented in the heatmap of Fig. 3.11. and in Table 3.3., in which it was possible to identify 12 DEGs related with DNA repair, 6 related with cell proliferation and 10 related with DNA replication, as well as 11 associated with the process of cell differentiation, 3 with cell death, and one associated with cell cycle arrest.

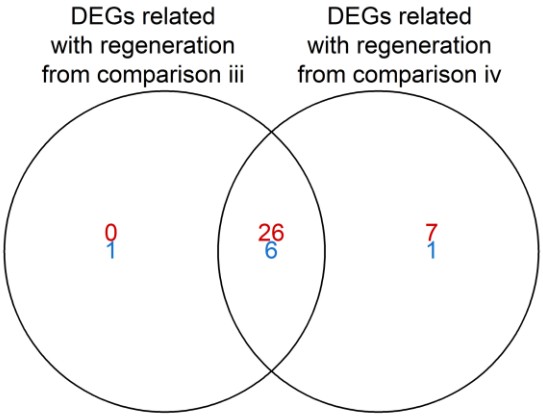

Figure 3.10. Venn diagram representing differentially-expressed genes (DEGs) associated with regeneration found for the comparison between the flanking pieces of regenerating and intact planarians (comparison iii) and the pole pieces of regenerating and intact planarians (comparison iv). Red and blue represent up- and down-regulated genes, respectively, based on the criteria logFC ≥ |1.5| and FDR-adjusted p-value ≤ 0.05.

Table 3.3. Relative expression levels represented in logFC of the 41 genes related with regeneration found for the comparison between the flanking pieces of regenerating and intact planarians and the pole pieces of regenerating and intact planarians (comparisons iii and iv, respectively). The FDR-adjusted p-value associated with each logFC value is also represented. The expression levels of regenerating planarians were measured 72-hours after transverse amputation.

Process	Gene	Comparison iii		Comparison iv	
		logFC	FDR p-value	logFC	FDR p-value
DNA repair	*DNAH7* - Dynein axonemal-like heavy chain 7	-2.0491	$5.5651e^{-42}$	-1.5990	$3.4852e^{-27}$
	MCM8 - DNA helicase MCM8-like	2.0348	$1.4426e^{-7}$	1.8521	$8.3120e^{-8}$
	Cre-ceh-43	1.4055	$5.2510e^{-21}$	1.6324	$8.5775e^{-28}$
	TONSL - Tonsoku-like protein	2.2226	0.0056	2.6597	0.0017
	PCNA - Proliferating cell nuclear antigen	2.2170	$6.8043e^{-47}$	2.8765	$5.4086e^{-68}$
	me31B - putative ATP-dependent RNA helicase me31b	2.6663	$4.5538e^{-10}$	1.9811	$2.1019e^{-5}$
	vas - ATP-dependent RNA helicase vasa, isoform A-like	-1.5420	$1.8497e^{-21}$	-1.9534	$1.1374e^{-32}$
	POLG - DNA polymerase gamma, catalytic subunit	0.3652	0.6711	2.6194	0.0007
	BLM - Bloom syndrome protein homolog	3.9705	0.0368	5.8496	0.0022
	ZNFX1 - NFX1-type zinc finger-containing protein 1-like	-0.9135	$2.1185e^{-5}$	-1.5542	$9.0038e^{-13}$
	DNAH6 - Dynein axonemal heavy chain 6	-1.9529	$2.1766e^{-41}$	-1.5194	$1.9655e^{-26}$
	RAD51	1.5291	$2.5952e^{-9}$	3.0934	$3.4584e^{-25}$
Cell proliferation	*FOXK1* - Forkhead box protein K1	2.2226	0.0056	2.6597	0.0017
	NFYB - Nuclear transcription factor Y subunit B	-1.9062	$2.3973e^{-13}$	-2.0274	$2.5079e^{-15}$
	TPH - Tryptophan hydroxylase	2.3236	$1.1090e^{-25}$	5.1953	$1.5880e^{-85}$
	sd-1 - Scalloped transcription factor homolog 1 (*TEAD1*)	1.4330	$5.8970e^{-27}$	1.5252	$3.3209e^{-29}$

	CycD - Cyclin D-like protein (*CCND1*)	2.3859	$1.4114e^{-39}$	2.3219	$4.1569e^{-38}$
	ACTB - Beta-actin	2.8688	$1.35e^{-84}$	3.0052	$9.9923e^{-93}$
DNA replication	*MCM4* - DNA replication licensing factor MCM4	2.0348	$1.4426e^{-7}$	1.8521	$8.3120e^{-8}$
	PPHLN1 - Periphilin-1	2.2226	0.0056	2.6597	0.0017
	MCM2 - DNA replication licensing factor MCM2-like, transcript variant X1	1.7349	$1.1785e^{-15}$	2.3620	$1.2395e^{-25}$
	PRIM1 - DNA primase small subunit	1.2178	$1.1108e^{-5}$	2.3518	$3.1845e^{-14}$
	MCM6 - DNA replication licensing factor MCM6-like	1.4364	$3.5392e^{-7}$	1.5759	$5.7404e^{-9}$
	MCM3 - DNA replication licensing factor MCM3	1.8193	$1.4608e^{-20}$	2.1812	$4.4178e^{-28}$
	MCM7 - DNA replication licensing factor MCM7 homologue	2.1627	$2.6516e^{-23}$	2.4999	$7.5507e^{-30}$
	MCM5 - DNA replication licensing factor MCM5	1.8973	$2.1263e^{-29}$	2.7014	$1.9735e^{-53}$
Cell differentiation	*BRAF* - Serine/threonine-protein kinase B-raf proto-oncogene, transcript variant X1	2.2226	0.0056	2.6597	0.0017
	17000600008005 GRN_PREHEP, Homo sapiens	5.2333	$3.0698e^{-5}$	4.0494	0.0007
	17000599933949 GRN_PRENEU, Homo sapiens	2.9424	$2.9228e^{-123}$	3.0366	$3.6340e^{-130}$
	NSD1 - Histone-lysine N-methyltransferase, H3 lysine-36 specific, like	1.9613	$3.9272e^{-27}$	1.9819	$1.4259e^{-27}$
	PIM1 - Serine/threonine-protein kinase pim-1	-2.5381	$2.9788e^{-37}$	-2.8438	2.8620e-45
	ntn2 - Netrin 2.1	4.4719	$1.0418e^{-73}$	5.1113	$4.8061e^{-80}$
	DjhistonH2B - Histone H2B.1	7.0499	$3.2839e^{-15}$	4.8144	$3.8766e^{-12}$
	DjhistonH2B - Histone H2B.2	2.9601	$8.7787e^{-35}$	3.8575	$5.4294e^{-48}$
	SUSFLECK Fat Cell	-2.7420	0.0032	-2.1316	0.0035
	ntn2 - Netrin 2.2	5.1229	$7.8234e^{-18}$	3.6307	$3.0343e^{-11}$
	DjmkpA - Mitogen-activated protein kinase phosphatase	1.4543	$2.7336e^{-35}$	1.5313	$1.6163e^{-39}$
Cell death	*ZMAT3* - Zinc finger matrin-type protein 3	2.2226	0.0056	2.6597	0.0017
	18S ribosomal RNA-like	5.2334	$3.0698e^{-5}$	4.0494	0.0007
	PDCD5 - Programmed cell death protein 5-like	-1.7119	$9.4399e^{-23}$	-1.3813	$7.2068e^{-16}$
Cell cycle arrest	*CDC42* - Cell division cycle 42	2.2226	0.0056	2.6597	0.0017

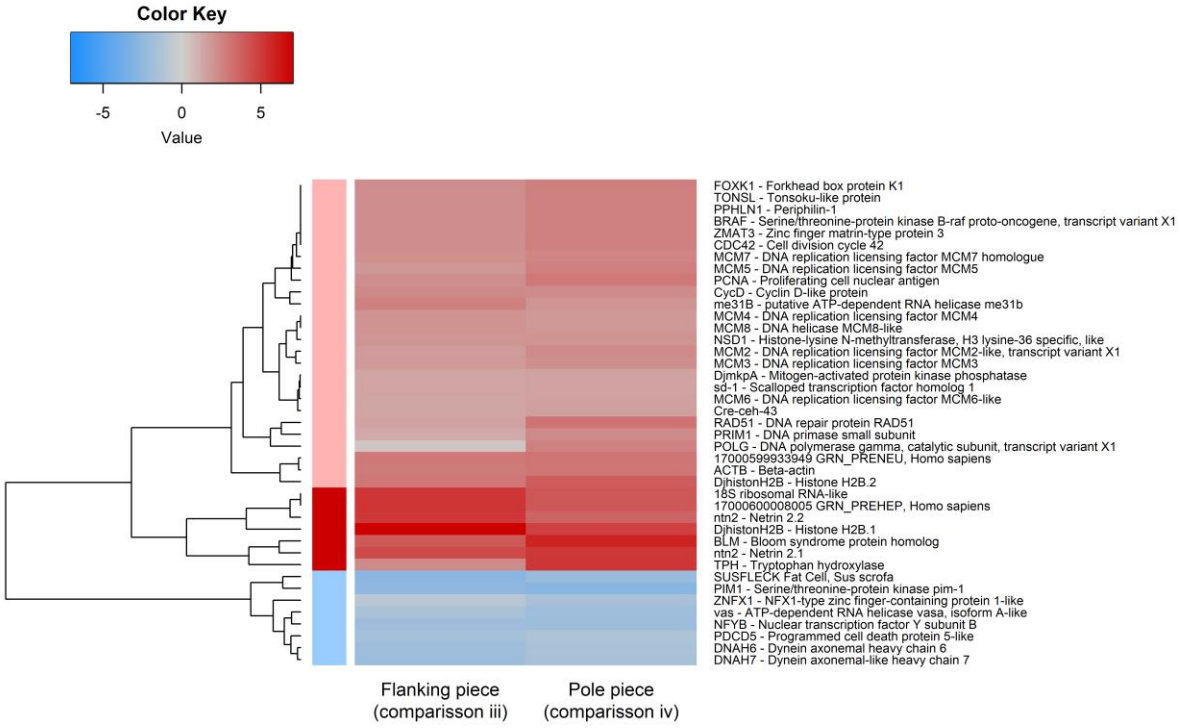

Figure 3.11. Heatmap showing differentially-expressed genes (DEGs) between the flanking pieces of regenerating and intact planarians (comparison iii) and the pole pieces of regenerating and intact planarians (comparison iv). The cluster analysis was made using Euclidean distances and complete linkage as metrics and amalgamation rule.

In this heatmap it was possible to notice that, generally, the pattern of expression was quite similar between the two regions under study (pole and flanking pieces), and it was possible to observe three clusters of genes, indicated by the vertical colored bar, with one cluster that corresponded to the down-regulated genes (represented in blue, which contained 8 genes) and two clusters of up-regulated genes in which the red one (consisting of 7 genes) had slightly higher levels of expression than the other (pink, consisting of 26 genes). The genes within these two clusters of up-regulated genes were analyzed to identify the connection between the proteins encoded by each one of them. From this list of 33 genes, 21 were found to have correspondence with homologous proteins in humans and the interactions between these proteins are represented in Fig. 3.12. Besides the general interactions, the biological processes in which these proteins were involved were also explored and the most evident were DNA replication, cell cycle regulation, and DNA repair (as presented in Fig. 3.13. A, B, and C, respectively). The same analysis was also performed for the list of down-regulated genes, however, the results obtained were not very conclusive and explanatory since the proteins did not seemed to interact with each other nor participate in any relevant process.

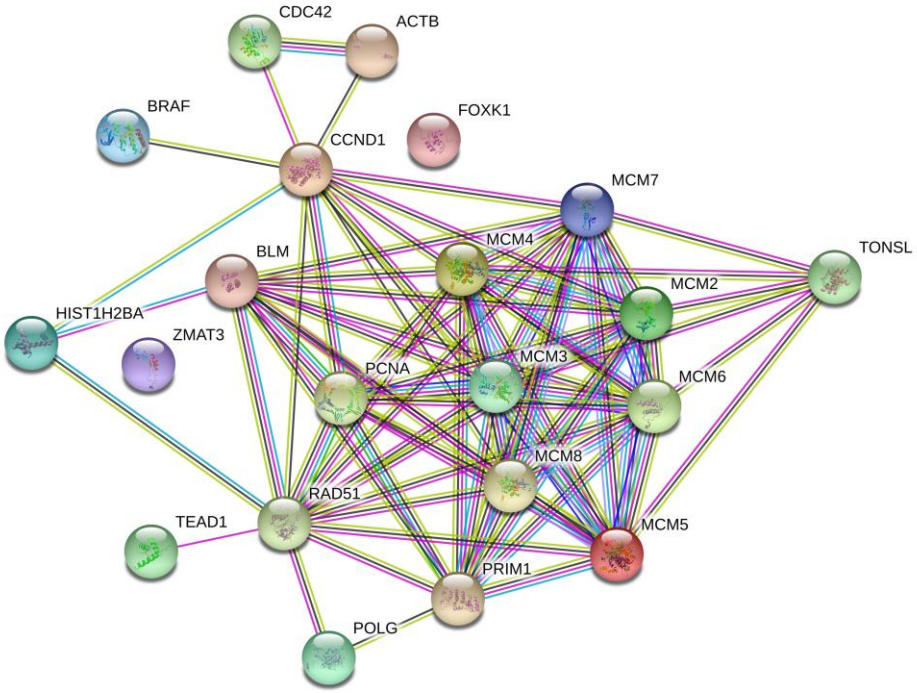

Figure 3.12. Protein interaction network between the up-regulated genes found in the comparisons between the flanking pieces of regenerating and intact planarians (comparison iii) and the pole pieces of regenerating and intact planarians (comparison iv). Each protein is represented by a circle and identified by its name. The lines connecting the proteins indicate that they jointly contribute to a shared function. There are different types of interactions represented by the different colors which may indicate known interactions from curated databases or experimentally determined interactions (retrieved from STRING database).

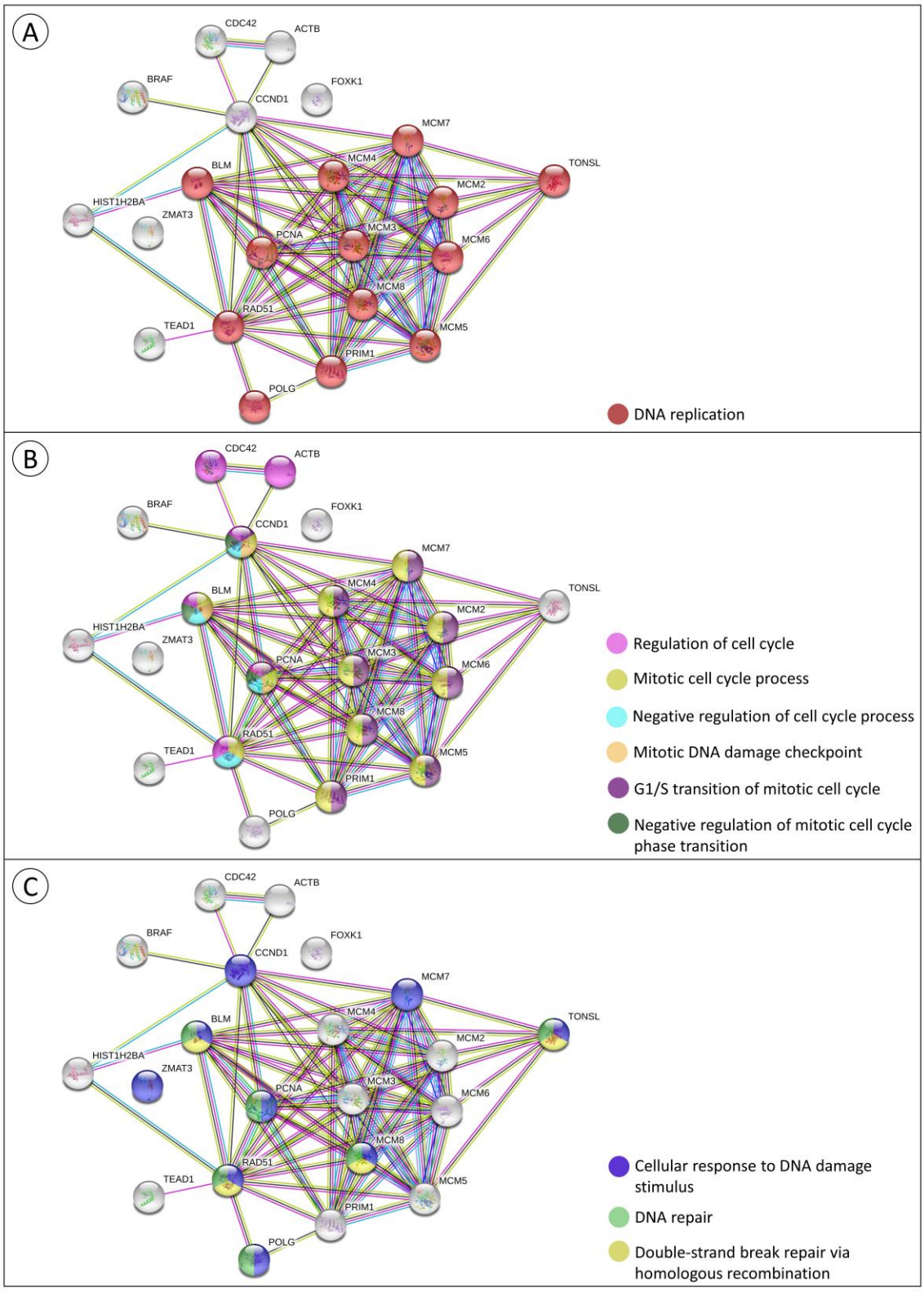

Figure 3.13. Protein interaction network of the up-regulated genes found in the flanking and pole pieces of regenerating planarians when compared with intact planarians (comparisons iii and iv) divided according to the different cellular processes in which they are involved. A) Proteins involved in DNA replication represented in red. **B)** Proteins related with cell cycle regulation processes, as it is indicated by the different colors. **C)** Proteins involved in DNA damage response and DNA repair, indicated by the legend. Proteins in grey in the three panels A, B, and C are not involved in any of the mentioned processes. Data retrieved from STRING database.

4. Discussion

4.1. Evaluation of the regeneration process

4.2.1. Anatomical modifications and cellular responses

The current findings demonstrate that *Leptoplana* endures a similar regeneration process to that described for other planarians such as *Dugesia* and *Schmidtea*. However, regeneration is seemingly a very complex and highly variable process without a clearly predictable outcome. This was shown by relatively high mortality and lack of an obvious pattern of time-dependent healing between individuals.

Regeneration in planarians is initiated by wound healing, which is mediated by local muscle contraction that allows contact between the remaining dorsal and ventral epidermis layers, which occurs right after a few minutes post-injury to prevent further damage and invasion of pathogens. Afterwards, the muscle relaxes, leading to the stretch of the epidermis, creating a thin membrane that closes the wound, called wound epidermis (Chandebois, 1980). In accordance, through the current experiments in *Leptoplana* sp. it was possible to observe that, 4 days post-injury, there were two epidermis ends close to each other (Fig. 3.3.E and 3.3.H), even though without complete contact at this stage. This small detachment of the two epidermis layers may be explained by an incomplete healing after 4 days or by the fact that the fixation process causes tissue shrinkage that may also contribute to draw apart these two ends. In turn, neoblasts and their proliferative capacity are also essential for the regenerative process (Orii et al. 2005). It has been described in most planarian species, for instance, in *Schmidtea mediterranea* and in *Girardia tigrina* (Baguñà, 1976; Saló & Baguñà, 1984) that neoblasts exhibit a coordinated biphasic pattern of mitotic activity during regeneration, where in the first 6 hours post-injury there is a “body-wide peak”, followed by a second peak restricted to the wound site between the second and fourth day after the injury in cases when there is tissue loss. This second peak leads to the production of non-dividing neoblasts in the pre-existing tissue in a specific area close to the wound site known as postblastema. These neoblasts will then migrate and accumulate in the new undifferentiated tissue (blastema) which will allow its growth (Saló & Baguñà, 1984). This second peak also seemed to occur in *Leptoplana* sp., once 4 days post-injury it was possible to notice the presence of neoblasts and the formation of the blastema in the region adjacent to the wound (recall Fig. 3.3.E).

The neoblasts in blastema will then differentiate into the various cell types needed to re-establish the lost body parts. This cell differentiation step happens due to the expression of cell-differentiation genes, which are activated ≈72 hours post-injury and continue to be expressed over time (Cebrià et al. 2018). In Fig. 3.3.F it was possible to notice that 4 days post-injury, there were neoblasts within the blastema of *Leptoplana* sp. with different sizes and in different differentiation stages. These findings indicated that neoblasts were beginning to organize themselves to restore the lost parts of the tissue and produce properly patterned structures. However, the neoblasts represented in Fig. 3.3.I did not seemed to have such a noticeable organization, which may be explained by the fact that they were located in a more anterior region, furthest from the wound site which seemed to be in a more advanced wound healing stage, making these neoblasts more advanced in the differentiation process.

The differentiation phase of neoblasts into other cell types is a continuous process over time and, after one month of regeneration, it was possible to notice a relatively large mass of small cells, namely fibroblasts, already dispersed in the mesenchymal tissue contrary to what was observed after the fourth day of regeneration in which the cells were much bigger and undifferentiated and their number was much lower. However, although the internal anatomy of these planarians appeared to be practically restored after one month into the regeneration process, the planarians had not yet acquired their original body size and did not appear to be regenerating. Considering that the normal regeneration time of the lost structures in *S. mediterranea* is about ≈1-2 weeks (Oviedo et al. 2008), possible explanations for the different results obtained for *Leptoplana* sp. after one month may be due to internal and/or environmental conditions, such as the quantity and quality of water since planarians were kept in Petri dishes with an amount of water that might not be enough, low concentration of essential nutrients, or even high salinity due to water evaporation that may have provided some bacteria growth and infections. Actually, except for the *Leptoplana* sp. samples fixed for histological observations, the remaining planarians from the regeneration experiments revealed failure to uphold tissue integrity or viability altogether, which is reflected in the survival time results in which the majority of the individuals and their respective fragments had a very short survival time. Despite this, and even with a low number of replicates that may have compromised the level of statistical significance mentioned in section 3.2.2., the results seemed to indicate that cutting the animals in two fragments was more efficient, which may be related to the fact that there were less structures to regenerate than when the animals were sectioned in three. In addition, another interesting result indicated that in some cases, the posterior fragment had a longer survival time than the anterior fragment, which had already been described in the literature for other planarian species or even polychaete worms (Baguñà & Romero, 1981; Oviedo et al. 2003; Bybee & Murray, 2017, pp. 573-574; Pandian, 2020, pp. 69). In an attempt to try to understand this result, it may be discussed that the vital organs such as brain, eyes and sensory organs are all situated in the anterior fragment of the body and these organs have a higher energy cost (i.e., they need more energy to maintain their normal function). Therefore, under stressful conditions and food deprivation, it may become difficult to maintain these organs and a process known in the literature as “degrowth” occurs, leading to a decrease in the body size of the animal (Baguñà & Romero, 1981; Oviedo et al. 2003). This effect indeed seemed to happen to some individuals in the current experiments, that ended up completely disintegrated. On the other hand, these organs are not present in the posterior fragment of the body, since the animal is still regenerating, so these fragments remain longer, hence they have a longer survival time.

It is also important to take into account that the regenerative capacity is different among the different planarian species (Owlarn & Bartscherer, 2016), which may also influence and explain the results obtained. In fact, information from Pandian (2020) shows that the Order Tricladida, in which *S. mediterranea* is included, has greater regenerative capacity than the Order Polycladida to which the species under study (*Leptoplana* sp.) belongs (Pandian, 2020, pp. 70-71). In addition, it is also mentioned that some species of the Order Polycladida, such as *L. tremellaris*, are capable of regenerating the posterior fragment but not the anterior fragment, and from the 2000 species included

in this Order, only 13 (0.65%) have regenerative capacity, and, as for the Order Tricladida, from the 1000 species that constitute it, 23 (2.3%) have regenerative capacity (Pandian, 2020, pp. 235). However, despite all these observations, in the regeneration experiments with *Leptoplana* sp., neoblasts have proven their extraordinary ability to replace lost and/or damaged cells which indeed demonstrates that they may have interesting therapeutic applications in terms of regenerative medicine and recovery of damaged cells, tissues, and organs.

4.2.2. Molecular changes underlying regeneration

The anatomical and cellular modifications essential for the regeneration process are downstream consequences of changes in the expression of a series of genes. It is known that there are constitutive genes whose expression is practically constant regardless whether the cells of the organism are in normal or pathophysiological conditions. However, on the other hand, there are genes whose expression changes depending on the situation, which are denominated differentially-expressed genes (DEGs). Thus, focusing on these DEGs and comparing the results of the Venn diagrams from the comparisons made in this study (pole piece versus flanking piece and intact planarians versus regenerating planarians), it is possible to verify that in the second set of comparisons (intact versus regenerating planarians, Fig. 3.7.) the total number of DEGs was much higher than in the first case (pole versus flanking, Fig. 3.6.) which makes sense since although the comparison was made between the corresponding regions, the animals were in different conditions, as one was intact and the other in regeneration. The same trend seemed to occur regarding the number of DEGs involved in regeneration (Fig. 3.8. and 3.10.) which was also supported by the results of the heatmap from Fig. 3.11. since the expression patterns of most of the DEGs involved in regeneration was quite similar between the two regions under study (pole versus flanking). All this suggests that it is not the body region that has the greatest impact in gene expression but the condition to which planarians are subjected (intact versus regenerating). The results obtained by Li et al. (2019) revealed that there were more pole-piece-enriched genes in intact animals than in regenerating animals (203 versus 86). This was an opposite result to the one obtained in this study, in which the number of up-regulated genes in the pole piece of regenerating planarians was more than twice of the number of up-regulated genes in pole piece of the intact planarians (79 versus 32, Fig. 3.6.). This difference may be explained by the different objectives of these two works, taking into account that the main focus of Li et al. (2019) was to study the patterning and establishment of the anterior-posterior axis in planarians, which allowed them to identify some putative organizers of the planarian head patterning, such as *notum*, *foxD*, *zic-1* and *follistatin*, which were up-regulated in the pole piece of both uninjured and regenerating animals. On the other hand, the present study had as principal focus the identification of genes and pathways involved in regeneration. Therefore, the following analyzes will be mainly focused on the DEGs involved in regeneration found in the comparisons iii and iv (intact versus regenerating planarians), whose expression levels from Table 3.3. indicated that 72-hours after transverse amputation, most of the genes found to be involved in the process of cell differentiation were up-regulated either in the flanking piece or in the pole piece of the regenerating planarians when compared to the intact planarians, which corroborates what was observed

in the histological analysis of *Leptoplana* sp., indicating that on the fourth day of regeneration there was already differentiation of stem cells occurring.

The results of gene expression from these comparisons also indicated that during the regeneration process there was an investment in DNA replication, and cell proliferation and growth, which was verified through the high expression of some genes involved in these processes, such as *FOXK1*, *PCNA*, *CycD* (*CCND1*) and *MCM2-7* genes. Mini chromosome maintenance (MCM) complex is a component of the pre-replication complex, which in turn is a component of the licensing factor. This MCM complex is a hexamer of six polypeptides (MCM2-7) evolutionarily conserved in all eukaryotes, which constitute a family of DNA helicases that have a role in the initiation and elongation phases of eukaryotic DNA replication (Tye, 1999). It is also the target of various cell cycle checkpoints, for instance in the G1/S transition of the mitotic cell cycle (Bochman & Schwacha, 2009). Due to these functions in which MCM complex is involved, it was expected that all these proteins (MCM2-7) had higher expression levels in regenerating planarians when compared with intact planarians, which was in fact confirmed by the results of Table 3.3., except for *MCM6* in the flanking piece, although its expression level was very close to 1.5 (value from which a gene was considered differentially-expressed). The fact that these MCM proteins are involved in DNA replication, cell proliferation, and cell cycle regulation demonstrates, as it is known, that these processes are closely related since it is essential to have a great control of the cell cycle to prevent an abnormal cell proliferation that may lead to the development of malignancies. Thus, there are some checkpoints that occur throughout the cell cycle, as, for instance, G1/S transition, G2 DNA damage checkpoint, or even cell cycle arrest, which are ensured by the expression of genes such as *PCNA*, *CycD* (*CCND1*), *MCM2-7*, *MCM8*, *PRIM1*, *RAD51*, and *BLM*, as the results demonstrated.

Cell cycle arrest is a checkpoint that stops the cell cycle progression in order to facilitate DNA damage repair before cell proliferation (Lo et al. 2015). Among the up-regulated genes in the regenerating planarians there were several involved in the DNA repair process, namely *PCNA*, which is involved in base-excision repair (BER), nucleotide-excision repair (NER), mismatch repair, and still in the cellular response to UV and other external factors that damage DNA (Shivji et al. 1992; Aboussekhra & Wood, 1995; Maga & Hübscher, 2003; Lehman & Mayo, 2012). In addition to this, there were also genes involved in the repair of double-strand breaks (DSBs), which is one of the most severe forms of DNA damage (Barghouth et al. 2019), such as *BLM*, *TONSL*, and *MCM8* genes. It is also known that TONSL and MCM8 proteins interact with RAD51, which is also involved in DSBs repair via homologous recombination (O'Donnell et al. 2010; Park et al. 2013). Still, despite the existence of multiple active DNA repair pathways, if the DNA damage is too severe and the cell cannot repair it, other signaling mechanisms will be activated to prevent malignancies, for instance, apoptosis or other forms of programmed cell death (Chatterjee & Walker, 2017). Therefore, the expression of some apoptosis-promoting genes was also observed in the tissues of regenerating planarians, such as *ZMAT3* and *Sd-1* (*TEAD1*). *Sd-1* is one of the two homologs of the Scalloped transcription factor identified in *S. mediterranea* (*Smed-sd-1* and *Smed-sd-2*). This transcription factor is the equivalent of TEAD1-4 (TEA Domain Transcription Factors 1-4) in vertebrates. Sd is a DNA-binding transcription factor that associates with Yki (YAP is the vertebrate homolog) to mediate the transcriptional output of the Hippo

growth-regulatory pathway, a pathway involved in stem cell regulation, organ size control, and tumor suppression by restricting proliferation and promoting cell death (Wu et al. 2008; Lin & Pearson, 2014). Therefore, Sd-1 is involved in cell proliferation control through several approaches regulated by the Hippo growth pathway, all of them essential for the regeneration process.

On the other hand, there were also some down-regulated genes found in the comparisons between the regenerating and intact planarians, such as *ZNFX1*, a transcription factor thought to be involved in chromatin silencing, and *PDCD5*, which is a negative regulator of cell proliferation and a promotor of apoptosis (Chen et al. 2006; Ishidate et al. 2018). Thus, despite appearing contradictory to what was said, the negative regulation of these two genes strengthens the fact that there was an investment in cell proliferation and in cell integrity maintenance in the regenerating planarians rather than a promotion of apoptosis, even though during regeneration, cell death also occurs, mediated by apoptotic and autophagic processes in order to remodeling the pre-existing tissues and to adjust body size proportions of the newly formed organism (González-Estévez et al. 2007; Pellettieri et al. 2010).

Making a global assessment of the three major processes mentioned in this study that are related with regeneration – DNA replication, cell cycle regulation, and DNA repair – it is possible to verify that there are four genes involved in all of them, which are *PCNA*, *MCM8*, *RAD51*, and *BLM*. MCM8, RAD51, and BLM proteins are all involved in DSBs repair via homologous recombination. MCM8 is involved in the recruitment of the effector RAD51 to DNA damage sites (Park et al. 2013) and there is also another protein, SPIDR (Scaffolding protein involved in DNA repair) which independently interacts with both BLM helicase and RAD51 recombinase proteins, leading to the formation of BLM/RAD51 complex, essential for the homologous recombination pathway (Wan et al. 2013). Unlike the previous genes, *PCNA* is not involved in DSBs repair, however, it is very important in detecting and repairing DNA lesions through various mechanisms, such as BER and NER as mentioned before. Moreover, PCNA is essential for DNA replication, functioning as an auxiliary protein to DNA polymerase δ, increasing its processability during the elongation phase (Bravo et al. 1987). A peculiarity of PCNA is that this protein is present in proliferating cells but not in differentiated cells, so it can be used as a marker to identify active proliferating cells and monitor its distribution during the regeneration process (Bravo & MacDonald-Bravo, 1985; Garcia et al. 1989). Thus, an additional way to assess the presence and distribution of neoblasts more accurately during the experiments in planarians and to assess the regeneration progress would be through PCNA expression analysis, such as Whole-mount in situ hybridization (WISH) or Whole-mount immunohistochemistry (WIHC), as already performed by Orii et al. (2005).

4.2. Predicting the impact of UV radiation on tissue regeneration and gene networks

Although the UV radiation experiments have not been carried out and their impact on regeneration, in planarians inclusively, is not yet fully understood, even though this is a very relevant topic in several fields of research, from ecology to regeneration and ageing medicine, it would be expected that in planarians, as in other organisms, UV radiation would affect DNA, leading to the formation of several mutagenic and cytotoxic lesions. These lesions would certainly impact normal cells, having consequences in the regeneration process, such as the emergence of some malformations or changes in the normal regeneration time and, therefore, one of the main objectives of these experiments would be to evaluate the capacity of neoblasts to recover damaged cells and tissues and maintain the ability of the organisms to regenerate.

As for the impact of UV radiation at a genetic level, it would be expected that the main genes affected by this environmental agent during the regeneration process would be those with the highest expression values, since UV radiation can lead to the introduction of errors and mutations in DNA, which may not be repaired and produce defective or non-functional proteins. Therefore, among the genes analyzed in this study, it would be expected that the 33 up-regulated genes involved in regeneration would be the most affected by the application of UV radiation on these animals, since, besides their high expression, there are interactions between almost all of them. Moreover, since most of these genes are involved in several processes essential for regeneration, this aptitude of planarians would be compromised. However, taking into account some interesting genes found in this study, such as *PCNA*, *BLM*, and *RAD51*, which seemed to indicate that these planarians are able to repair lesions in DNA and also to respond to radiation exposure (Shivji et al. 1992; Aboussekhra & Wood, 1995; Maga & Hübscher, 2003; Park et al. 2013; Wan et al. 2013), these abilities could counterbalance this trend. This ability of DNA repair may be comparable to that existing in other organisms, since planarians display a high functional and evolutionary conservation of the DNA repair mechanisms and signaling pathways that regulate adult stem cells (Barghouth et al. 2019). Therefore, after UV exposure, it would be expected that the nucleotide-excision repair (NER) mechanism would be activated, since the main damages caused by UV are dimers and adducts.

5. Conclusions and final remarks

The present study revealed that regeneration is a very complex process, not always successful, involving a series of genes and pathways which determinate the fate of cells, namely for neoblasts. The ability of these cells to replace several cell types of the organism and to resist to the damages caused by amputation, even in species with more limited regenerative capacity, has been proved. This demonstrated that, in fact, *Leptoplana* sp. has regenerative capacity, although it is not as evident nor as fast or successful as in other planarian species such as *S. mediterranea* or *G. tigrina*, although there are many similarities between the processes that occur in these species.

This work also helped to evidence the importance and interest of neoblasts and regeneration in potential therapeutic applications for regenerative medicine and recovery of damaged cells, tissues, and organs. Thus, a potential future study could be focused on genetically engineering these adult pluripotent stem cells in vitro and then evaluate their capacity of regenerating damaged tissues from other organisms with more limited regenerative capacities, such as mammals.

The current findings also indicated that during regeneration there was an investment in maintaining the normal functioning of certain pathways and processes, such as DNA replication, cell proliferation and growth, and cell differentiation, as well as in regulating all these processes through the control of cell cycle and the repair of damages in DNA in order to avoid the emergence of problems and disorders during this process. On the other hand, there was a negative regulation of unnecessary processes related to cell death, which demonstrated the effort of planarians in maintaining their integrity. It has also been shown that all these molecular processes observed during regeneration in *S. mediterranea* were also observed at a morphological level in regeneration of *Leptoplana* sp., namely in what concerns proliferation and differentiation of neoblasts and cell and tissue healing and reorganization.

Finally, planarians, which were initially seen as a model to study tissue regeneration and stem cell manipulation, may also represent an alternative to understand and manipulate DNA damage, since there are several mechanisms of DNA repair in planarians that can be studied.

6. References

Aboussekhra, A., & Wood, R. D. (1995). Detection of nucleotide excision repair incisions in human fibroblasts by immunostaining for PCNA. *Experimental Cell Research*, 221(2), 326-332. https://doi.org/10.1006/excr.1995.1382

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Alvarado, A. S. (2004). Planarians. *Current Biology*, 14(18), R737-R738. https://doi.org/10.1016/j.cub.2004.09.005

Baguñà, J. (1976). Mitosis in the intact and regenerating planarian *Dugesia mediterranea* n. sp. II. Mitotic studies during regeneration, and a possible mechanism of blastema formation. *Journal of Experimental Zoology*, 195(1), 65-79. https://doi.org/10.1002/jez.1401950107

Baguñà, J., & Romero, R. (1981). Quantitative analysis of cell types during growth, degrowth and regeneration in the planarians *Dugesia mediterranea* and *Dugesia tigrina*. *Hydrobiologia*, 84(1), 181-194. https://doi.org/10.1007/BF00026179

Barghouth, P. G., Thiruvalluvan, M., LeGro, M., & Oviedo, N. J. (2019). DNA damage and tissue repair: What we can learn from planaria. In *Seminars in Cell & Developmental Biology* (Vol. 87, pp. 145-159). Academic Press. https://doi.org/10.1016/j.semcdb.2018.04.013

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, 41(D1), D36-D42. https://doi.org/10.1093/nar/gks1195

Bochman, M. L., & Schwacha, A. (2009). The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiology and Molecular Biology Reviews: MMBR*, 73(4), 652-683. https://doi.org/10.1128/MMBR.00019-09

Bravo, R., & Macdonald-Bravo, H. (1985). Changes in the nuclear distribution of cyclin (PCNA) but not its synthesis depend on DNA replication. *The EMBO Journal*, 4(3), 655-661.

Bravo, R., Frank, R., Blundell, P. A., & Macdonald-Bravo, H. (1987). Cyclin/PCNA is the auxiliary protein of DNA polymerase-delta. *Nature*, 326(6112), 515-517. https://doi.org/10.1038/326515a0

Bybee, D. R., & Murray, J. M. (2017). Polychaetes. In *Marine Ornamental Species Aquaculture.* (Eds. R. Calado, I. Olivotto, M. P. Oliver, & G. J. Holt). (pp. 573-574). John Wiley & Sons. https://doi.org/10.1002/9781119169147.ch24

Caplan, A. I. (1991). Mesenchymal stem cells. *Journal of Orthopaedic Research*, 9(5), 641-650. https://doi.org/10.1002/jor.1100090504

Cebrià, F., Adell, T., & Saló, E. (2018). Rebuilding a planarian: from early signaling to final shape. *International Journal of Developmental Biology*, 62(6-7-8), 537-550. https://doi.org/10.1387/ijdb.180042es

Chandebois, R. (1980). The dynamics of wound closure and its role in the programming of planarian regeneration. II—Distalization. *Development, Growth & Differentiation*, 22(4), 693-704. https://doi.org/10.1111/j.1440-169X.1980.00693.x

Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis*, 58(5), 235-263. https://doi.org/10.1002/em.22087

Chen, L. N., Wang, Y., Ma, D. L., & Chen, Y. Y. (2006). Short interfering RNA against the PDCD5 attenuates cell apoptosis and caspase-3 activity induced by Bax overexpression. *Apoptosis*, 11(1), 101-111. https://doi.org/10.1007/s10495-005-3134-y

D'Orazio, J., Jarrett, S., Amaro-Ortiz, A., & Scott, T. (2013). UV radiation and the skin. *International Journal of Molecular Sciences*, 14(6), 12222-12248. https://doi.org/10.3390/ijms140612222

Fields, C., & Levin, M. (2020). Does regeneration recapitulate phylogeny? Planaria as a model of body-axis specification in ancestral eumetazoa. *Communicative & Integrative Biology*, 13(1), 27-38. https://doi.org/10.1080/19420889.2020.1729601

Garcia, R. L., Coltrera, M. D., & Gown, A. M. (1989). Analysis of proliferative grade using anti-PCNA/cyclin monoclonal antibodies in fixed, embedded tissues. Comparison with flow cytometric analysis. *The American Journal of Pathology*, 134(4), 733-739.

Gentile, L., Cebrià, F., & Bartscherer, K. (2011). The planarian flatworm: an in vivo model for stem cell biology and nervous system regeneration. *Disease Models & Mechanisms*, 4(1), 12-19. https://doi.org/10.1242/dmm.006692

Gibson, R., & Knight-Jones, E. W. (2017). Flatworms and Ribbon Worms. In *Handbook of the Marine Fauna of North-West Europe*. (Eds. P. J. Hayward, & J. S. Ryland). (pp. 135-141). Oxford University Press. http://doi.org/10.1093/acprof:oso/9780199549443.003.0005

González-Estévez, C., Felix, D. A., Aboobaker, A. A., & Saló, E. (2007). Gtdap-1 promotes autophagy and is required for planarian remodeling during regeneration and starvation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(33), 13373-13378. https://doi.org/10.1073/pnas.0703588104

Häder, D. P., & Sinha, R. P. (2005). Solar ultraviolet radiation-induced DNA damage in aquatic organisms: potential environmental impact. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 571(1-2), 221-233. https://doi.org/10.1016/j.mrfmmm.2004.11.017

Hoeijmakers, J. H. (2009). DNA damage, aging, and cancer. *New England Journal of Medicine*, 361(15), 1475-1485. https://doi.org/10.1056/NEJMra0804615

Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314. https://www.tandfonline.com/doi/abs/10.1080/10618600.1996.10474713

Ishidate, T., Ozturk, A. R., Durning, D. J., Sharma, R., Shen, E. Z., Chen, H., Seth, M., Shirayama, M., & Mello, C. C. (2018). ZNFX-1 functions within perinuclear nuage to balance epigenetic signals. *Molecular Cell*, 70(4), 639-649. https://doi.org/10.1016/j.molcel.2018.04.009

Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature*, 461(7267), 1071-1078. https://doi.org/10.1038/nature08467

Knakievicz, T., & Ferreira, H. B. (2008). Evaluation of copper effects upon *Girardia tigrina* freshwater planarians based on a set of biomarkers. *Chemosphere*, 71(3), 419-428. https://doi.org/10.1016/j.chemosphere.2007.11.004

Koronakis, P. S., Sfantos, G. K., Paliatsos, A. G., Kaldellis, J. K., Garofalakis, J. E., & Koronaki, I. P. (2002). Interrelations of UV-global/global/diffuse solar irradiance components and UV-global attenuation on air pollution episode days in Athens, Greece. *Atmospheric Environment*, 36(19), 3173-3181. https://doi.org/10.1016/S1352-2310(02)00233-9

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), 1547-1549. https://doi.org/10.1093/molbev/msy096

Lehman, J. A., & Mayo, L. D. (2012). Integration of DNA damage and repair with murine double-minute 2 (Mdm2) in tumorigenesis. *International Journal of Molecular Sciences*, 13(12), 16373-16386. https://doi.org/10.3390/ijms131216373

Li, D. J., McMann, C. L., & Reddien, P. W. (2019). Nuclear receptor NR4A is required for patterning at the ends of the planarian anterior-posterior axis. *eLife*, 8, e42015. https://doi.org/10.7554/eLife.42015.001

Lin, A. Y., & Pearson, B. J. (2014). Planarian yorkie/YAP functions to integrate adult stem cell proliferation, organ homeostasis and maintenance of axial patterning. *Development*, 141(6), 1197-1208. https://doi.org/10.1242/dev.101915

Lo, A., Serravallo, M., & Jagdeo, J. (2015). Epigenetic Mechanisms of Sirtuins in Dermatology. In *Epigenetics and Dermatology.* (Eds. Q. Lu, C. C. Chang, & B. C. Richardson). (pp. 137-175). Academic Press. https://doi.org/10.1016/B978-0-12-800957-4.00008-4

Lobo, J., Costa, P. M., Teixeira, M. A., Ferreira, M. S., Costa, M. H., & Costa, F. O. (2013). Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. *BMC Ecology*, 13(1), 34. https://doi.org/10.1186/1472-6785-13-34

Lord, C. J., & Ashworth, A. (2012). The DNA damage response and cancer therapy. *Nature*, 481(7381), 287-294. https://doi.org/10.1038/nature10760

Maga, G., & Hübscher, U. (2003). Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *Journal of Cell Science*, 116(15), 3051-3060. https://jcs.biologists.org/content/116/15/3051

Mao, A. S., & Mooney, D. J. (2015). Regenerative medicine: current therapies and future directions. *Proceedings of the National Academy of Sciences of the United States of America*, 112(47), 14452-14459. https://doi.org/10.1073/pnas.1508520112

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10), 4288-4297. https://doi.org/10.1093/nar/gks042

Mehta, A. S., & Singh, A. (2019). Insights into regeneration tool box: An animal model approach. *Developmental Biology*, 453(2), 111-129. https://doi.org/10.1016/j.ydbio.2019.04.006

Minguell, J. J., Erices, A., & Conget, P. (2001). Mesenchymal stem cells. *Experimental Biology and Medicine*, 226(6), 507-520. https://doi.org/10.1177/153537020122600603

Morgan, T. H. (1898). Experimental studies of the regeneration of *Planaria maculata*. *Roux's Archives of Developmental Biology*, *7*(2-3), 364-397.

Newmark, P. A., & Alvarado, A. S. (2002). Not your father's planarian: a classic model enters the era of functional genomics. *Nature Reviews Genetics*, 3(3), 210-219. https://doi.org/10.1038/nrg759

O'Donnell, L., Panier, S., Wildenhain, J., Tkach, J. M., Al-Hakim, A., Landry, M. C., Escribano-Diaz, C., Szilard, R.K., Young, J.T., Munro, M., & Canny, M. D. (2010). The MMS22L-TONSL complex mediates recovery from replication stress and homologous recombination. *Molecular Cell*, 40(4), 619-631. https://doi.org/10.1016/j.molcel.2010.10.024

Orii, H., Sakurai, T., & Watanabe, K. (2005). Distribution of the stem cells (neoblasts) in the planarian *Dugesia japonica*. *Development Genes and Evolution*, 215(3), 143-157. https://doi.org/10.1007/s00427-004-0460-y

Oviedo, N. J., Nicolas, C. L., Adams, D. S., & Levin, M. (2008). Planarians: a versatile and powerful model system for molecular studies of regeneration, adult stem cell regulation, aging, and behavior. *Cold Spring Harbor Protocols*, 2008(10), pdb-emo101. https://doi.org/10.1101/pdb.emo101

Oviedo, N. J., Newmark, P. A., & Sánchez Alvarado, A. (2003). Allometric scaling and proportion regulation in the freshwater planarian *Schmidtea mediterranea*. *Developmental Dynamics*, 226(2), 326-333. https://doi.org/10.1002/dvdy.10228

Owlarn, S., & Bartscherer, K. (2016). Go ahead, grow a head! A planarian's guide to anterior regeneration. *Regeneration*, 3(3), 139-155. https://doi.org/10.1002/reg2.56

Pandian, T. J. (2020). Turbellaria. Comparison and Highlights. In *Reproduction and Development in Platyhelminthes (Reproduction and Development in Aquatic Invertebrates)*. (pp. 69-71; 235). CRC Press.

Park, J., Long, D. T., Lee, K. Y., Abbas, T., Shibata, E., Negishi, M., Luo, Y., Schimenti, J. C., Gambus, A., Walter, J. C., & Dutta, A. (2013). The MCM8-MCM9 complex promotes RAD51 recruitment at DNA

damage sites to facilitate homologous recombination. *Molecular and Cellular Biology*, 33(8), 1632-1644. https://doi.org/10.1128/MCB.01503-12

Pattison, D. I., & Davies, M. J. (2006). Actions of ultraviolet light on cellular structures. In *Cancer: Cell Structures, Carcinogens and Genomic Instability.* Experientia Supplementum. (Eds. L. P. Bignold). (pp. 131-157). Birkhäuser Basel. https://doi.org/10.1007/3-7643-7378-4_6

Pellettieri, J., Fitzgerald, P., Watanabe, S., Mancuso, J., Green, D. R., & Sánchez Alvarado, A. (2010). Cell death and tissue remodeling in planarian regeneration. *Developmental Biology*, 338(1), 76-85. https://doi.org/10.1016/j.ydbio.2009.09.015

Randolph, H. (1897). Observations and experiments on regeneration in planarians. *Archiv für Entwicklungsmechanik der Organismen*, 5(2), 352-372. https://doi.org/10.1007/BF02162271

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, 7(3), 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ravanat, J. L., & Douki, T. (2016). UV and ionizing radiations induced DNA damage, differences and similarities. *Radiation Physics and Chemistry*, 128, 92-102. https://doi.org/10.1016/j.radphyschem.2016.07.007

Reddien, P. W., & Alvarado, A. S. (2004). Fundamentals of planarian regeneration. *Annual Review of Cell and Developmental Biolo*gy, 20, 725-757. https://doi.org/10.1146/annurev.cellbio.20.010403.095114

Reddien, P. W., Bermange, A. L., Murfitt, K. J., Jennings, J. R., & Alvarado, A. S. (2005). Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Developmental Cell*, 8(5), 635-649. https://doi.org/10.1016/j.devcel.2005.02.014

Ritchie, M. E., Phipson, B., Wu, D. I., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47-e47. https://doi.org/10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. https://doi.org/10.1093/bioinformatics/btp616

Saló, E., & Baguñà, J. (1984). Regeneration and pattern formation in planarians: I. The pattern of mitosis in anterior and posterior regeneration in *Dugesia (G) tigrina*, and a new proposal for blastema formation. *Development*, 83(1), 63-80.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671-675. https://doi.org/10.1038/nmeth.2089

Shivji, K. K., Kenny, M. K., & Wood, R. D. (1992). Proliferating cell nuclear antigen is required for DNA excision repair. *Cell*, 69(2), 367-374. https://doi.org/10.1016/0092-8674(92)90416-a

Sinha, R. P., Singh, N., Kumar, A., Kumar, H. D., Häder, M., & Häder, D. P. (1996). Effects of UV irradiation on certain physiological and biochemical processes in cyanobacteria. *Journal of Photochemistry and Photobiology B: Biology*, 32(1-2), 107-113. https://doi.org/10.1016/1011-1344(95)07205-5

Stoick-Cooper, C. L., Moon, R. T., & Weidinger, G. (2007). Advances in signaling in vertebrate regeneration as a prelude to regenerative medicine. *Genes & Development*, 21(11), 1292-1315. https://doi.org/10.1101/gad.1540507

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1), D607-D613. https://doi.org/10.1093/nar/gky1131

Tye B. K. (1999). MCM proteins in DNA replication. *Annual Review of Biochemistry*, 68, 649-686. https://doi.org/10.1146/annurev.biochem.68.1.649

Wan, L., Han, J., Liu, T., Dong, S., Xie, F., Chen, H., & Huang, J. (2013). Scaffolding protein SPIDR/KIAA0146 connects the Bloom syndrome helicase with homologous recombination repair. *Proceedings of the National Academy of Sciences of the United States of America*, 110(26), 10646-10651. https://doi.org/10.1073/pnas.1220921110

Wu, S., Liu, Y., Zheng, Y., Dong, J., & Pan, D. (2008). The TEAD/TEF family protein Scalloped mediates transcriptional output of the Hippo growth-regulatory pathway. *Developmental Cell*, 14(3), 388-398. https://doi.org/10.1016/j.devcel.2008.01.007

7. Annexes

Annex 7.1. A) Consensus sequence of Lep1 (609 nts). **B)** Consensus sequence of Lep2 (625 nts). **C)** Consensus sequence of the unidentified planarian (U) (543 nts).

A)

GGTACTTTATACCTACTATTCGGAATATGATCAGGGCTTGTAGGTACAGCCTTTAGTTTTCTTATACGTTCAGAGTTATCCCA
ACCTGGCAGTATACTAAAAGATTCACAATTATATAATAGAATTATTACGGCACATGGCCTTATAATGATTTTCTTCTTTGTGA
TGCCTGTACTAATTGGTGGTTTTGGTAACTGGTTAATACCTTTATATTTAACTAGACCGGATATGGCCTTCCCCCGCCTTAAA
AATATGAGGTTTTGGCTTTTGCCACCAGCCTTATTTTTATTACTAGGATCTTTCGTAGTTGAAGGTGGTGTGGGAACAGGTTG
AACAATATATCCTCCTTTATCGTCAAATATAGCCCAAAGAGGACCAAGGGTGGATTTAGCTATATTTTCTTTACACCTCGCAG
GTGTGAGGTCAATTCTAGGCTCAATTAAATTCATTACTACAATAGTGAATGCCAAGATCCAAGTTTCATGAGGCCAATTACCT
TTATTCCTATGGGCTGTTATGGTTACCGCCTATATGCTGGTTCTTTCTCTGCCCGTTTTGGCTGGTGGCCTAACAATGTTGTT
AACAGATCGAAAATTTAATACAACATTT

B)

TCGGTACTTTATACCTACTATTCGGAATATGATCAGGGCTTGTAGGTACAGCCTTTAGTTTTCTTATACGTTCAGAGTTATCC
CAACCTGGCAGTATACTAAAAGATTCACAATTATATAATAGAATTATTACGGCACATGGCCTTATAATGATTTTCTTCTTTGT
GATGCCTGTACTAATTGGTGGTTTTGGTAACTGGATAATACCTCTATATTTAACTAGACCGGATATGGCCTTCCCCCGCCTTA
AAAATATGAGGTTTTGGCTTTTGCCACCAGCCTTATTTTTATTACTAGGATCTTTCGTAGTTGAAGGTGGTGTGGGAACAGGT
TGAACAATATATCCTCCTTTATCGTCAAATATAGCCCAAAGAGGACCAAGGGTGGATTTAGCTATATTTTCTTTACACCTCGC
AGGTGTAAGGTCAATTCTAGGCTCAATTAAATTCATTACTACAATAGTGAATGCCAAGATCCAAGTTTCATGAGGCCAATTAC
CTTTATTCCTATGGGCTGTTATGGTTACCGCCTATATGCTGGTTCTTTCTCTGCCCGTTTTGGCCGGTGGCCTAACAATGTTG
TTAACAGATCGAAAATTTAATACAACATTTTTCGATCCGGGAGG

C)

TATAAAAGGGTTATTACAGCACATGGCTTAGTAATGATATTTTTTTTTGTAATGCCGGTTATGATAGGAGGTTTCGGAAACTG
ACTGATACCTATTTATATGGGAGTAGCAGATATGAACTTTCCTCGGCTGAAAAATTTAAGTTTATGATTACAAATGCCTTCAG
TTACTTTACTAATAGGTTCATTCTTTGCAAGTAAAGGAGTAGGGGGTGGTTGAACTATCTACCCTCCTTTATCAAGAGGGATA
GCTCATGCAGGGAGTAGGGTGGATATGGCTATATTCTCCTTACATCTTGCTGGGGCGAGTTCAATATTAGGGTCGATTAAATT
TATTAGCACCATAGGCCAGAGTAATAAGAGGGGTATGGCATGGTATCGTTTCCCATTATTTATTTGGGCCATGACGATAACAG
CTTACATGTTAGTACTTTCCTTACCCGTATTGGCAGCTGGTATAACAATGTTATTAACAGATCGTAAATTTAATACAACATTC
TTCGACCCCTCTGGGGGAGGGGACCCAATATTATTCCAACATATA

Annex 7.2. Taxonomic identification of Lep1 sample adapted from Bold Systems analysis.

Phylum	Class	Order	Family	Genus	Similarity (%)
Platyhelminthes					77.66
Platyhelminthes	Rhabditophora	Polycladida			76.9
Platyhelminthes	Rhabditophora	Polycladida	Leptoplanidae	*Hoploplana*	76.46

Annex 7.3. Taxonomic identification of Lep1 sample adapted from GenBank analysis.

Classification	Number of Hits	Score
Phylum Platyhelminthes	100	
Order Polycladida	98	
Suborder Acotylea	94	
unclassified Acotylea	19	
Acotylea sp. WtP 1	8	523

Annex 7.4. Specimen identification of Lep1 sample adapted from GenBank analysis.

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Acotylea sp. WtP 1 isolate WtP-19-10 cytochrome c oxidase subunit I (*COI*) gene, partial cds; mitochondrial	523	523	99%	$7e^{-146}$	79.08%	MN013763.1
Acotylea sp. WtP 1 isolate WtP-19-14 cytochrome c oxidase subunit I (*COI*) gene, partial cds; mitochondrial	518	518	99%	$9e^{-145}$	78.91%	MN013766.1
Acotylea sp. WtP 1 isolate WtP-19-15 cytochrome c oxidase subunit I (*COI*) gene, partial cds; mitochondrial	516	516	99%	$1e^{-143}$	78.94%	MN013767.1
Acotylea sp. WtP 1 isolate WtP-19-13 cytochrome c oxidase subunit I (*COI*) gene, partial cds; mitochondrial	514	514	99%	$4e^{-143}$	78.75%	MN013765.1
Acotylea sp. WtP 1 isolate WtP-19-11 cytochrome c oxidase subunit I (*COI*) gene, partial cds; mitochondrial	514	514	99%	$4e^{-143}$	78.75%	MN013764.1

Annex 7.5. Taxonomic identification of Lep2 sample adapted from Bold Systems analysis.

Phylum	Class	Order	Family	Genus	Species	Similarity (%)
Platyhelminthes						78.27
Platyhelminthes	Rhabditophora	Polycladida	Leptoplanidae	*Hoploplana*		78.27
Platyhelminthes	Rhabditophora	Polycladida				77.94
Platyhelminthes	Rhabditophora	Polycladida	Stylochidae	*Stylochus*	*ellipticus*	77.45
Platyhelminthes	Rhabditophora	Polycladida	Notoplanidae	*Notocomplana*	*septentrionalis*	77.29

Annex 7.6. Taxonomic identification of Lep2 sample adapted from GenBank analysis.

Classification	Number of Hits	Score
Phylum Platyhelminthes	83	
Order Polycladida	81	
Suborder Acotylea	78	
Euplana gracilis	1	526
Acotylea sp. WtP 1	8	524
Notocomplana septentrionalis	13	521

Annex 7.7. Specimen identification of Lep2 sample adapted from GenBank analysis.

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Euplana gracilis voucher SNM:IZ:1287161 cytochrome oxidase subunit 1 (*COI*) gene, partial cds; mitochondrial	526	526	99%	$7e^{-147}$	78.74%	KU905935.1
Acotylea sp. WtP 1 isolate WtP-19-10 cytochrome c oxidase subunit I (*COI*) gene, partial cds; mitochondrial	524	1048	98%	$1e^{-146}$	78.86%	MN013763.1
Acotylea sp. WtP 1 isolate WtP-19-14 cytochrome c oxidase subunit I (*COI*) gene, partial cds; mitochondrial	521	1043	99%	$3e^{-145}$	78.58%	MN013766.1
Acotylea sp. WtP 1 isolate WtP-19-08 cytochrome c oxidase subunit I (*COI*) gene, partial cds; mitochondrial	517	1034	99%	$3e^{-144}$	78.42%	MN013762.1
Notocomplana septentrionalis mitochondrial *COI* gene for cytochrome c oxidase subunit I, partial cds, isolate: YO20160606_02	517	517	99%	$3e^{-144}$	78.42%	LC176030.1

Annex 7.8. Taxonomic identification of the unidentified planarian adapted from Bold Systems analysis.

Phylum	Class	Order	Family	Genus	Species	Similarity (%)
Platyhelminthes						80.56
Platyhelminthes	Rhabditophora	Polycladida	Prosthiostomidae	*Prosthiostomum*	*siphunculus*	79.63
Platyhelminthes	Rhabditophora	Polycladida				77.94
Platyhelminthes	Rhabditophora	Polycladida	Prosthiostomidae			71.6

Annex 7.9. Taxonomic identification of the unidentified planarian adapted from GenBank analysis.

Classification	Number of Hits	Score
Phylum Platyhelminthes	105	
Class Rhabditophora	98	
Order Polycladida	76	
Suborder Cotylea	46	
Family Prosthiostomidae	5	
Prosthiostomum siphunculus	1	534

Annex 7.10. Specimen identification the unidentified planarian adapted from GenBank analysis.

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Prosthiostomum siphunculus mitochondrion, complete genome	534	534	100%	$4e^{-149}$	81.77%	KT363736.1
Pseudoceros sp. Ps-18-02 cytochrome c oxidase subunit I (*COI*) gene, partial cds; mitochondrial	361	361	99%	$1e^{-97}$	74.91%	MN013747.1
Pseudoceros sp. m SCK-2017 isolate PCDA_KUO_271 cytochrome oxidase subunit I gene, partial cds; mitochondrial	347	347	100%	$3e^{-93}$	74.50%	MK820717.1
Pseudoceros sp. c SCK-2019 isolate PCDA_KUO_306 cytochrome oxidase subunit I gene, partial cds; mitochondrial	347	347	99%	$3e^{-93}$	74.17%	MK820715.1
Pseudoceros sp. m SCK-2017 isolate PCDA_KUO_231 cytochrome oxidase subunit I (*COI*) gene, partial cds; mitochondrial	347	347	100%	$3e^{-93}$	74.50%	KY421506.1

Annex 7.11. R Script for the bioinformatics analysis of the RNA-seq data.

```
#Package loading
library(tximport)
library(limma)
library(edgeR)
library(seqinr)
library(gplots)
library(RColorBrewer)

#Set working directory
folder<-"C:/smed/"
setwd("C:/smedFiles/")

#Create table containing the experimental design
sampleName<-c(
 "D14-783 [Ai1]",
 "D14-784 [Ai2]",
 "D14-785 [Ai3]",
 "D14-786 [NAi1]",
 "D14-787 [NAi2]",
 "D14-788 [NAi3]",
 "D14-1368 [Ar1]",
 "D14-1369 [Ar2]",
 "D14-1370 [Ar3]",
 "D14-1371 [NAr1]",
 "D14-1372 [NAr2]",
 "D14-1373 [NAr3]"
)
treatment<-c(
 "uninjured or regenerating animal: uninjured",
 "uninjured or regenerating animal: uninjured",
 "uninjured or regenerating animal: uninjured",
 "uninjured or regenerating animal: uninjured",
 "uninjured or regenerating animal: uninjured",
 "uninjured or regenerating animal: uninjured",
 "uninjured or regenerating animal: regenerating animal 72-hours-post transverse amputation",
 "uninjured or regenerating animal: regenerating animal 72-hours-post transverse amputation",
 "uninjured or regenerating animal: regenerating animal 72-hours-post transverse amputation",
 "uninjured or regenerating animal: regenerating animal 72-hours-post transverse amputation",
 "uninjured or regenerating animal: regenerating animal 72-hours-post transverse amputation",
 "uninjured or regenerating animal: regenerating animal 72-hours-post transverse amputation"
)
tissue<-c(
 "tissue: Pole-containing head tip region",
 "tissue: Pole-containing head tip region",
 "tissue: Pole-containing head tip region",
 "tissue: Non-pole-containing head tip flanking region",
 "tissue: Non-pole-containing head tip flanking region",
 "tissue: Non-pole-containing head tip flanking region",
 "tissue: Pole-containing anterior blastema region",
 "tissue: Pole-containing anterior blastema region",
 "tissue: Pole-containing anterior blastema region",
 "tissue: Non-pole-containing anterior blastema flanking region",
 "tissue: Non-pole-containing anterior blastema flanking region",
 "tissue: Non-pole-containing anterior blastema flanking region"
)
smedLevel<-c(
 "Ai",
 "Ai",
 "Ai",
 "NAi",
 "NAi",
 "NAi",
 "Ar",
 "Ar",
 "Ar",
 "NAr",
 "NAr",
 "NAr"
)
geoAccession<-c(
 "GSM3425017",
 "GSM3425018",
 "GSM3425019",
 "GSM3425020",
```

```
 "GSM3425021",
 "GSM3425022",
 "GSM3425023",
 "GSM3425024",
 "GSM3425025",
 "GSM3425026",
 "GSM3425027",
 "GSM3425028"
)
runNumber<-c(
 "SRR7992066",
 "SRR7992067",
 "SRR7992068",
 "SRR7992069",
 "SRR7992070",
 "SRR7992071",
 "SRR7992072",
 "SRR7992073",
 "SRR7992074",
 "SRR7992075",
 "SRR7992076",
 "SRR7992077"
)
dataMatrix<-as.data.frame(
 cbind(
  sampleName,
  smedLevel,
  geoAccession,
  runNumber,
  treatment,
  tissue
 )
)

write.table(dataMatrix, "dataMatrix.csv", sep=";", col.names=NA)

#Create file containing the abundance information of the 12 samples in H5 files
samples<-dir(folder)
file<-c(paste(sep="",folder,samples,"/","abundance.h5"))
names(file)=samples
names(file)

#Convert raw data from H5 or tsv files into contigs
smedh5<-tximport(file, type = "kallisto", txOut = TRUE, countsFromAbundance =
"lengthScaledTPM")
head(smedh5$counts)
dim(smedh5$counts)
colnames(smedh5$counts)

#Add the sequence of each gene to the respective "name" of the contigs (e.g. SMED30000001)
using the reference transcriptome (in fasta format)
file<-"smed_20140614.nt.gz"
smedFa<-read.fasta(file, as.string = TRUE)
head(smedFa)
smedTxID<-getName(smedFa)
smedTxSEQ<-unlist(getSequence(smedFa,as.string=TRUE))
smedTxSEQ<-as.data.frame(cbind(smedTxID,smedTxSEQ))
head(smedTxSEQ)
colnames(smedTxSEQ)[1]<-"ID"
nrow(smedTxSEQ)
smedCounts<-smedh5$counts
smedCounts<-cbind(as.data.frame(row.names(smedCounts)),smedCounts)
colnames(smedCounts)[1]<-"ID"
smedFull <- merge(smedCounts,smedTxSEQ,by="ID")
colnames(smedFull)<-c("ID", sampleName, "sequence")
write.table(smedFull, "smedFull.csv", sep=";", col.names=NA)
#smedFull contains the raw data of the number of reads per contig of each of the 36 035 genes
in study. The columns contain the ID´s of the genes, samples' name and gene sequences of each
contig


#Statistics
#Create an object (smedData) containing the list of differentially-expressed genes to be used
in the statistical analysis
smedData<-DGEList(counts=smedFull[,2:13], group=smedLevel,genes=smedFull[,1])
```

```
smedData<-calcNormFactors(smedData)
smedDesign<-model.matrix(~0+smedLevel,data=smedData$samples)
colnames(smedDesign) <- levels(smedData$samples$group)
smedData<-estimateDisp(smedData,smedDesign)
smedfit <- glmFit(smedData,smedDesign)

#Deisgn of the comparisons to be made
#First comparison
#Make paired comparisons using the expression levels from the flanking piece of both
planarians as a control
AivNAi <- makeContrasts(Ai-NAi, levels=smedDesign)
ArvNAr <- makeContrasts(Ar-NAr, levels=smedDesign)
#Estimate the relative expression for each comparison using the glmLRT function
smedlrtAi <- glmLRT(smedfit,contrast=AivNAi)
topTags(smedlrtAi)
smedlrtAr <- glmLRT(smedfit,contrast=ArvNAr)
topTags(smedlrtAr)
#Create a table with the expression data (and the respective parameters)
smedResults<-cbind(smedFull[,1], smedlrtAi$table, smedlrtAr$table)
colnames(smedResults)<-c("ID", "logFC-Ai", "logCPM-Ai","LR-Ai","FDRp-Ai", "logFC-Ar", "logCPM-
Ar","LR-Ar","FDRp-Ar")
write.table(smedResults,"smedResults.csv",sep=";",col.names=NA)

#Second comparison
#Make paired comparisons using the expression levels of the intact planarians as a control
NArvNAi <- makeContrasts(NAr-NAi, levels=smedDesign)
ArvAi <- makeContrasts(Ar-Ai, levels=smedDesign)
smedlrtNA <- glmLRT(smedfit,contrast=NArvNAi)
topTags(smedlrtNA)
smedlrtA <- glmLRT(smedfit,contrast=ArvAi)
topTags(smedlrtA)
smedResults2<-cbind(smedFull[,1], smedlrtNA$table, smedlrtA$table)
colnames(smedResults2)<-c("ID", "logFC-NA", "logCPM-NA","LR-NA","FDRp-NA", "logFC-A", "logCPM-
A","LR-A","FDRp-A")
write.table(smedResults2,"smedResults2.csv",sep=";",col.names=NA)


#Post treatment
#First comparison
#Assignment of values 0, 1 and -1 according to the relative expression levels between the pole
and the flanking piece (meaning unaltered, over- or under-expressed, respectively), using as
criteria a significant FDR adjusted p-value and logFC values ≥ |1.5| and then create a sub
table only with the up and down regulated genes (1 and -1, respectively)
expressionTable<-decideTests(smedResults[,grepl("FDRp",
colnames(smedResults))],coefficients=smedResults[,grepl("logFC",
colnames(smedResults))],lfc=1.5,adjust.method="fdr")
colnames(expressionTable)<-c("Ai","Ar")
head(expressionTable)
nrow(expressionTable)
#Merge the expression data and the annotation data
smedDEG<-cbind(smedResults,expressionTable)
degTable<-smedDEG[which(abs(smedDEG$Ai) == 1 | abs(smedDEG$Ar) == 1),]
head(degTable)
nrow(degTable)
write.table(degTable,"degTable.csv",sep=";",col.names=NA)

#Second comparison
#Assignment of values 0, 1 and -1 according to the relative expression levels between the
regenerating and intact planarians, using as criteria a significant FDR adjusted p-value and
logFC values ≥ |1.5| and then create a sub table only with the up and down regulated genes (1
and -1, respectively)
expressionTable2<-decideTests(smedResults2[,grepl("FDRp",
colnames(smedResults2))],coefficients=smedResults2[,grepl("logFC",
colnames(smedResults2))],lfc=1.5,adjust.method="fdr")
colnames(expressionTable2)<-c("nA","A")
head(expressionTable2)
nrow(expressionTable2)
smedDEG2<-cbind(smedResults2,expressionTable2)
degTable2<-smedDEG2[which(abs(smedDEG2$nA) == 1 | abs(smedDEG2$A) == 1),]
head(degTable2)
nrow(degTable2)
write.table(degTable2,"degTable2.csv",sep=";",col.names=NA)
```

```
#Venn diagram plots
#First comparison
vennCounts(degTable)
windows()
vennDiagram(
 degTable[,c("Ai", "Ar")],
 names=c("DEGs from\ncomparison i\n","DEGs from\ncomparison ii\n"),
 include=c("up", "down"),
 counts.col=c("red3","dodgerblue3"),
 circle.col=NULL,
 show.include=TRUE,
 cex=c(1.3,1.3,1.3),
 lwd=1.5)

fileName<-"Venn_i_ii.tif"
path<-"C:/smedFiles/"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 22,
          width = 22,
          units = 'cm',
          type="windows",
          res=600)


#Second comparison
vennCounts(degTable2)
windows()
vennDiagram(
  degTable2[,c("nA", "A")],
  names=c("DEGs from\ncomparison iii\n", "DEGs from\ncomparison iv\n"),
  include=c("up", "down"),
  counts.col=c("red3","dodgerblue3"),
  circle.col=NULL,
  show.include=TRUE,
  cex=c(1.3,1.3,1.3),
  lwd=1.5)

fileName<-"Venn_iii_iv.tif"
path<-"C:/smedFiles/"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 22,
          width = 22,
          units = 'cm',
          type="windows",
          res=600)


#Annotation
#Create a fasta file containing the sequences to be annotated
write.fasta(sequences=as.list(smedFull$sequence), names=smedFull$ID,
file.out="smedSequences.fasta")

#Get the fasta files of the databases of interest from the NCBI repertoire
#Create blast database - Cell Proliferation
makeblastdb -in CellProliferationEumetazoa.fasta -dbtype nucl -out
CellProliferationEumetazoa.fa
#Contrast CellProfliferationEumetazoa database with smedSequences
blastn -query smedSequences.fasta -db CellProliferationEumetazoa.fa -max_target_seqs 1 -
max_hsps 1 -outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5 -
num_threads 20 > CellProliferationSMED.csv

#Do the same for the remaining databases
#DNA Replication
makeblastdb -in DNAReplicationEumetazoa.fasta -dbtype nucl -out DNAReplicationEumetazoa.fa
blastn -query smedSequences.fasta -db DNAReplicationEumetazoa.fa -max_target_seqs 1 -max_hsps
1 -outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5 -
num_threads 20 > DNAReplicationSMED.csv

#Cell Differentiation
makeblastdb -in CellDifferentiationEumetazoa.fasta -dbtype nucl -out
CellDifferentiationEumetazoa.fa
```

```
blastn -query smedSequences.fasta -db CellDifferentiationEumetazoa.fa -max_target_seqs 1 -
max_hsps 1 -outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5 -
num_threads 20 > CellDifferentiationSMED.csv

#DNA Repair
makeblastdb -in DNARepairEumetazoa.fasta -dbtype nucl -out DNARepairEumetazoa.fa
blastn -query smedSequences.fasta -db DNARepairEumetazoa.fa -max_target_seqs 1 -max_hsps 1 -
outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5 -num_threads
20 > DNARepairSMED.csv

#Cell Cycle Arrest
makeblastdb -in CellCycleArrestEumetazoa.fasta -dbtype nucl -out CellCycleArrestEumetazoa.fa
blastn -query smedSequences.fasta -db CellCycleArrestEumetazoa.fa -max_target_seqs 1 -max_hsps
1 -outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5 -
num_threads 20 > CellCycleArrestSMED.csv

#Cell Death
makeblastdb -in CellDeathEumetazoa.fasta -dbtype nucl -out CellDeathEumetazoa.fa
blastn -query smedSequences.fasta -db CellDeathEumetazoa.fa -max_target_seqs 1 -max_hsps 1 -
outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5 -num_threads
20 > CellDeathSMED.csv

#Tumor Supressor
makeblastdb -in TumorSupressorEumetazoa.fasta -dbtype nucl -out TumorSupressorEumetazoa.fa
blastn -query smedSequences.fasta -db TumorSupressorEumetazoa.fa -max_target_seqs 1 -max_hsps
1 -outfmt "6 delim=; qseqid sseqid pident nident length evalue stitle" -evalue 1e-5 -
num threads 20 > TumorSupressorSMED.csv


#Contrast the genes in study with subsets obtained from databases containing genes involved in
the regeneration process
#Upload the cell proliferation gene list
fileName<-"CellProliferationSMED.csv"
annotationData<-read.table(fileName,sep=";",header=FALSE)
colnames(annotationData)<-c("ID","sseqid","pident","nident","length","evalue","stitle")
head(annotationData)
colnames(annotationData)

#Contrast the cell proliferation gene list with the annotated S. mediterranea transcriptome of
36 035 genes. The match is made according to the ID
smedCellProliferationGenes<-merge(smedDEG, annotationData, by="ID")
View(smedCellProliferationGenes)
write.table(smedCellProliferationGenes,"smedCellProliferationGenes.csv",sep=";",col.names=NA)

#Contrast the cell proliferation gene list with the one containing the up and down regulated
genes of each comparison. The match is made according to the ID
#First comparison
smedCellProliferationDEG<-merge(degTable, annotationData, by="ID")
View(smedCellProliferationDEG)
write.table(smedCellProliferationDEG,"smedCellProliferationDEG.csv",sep=";",col.names=NA)

#Second comparison
smedCellProliferationDEG2<-merge(degTable2, annotationData, by="ID")
View(smedCellProliferationDEG2)
write.table(smedCellProliferationDEG2,"smedCellProliferationDEG2.csv",sep=";",col.names=NA)


#Upload the DNA replication gene list
fileName<-"DNAReplicationSMED.csv"
annotationData<-read.table(fileName,sep=";",quote="",header=FALSE)
colnames(annotationData)<-c("ID","sseqid","pident","nident","length","evalue","stitle")
head(annotationData)
colnames(annotationData)

#Contrast the DNA replication gene list with the annotated S. mediterranea transcriptome of 36
035 genes. The match is made according to the ID
smedDNAReplicationGenes<-merge(smedDEG, annotationData, by="ID")
View(smedCellCycleArrestGenes)
write.table(smedDNAReplicationGenes,"smedDNAReplicationGenes.csv",sep=";",col.names=NA)


#Contrast the DNA replication gene list with the one containing the up and down regulated
genes of each comparison. The match is made according to the ID
#First comparison
smedDNAreplicationDEG<-merge(degTable, annotationData, by="ID")
```

```
View(smedDNAreplicationDEG)
#There were no DEGs related with this process in this comparison

#Second comparison
smedDNAreplicationDEG2<-merge(degTable2, annotationData, by="ID")
View(smedDNAreplicationDEG2)
write.table(smedDNAreplicationDEG2,"smedDNAreplicationDEG2.csv",sep=";",col.names=NA)


#Upload the cell differentiation gene list
fileName<-"CellDifferentiationSMED.csv"
annotationData<-read.table(fileName,sep=";",quote="",header=FALSE)
colnames(annotationData)<-c("ID","sseqid","pident","nident","length","evalue","stitle")
head(annotationData)
colnames(annotationData)

#Contrast the cell differentiation gene list with the annotated S. mediterranea transcriptome
of 36 035 genes. The match is made according to the ID
smedCellDifferentiationGenes<-merge(smedDEG, annotationData, by="ID")
View(smedCellDifferentiationGenes)
write.table(smedCellDifferentiationGenes,"smedCellDifferentiationGenes.csv",sep=";",col.names=
NA)

#Contrast the cell differentiation gene list with the one containing the up and down regulated
genes of each comparison. The match is made according to the ID
#First comparison
smedCellDifferentiationDEG<-merge(degTable, annotationData, by="ID")
View(smedCellDifferentiationDEG)
write.table(smedCellDifferentiationDEG,"smedCellDifferentiationDEG.csv",sep=";",col.names=NA)
#The gene obtained was TPH which was already on the list of genes involved in cell
proliferation

#Second comparison
smedCellDifferentiationDEG2<-merge(degTable2, annotationData, by="ID")
View(smedCellDifferentiationDEG2)
write.table(smedCellDeathDEG2,"smedCellDeathDEG2.csv",sep=";",col.names=NA)


#Upload the DNA repair gene list
fileName<-"DNARepairSMED.csv"
annotationData<-read.table(fileName,sep="@",quote="",header=FALSE)
colnames(annotationData)<-c("ID", "sseqid", "pident", "nident", "length", "evalue", "stitle")
head(annotationData)
colnames(annotationData)

#Contrast the DNA repair gene list with the annotated S. mediterranea transcriptome of 36 035
genes. The match is made according to the ID
smedDNArepairGenes<-merge(smedDEG, annotationData, by="ID")
View(smedDNArepairGenes)
write.table(smedDNArepairGenes,"smedDNArepairGenes.csv",sep=";",col.names=NA)

#Contrast the DNA repair gene list with the one containing the up and down regulated genes of
each comparison. The match is made according to the ID
#First comparison
smedDNArepairDEG<-merge(degTable, annotationData, by="ID")
View(smedDNArepairDEG)
write.table(smedDNArepairDEG,"smedDNArepairDEG.csv",sep=";",col.names=NA)

#Second comparison
smedDNArepairDEG2<-merge(degTable2, annotationData, by="ID")
View(smedDNArepairDEG2)
write.table(smedDNArepairDEG2,"smedDNArepairDEG2.csv",sep=";",col.names=NA)


#Upload the cell cycle arrest gene list
fileName<-"CellCycleArrestSMED.csv"
annotationData<-read.table(fileName,sep=";",header=FALSE)
colnames(annotationData)<-c("ID","sseqid","pident","nident","length","evalue","stitle")
head(annotationData)
colnames(annotationData)

#Contrast the cell cycle arrest gene list with the annotated S. mediterranea transcriptome of
36 035 genes. The match is made according to the ID
smedCellCycleArrestGenes<-merge(smedDEG, annotationData, by="ID")
View(smedCellCycleArrestGenes)
```

```
write.table(smedCellCycleArrestGenes,"smedCellCycleArrestGenes.csv",sep=";",col.names=NA)
#Contrast the cell cycle arrest gene list with the one containing the up and down regulated
genes of each comparison. The match is made according to the ID
#First comparison
smedCellCycleArrestDEG<-merge(degTable, annotationData, by="ID")
View(smedCellCycleArrestDEG)
#There were no DEGs related with this process in this comparison

#Second comparison
smedCellCycleArrestDEG2<-merge(degTable2, annotationData, by="ID")
View(smedCellCycleArrestDEG2)
write.table(smedCellCycleArrestDEG2,"smedCellCycleArrestDEG2.csv",sep=";",col.names=NA)


#Upload the cell death gene list
fileName<-"CellDeathSMED.csv"
annotationData<-read.table(fileName,sep=";",quote="",header=FALSE)
colnames(annotationData)<-c("ID","sseqid","pident","nident","length","evalue","stitle")
head(annotationData)
colnames(annotationData)

#Contrast the cell death gene list with the annotated S. mediterranea transcriptome of 36 035
genes. The match is made according to the ID
smedCellDeathGenes<-merge(smedDEG, annotationData, by="ID")
View(smedCellDeathGenes)
write.table(smedCellDeathGenes,"smedCellDeathGenes.csv",sep=";",col.names=NA)

#Contrast the cell death gene list with the one containing the up and down regulated genes of
each comparison. The match is made according to the ID
#First comparison
smedCellDeathDEG<-merge(degTable, annotationData, by="ID")
View(smedCellDeathDEG)
#There were no DEGs related with this process in this comparison

#Second comparison
smedCellDeathDEG2<-merge(degTable2, annotationData, by="ID")
View(smedCellDeathDEG2)
write.table(smedCellDeathDEG2,"smedCellDeathDEG2.csv",sep=";",col.names=NA)


#Assemble the list of DEGs associated with regeneration of each comparison in a single table
#First comparison
GenesRegeneracao<-rbind(smedDNArepairDEG, smedCellProliferationDEG)
write.table(GenesRegeneracao,"GenesRegeneracao.csv",sep=";",col.names=NA)
fileName<-"NomesGenesRegeneracao.txt"
File<-paste0(path="C:/Users/Márcia/Documents/FCT - Genética Molecular e
Biomedicina/Tese/Bioinformática/GSE121045/",fileName)
NomesGenesRegeneracao<-read.table(File,sep=";",header=FALSE)
rownames(GenesRegeneracao)=NomesGenesRegeneracao[,1]
write.table(GenesRegeneracao,"GenesRegeneracao.csv",sep=";",col.names=NA)

#Second comparison
GenesRegeneracao2<-rbind(smedDNArepairDEG2, smedCellProliferationDEG2, smedDNAreplicationDEG2,
smedCellDifferentiationDEG2, smedCellDeathDEG2, smedCellCycleArrestDEG2)
write.table(GenesRegeneracao2,"GenesRegeneracao2.csv",sep=";",col.names=NA)
#Delete lines with repeated genes (PCNA, POLG e TPH)
GenesRegeneracao2<-GenesRegeneracao2[-c(21,25,34),]
write.table(GenesRegeneracao2,"GenesRegeneracao2.csv",sep=";",col.names=NA)
fileName<-"NomesGenesRegeneracao2.txt"
File<-paste0(path="C:/Users/Márcia/Documents/FCT - Genética Molecular e
Biomedicina/Tese/Bioinformática/GSE121045/",fileName)
NomesGenesRegeneracao2<-read.table(File,sep=";",header=FALSE)
rownames(GenesRegeneracao2)=NomesGenesRegeneracao2[,1]
write.table(GenesRegeneracao2,"GenesRegeneracao2.csv",sep=";",col.names=NA)


#Assemble the list of all genes associated with regeneration found in this study
GenesRegeneracao3<-rbind(smedDNArepairGenes,
smedCellProliferationGenes,smedCellDeathGenes,smedCellCycleArrestGenes,
smedDNAReplicationGenes, smedCellDifferentiationGenes)
write.table(GenesRegeneracao3,"GenesRegeneracao3.csv",sep=";",col.names=NA)
fileName<-"NomesGenesRegeneracao3.txt"
File<-paste0(path="C:/Users/Márcia/Documents/FCT - Genética Molecular e
Biomedicina/Tese/Bioinformática/GSE121045/",fileName)
NomesGenesRegeneracao3<-read.table(File,sep=";",header=FALSE)
```

```
rownames(GenesRegeneracao3)=NomesGenesRegeneracao3[,1]
write.table(GenesRegeneracao3,"GenesRegeneracao3.csv",sep=";",col.names=NA)


#Heatmaps
#First comparison
heatData<-as.matrix(GenesRegeneracao[,c(2,6)])
rownames(heatData)=rownames(GenesRegeneracao)
write.table(heatData,"heatData.csv",sep=";",col.names=NA)

heatColour<-colorRampPalette(c("#1E90FF", "#D1D0CE", "#CD0000"))(n = 500) #blue, gray, red
windows()
heatmap.2(
  heatData,
  density.info="none",
  key=TRUE,
  col=heatColour,
  labCol=c("Pole piece from\nintact planarians\n(comparison i)","Pole piece from\nregenerating
planarians\n(comparison ii)"),
  trace="none",
  scale="none",
  Colv=FALSE,
  Rowv=FALSE,
  dendogram="none",
  margins = c(10,20),
  srtCol=360,
  adjCol = c(0.5,1),
  cexRow=1.2,
  cexCol=1.2,
  lwid = c(1.5,4.5),
  lhei = c(1.5,4))

fileName<-"Heatmap_i_ii.tif"
path<-"C:/smedFiles/"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 13,
          width = 24,
          units = 'cm',
          type="windows",
          res=600)


#Second comparison
heatData2<-as.matrix(GenesRegeneracao2[,c(2,6)])
rownames(heatData2)=rownames(GenesRegeneracao2)
write.table(heatData2,"heatData2.csv",sep=";",col.names=NA)

#Clustering of the genes
clustFunction <- function(x) hclust(x, method="complete")
distFunction <- function(x) dist(x,method="euclidean")
rowCutoff = 5
cRow<-colorRampPalette(c("#99ccff", "#ffb3b3", "#cc0000")) #blue, pink, red
rowFit<-clustFunction(distFunction(heatData2))
rClusters<-cutree(rowFit,h=rowCutoff)
rHeight<-length(unique(as.vector(rClusters)));
rowHeight = cRow(rHeight)

heatColour<-colorRampPalette(c("#1E90FF", "#D1D0CE", "#CD0000"))(n = 500) #blue gray, red
windows()
heatmap.2(
  heatData2,
  hclust=clustFunction,
  distfun=distFunction,
  RowSideColors=rowHeight[rClusters],
  density.info="none",
  key=TRUE,
  col=heatColour,
  labCol=c("Flanking piece\n(comparison iii)", "Pole piece\n(comparison iv)"),
  trace="none",
  scale="none",
  Colv=FALSE,
  dendogram="row",
  margins = c(5,30),
  srtCol=360,
```

```
  adjCol = c(0.5,1),
  cexRow=0.9,
  cexCol=1.3,
  lwid = c(1.5,4),
  lhei = c(1.3,4))

fileName<-"Heatmap_iii_iv.tif"
path<-"C:/smedFiles/"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 15,
          width = 24,
          units = 'cm',
          type="windows",
          res=600)


#Venn diagrams from the differentially-expressed genes involved in regeneration
#First comparison
vennCounts(GenesRegeneracao)
windows()
vennDiagram(
  GenesRegeneracao[,c("Ai", "Ar")],
  names=c("DEGs related\nwith regeneration\nfrom comparison i\n","DEGs related\nwith
regeneration\nfrom comparison ii\n"),
  include=c("up", "down"),
  counts.col=c("red3","dodgerblue3"),
  circle.col=NULL,
  show.include=TRUE,
  cex=c(1.3,1.3,1.3),
  lwd=1.5)

fileName<-"Venn_i_ii_2.tif"
path<-"C:/smedFiles/"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 22,
          width = 22,
          units = 'cm',
          type="windows",
          res=600)


#Second comparison
vennCounts(GenesRegeneracao2)
windows()
vennDiagram(
  GenesRegeneracao2[,c("nA", "A")],
  names=c("DEGs related\nwith regeneration\nfrom comparison iii\n", "DEGs related\nwith
regeneration\nfrom comparison iv\n"),
  include=c("up", "down"),
  counts.col=c("red3","dodgerblue3"),
  circle.col=NULL,
  show.include=TRUE,
  cex=c(1.3,1.3,1.3),
  lwd=1.5)

fileName<-"Venn_iii_iv_2.tif"
path<-"C:/smedFiles/"
dev.print(tiff,
          paste(sep="",path,fileName),
          height = 22,
          width = 22,
          units = 'cm',
          type="windows",
          res=600)
```

Annex 7.12. List of 258 genes associated with regeneration grouped by biological process. Species and accession number information retrieved from NCBI database. Biological process information obtained in UniProt databases. DEGs classification based in the bioinformatic analysis.

Process	Gene	Species	Biological process	DEGs	Accession number
Cell proliferation	*YKI* - Transcriptional coactivator yorkie	*Schmidtea mediterranea*	Transcription, Transcription regulation	No	KF990477.1
	Ppp2r2d - Serine/threonine-protein phosphatase 2, regulatory subunit B, delta, transcript variant 2	*Mus musculus*	Cell cycle, Cell division, Mitosis	No	NM_001347618.1
	FOXK1 - Forkhead box K1	*Homo sapiens*	Differentiation, Host-virus interaction, Myogenesis, Transcription, Transcription regulation	Yes - up-regulated comparisons iii and iv	NM_001037165.2
	NF-YB - Nuclear factor-Y subunit B	*Schmidtea mediterranea*	Transcription, Transcription regulation	Yes - up-regulated comparisons iii and iv	KU366699.1
	Tubb3 - Tubulin, beta 3 class III	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle, Neuron differentiation	No	NM_023279.3
	MER	*Schmidtea mediterranea*	Cell-cell signaling, Cell migration, Multicellular organism development, Nervous system development	No	KF990480.1
	SD1	*Schmidtea mediterranea*	Transcription, Transcription regulation	Yes - up-regulated comparison iv	KF990481.1
	Acta1 - Actin, alpha 1, skeletal muscle, transcript variant 2	*Mus musculus*	Muscle protein (Structural constituent of cytoskeleton)	No	NM_009606.3
	WTS - Serine/threonine-protein kinase Warts	*Schmidtea mediterranea*	Kinase, Serine/threonine-protein kinase, Transferase	No	KF990478.1
	HSP90AA1 - Heat shock protein 90 alpha family class A member 1, transcript variant 1	*Homo sapiens*	Chaperone, ATPase activity, Cellular response to heat	No	NM_001017963.3
	EGFR-1 - Epidermal growth factor receptor 1	*Schmidtea mediterranea*	Developmental protein, Cell differentiation, Kinase, Transferase, Tyrosine-protein kinase	No	HM777018.1
	MATS2	*Schmidtea mediterranea*	Hippo signaling, Regulation of protein autophosphorylation	No	KF990484.1
	TH - Tyrosine hydroxylase	*Schmidtea mediterranea*	Tyrosine 3-monooxygenase activity	No	HM777014.1
	TPH - Tryptophan hydroxylase	*Schmidtea mediterranea*	Serotonin biosynthesis	Yes - down-regulated comparison i; up-regulated comparisons iii and iv	HM777015.1
	MER	*Schmidtea mediterranea*	Cell-cell signaling, Cell migration, Multicellular organism development, Nervous system development	No	KF990480.1
	NF-YA2 - Nuclear factor-Y subunit A2	*Schmidtea mediterranea*	Transcription, Transcription regulation	No	KU366702.1
	Smarca2 - SWI/SNF related, matrix associated, actin dependent regulator of	*Mus musculus*	Transcription, Transcription regulation, Chromatin regulator	No	NM_011416.2

chromatin, subfamily a, member 2, transcript variant 1				
SETD6 - SET domain containing 6, protein lysine methyltransferase, transcript variant 1	*Homo sapiens*	Stem cell differentiation, Stem cell population maintenance, Regulation of inflammatory response	No	NM_001160305.4
DP - E2F dimerization protein	*Schmidtea mediterranea*	Transcription, Transcription regulation	No	JX967268.1
NF-YB2 - Nuclear factor-Y subunit B2	*Schmidtea mediterranea*	Transcription, Transcription regulation	Yes - up-regulated comparisons iii and iv	KU366700.1
EGFR-3 - Epidermal growth factor receptor 3	*Schmidtea mediterranea*	Developmental protein, Cell differentiation, Kinase, Transferase, Tyrosine-protein kinase	No	HM777016.1
NF-YA1 - Nuclear factor-Y subunit A1	*Schmidtea mediterranea*	Transcription, Transcription regulation	No	KU366701.1
RB - Retinoblastoma-like protein	*Schmidtea mediterranea*	Cell cycle, Transcription, Transcription regulation, Chromatin regulator, DNA-binding	No	JX967264.1
Tubb3 - Tubulin, beta 3 class III	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle, Neuron differentiation	No	NM_023279.3
SD2	*Schmidtea mediterranea*	Transcription, Transcription regulation	No	KF990482.1
SAV	*Schmidtea mediterranea*	Apoptosis	No	KF990485.1
DP - E2F dimerization protein	*Schmidtea mediterranea*	Transcription, Transcription regulation	No	JX967268.1
SmB - Small ribonucleoprotein	*Schmidtea mediterranea*	mRNA processing, mRNA splicing	No	GU562964.1
NF-YC - Nuclear factor-Y subunit C	*Schmidtea mediterranea*	Transcription, Transcription regulation	No	KU366703.1
TH - Tyrosine hydroxylase	*Schmidtea mediterranea*	Tyrosine 3-monooxygenase activity	No	HM777014.1
E2F4-1 - E2F4/5-like protein	*Schmidtea mediterranea*	Cell cycle, Cilium biogenesis/degradation, Transcription, Transcription regulation	No	JX967265.1
SAV	*Schmidtea mediterranea*	Apoptosis	No	KF990485.1
Tubb4b - Tubulin, beta 4B class Ivb	*Danio rerio*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	NM_198809.2
SMED-P53	*Schmidtea mediterranea*	Apoptosis, Biological rhythms, Cell cycle, Host-virus interaction, Necrosis, Transcription, Transcription regulation	No	AY068713.3
SD1	*Schmidtea mediterranea*	Transcription, Transcription regulation	Yes - up-regulated comparison iv	KF990481.1
LOC257642 - rRNA promoter binding protein	*Ratus norvegicus*		No	U77931.1
Tubb3 - Tubulin, beta 3 class III	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle, Neuron differentiation	No	NM_023279.3
CycD - Cyclin D-like protein	*Schmidtea mediterranea*	Mitotic cell cycle phase transition, Positive regulation of cell growth	Yes - up-regulated comparisons iii and iv	JX967267.1

	EGFR-2 - Epidermal growth factor receptor 2	*Schmidtea mediterranea*	Developmental protein, Cell differentiation, Kinase, Transferase, Tyrosine-protein kinase	No	HM777017.1
	MATS1	*Schmidtea mediterranea*	Hippo signaling, Regulation of protein autophosphorylation	No	KF990483.1
	ACTB - Beta-actin	*Haplochromis burtoni*	Cell motility and contraction	Yes - up-regulated comparisons iii and iv	JF826504.1
	HPO - Serine/threonine-protein kinase hippo	*Schmidtea mediterranea*	Apoptosis, Negative regulation of cell population proliferation, Hippo signaling	No	KF990479.1
	CBX3 - Chromobox 3, transcript variant 2	*Homo sapiens*	Biological rhythms, Transcription, Transcription regulation, Chromatin regulator	No	NM_016587.4
	Erf1 - ETS-related transcription factor ERF	*Mus musculus*	Transcription, Transcription regulation	No	AY036118.1
	EEF1A1 - Eukaryotic translation elongation factor 1 alpha 1	*Homo sapiens*	Protein biosynthesis, Transcription, Transcription regulation	No	NM_001402.6
	NF-YB2 - Nuclear factor-Y subunit B2	*Schmidtea mediterranea*	Transcription, Transcription regulation	Yes - up-regulated comparisons iii and iv	KU366700.1
	Tubb3 - Tubulin, beta 3 class III	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle, Neuron differentiation	No	NM_023279.3
	HDAC1 - Histone deacetylase 1-like protein	*Schmidtea mediterranea*	Biological rhythms, Transcription, Transcription regulation, Chromatin regulator, Hydrolase	No	JX967266.1
	Tubb6 - Tubulin, beta 6 class V	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	NM_026473.2
DNA Replication	*MCM4* - DNA replication licensing factor MCM4	*Meloidogyne arenaria*	Cell cycle, DNA replication	Yes - up-regulated comparisons iii and iv	CF358483.1
	PPHLN1 - Periphilin 1, transcript variant 17	*Homo sapiens*	Keratinization, Transcription, Transcription regulation	Yes - up-regulated comparisons iii and iv	NM_001364830.2
	POLA1 - DNA polymerase alpha catalytic subunit	*Acyrthosiphon pisum*	DNA replication, Host-virus interaction	No	XM_003244406.4
	PCNA - Proliferating cell nuclear antigen	*Schmidtea mediterranea*	DNA damage, DNA repair, DNA replication, Host-virus interaction	Yes - up-regulated comparisons iii and iv	EU856391.1
	MCM7 - DNA replication licensing factor mcm7 homologue	*Schmidtea mediterranea*	Cell cycle, DNA replication	Yes - up-regulated comparisons iii and iv	EG352022.1
	MCM2 - DNA replication licensing factor mcm2-like, transcript variant X1	*Stegodyphus dumicola*	Cell cycle, DNA replication	Yes - up-regulated comparisons iii and iv	XM_035371495.1
	CCNK - Cyclin-K	*Pieris rapae*	Cell cycle, Cell division, Mitosis, Transcription, Transcription regulation	No	XM_022268935.1
	PRIM1 - DNA primase small subunit	*Ciona intestinalis*	DNA replication, Transcription	Yes - up-regulated comparison iv	XM_002129841.5
	POLE - DNA polymerase epsilon catalytic subunit A-like	*Dendronephthya gigantea*	DNA damage, DNA repair, DNA replication	No	XM_028544732.1
	MCM6 - DNA replication licensing factor mcm6-like	*Centruroides sculpturatus*	Cell cycle, DNA replication	Yes - up-regulated comparison iv	XM_023376582.1
	POLG - DNA polymerase gamma, catalytic subunit, transcript variant X1	*Sus scrofa*	DNA replication	Yes - up-regulated comparison iv	XM_001927064.5

	LOC102225078	*Xiphophorus maculatus*		No	XM_023339806.1
	Mcm6p	*Schmidtea mediterranea*	Cell cycle, DNA replication	Yes - up-regulated comparison iv	EE672843.1
	GINS4 - DNA replication complex GINS protein SLD5	*Frankliniella occidentalis*	DNA replication	No	XM_026421628.1
	RBMS2 - RNA binding motif single stranded interacting protein 2	*Homo sapiens*	RNA processing	No	NM_002898.4
	Pol polyprotein	*Schmidtea mediterranea*	DNA integration, DNA recombination, Nucleotide metabolism, Viral genome integration, Virus entry into host cell	No	GU591867.1
	P1-MCM3 - DNA replication licensing factor MCM3 (DNA polymerase alpha holoenzyme-associated protein P1)	*Schmidtea mediterranea*	Cell cycle, DNA replication	Yes - up-regulated comparisons iii and iv	EG343006.1
	Ino80 - INO80 complex ATPase subunit	*Rattus norvegicus*	Cell cycle, Cell division, DNA damage, DNA recombination, DNA repair, Mitosis	No	NM_001261404.1
	MCM7 - DNA replication licensing factor mcm7 homologue	*Schmidtea mediterranea*	Cell cycle, DNA replication	Yes - up-regulated comparisons iii and iv	EG352022.1
	MCM5 - DNA replication licensing factor MCM5	*Schmidtea mediterranea*	Cell cycle, DNA replication	Yes - up-regulated comparisons iii and iv	EE667998.1
	Mcm6p	*Schmidtea mediterranea*	Cell cycle, DNA replication	Yes - up-regulated comparison iv	EG350996.1
Cell Differentiation	*ntn5* - Netrin-5	*Schmidtea mediterranea*	Neurogenesis	No	MK430180.1
	603304885F1 NCI_CGAP_Mam4 Mus musculus	*Mus musculus*		No	BI661820.1
	Calcium-dependent cysteine protease	*Dugesia japonica*	Hydrolase, Protease	No	KY214379.1
	DjSTK5 - non receptor serine/threonine kinase	*Dugesia japonica*	Protein kinase activity, Transferase	No	AB014506.1
	Ppp2r2d - Protein phosphatase 2, regulatory subunit B, delta, transcript variant 1	*Mus musculus*	Cell cycle, Cell division, Mitosis	No	NM_026391.2
	GATA456-2	*Schmidtea mediterranea*	DNA- binding transcription factor	No	KX827244.1
	17000531877137 GRN_ES Homo sapiens	*Homo sapiens*		No	CN278574.1
	17000531927517 GRN_EB Homo sapiens	*Homo sapiens*		No	CN289152.1
	17000532198181 GRN_EB Homo sapiens	*Homo sapiens*		No	CN411900.1
	BRAF - B-Raf proto-oncogene, serine/threonine kinase, transcript variant 1	*Homo sapiens*	Cell proliferation, Cell differentiation, Cell migration, apoptosis	Yes - up-regulated comparisons iii and iv	NM_004333.6
	17000417981260 GRN_EB Homo sapiens	*Homo sapiens*		No	CN346130.1
	PIM-1-like protein	*Schmidtea mediterranea*	Apoptosis, Cell cycle	Yes - down-regulated comparisons iii and iv	KY849970.1
	Tubb4b.L - Tubulin beta 4B class IVb L homeolog	*Xenopus laevis*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	NM_001087097.2

	PIM-1-like protein	*Schmidtea mediterranea*	Apoptosis, Cell cycle	Yes - down-regulated comparisons iii and iv	KY849970.1
	17000600008005 GRN_PREHEP Homo sapiens	*Homo sapiens*		Yes - up-regulated comparisons iii and iv	CN272630.1
	Tubb4b - Beta 4B class IVB	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	NM_146116.2
	SUSFLECK Fat Cell Normalized	*Sus scrofa*		No	FD633048.1
	Adipocyte Zap Express Library After differentiation	*Sus scrofa*		No	EH005895.1
	17000424025775 GRN_EB Homo sapiens	*Homo sapiens*		No	CN257111.1
	not7a pseudogene	*Schmidtea mediterranea*		No	KF781124.1
	HSP90B1 - Heat shock protein 90 beta family member 1	*Canis lupus familiaris*	Chaperone, ATPase activity, Cellular response to heat	No	NM_001003327.2
	PIM-1-like protein	*Schmidtea mediterranea*	Apoptosis, Cell cycle	Yes - down-regulated comparisons iii and iv	KY849970.1
	17000599933949 GRN_PRENEU Homo sapiens	*Homo sapiens*		Yes - up-regulated comparisons iii and iv	CN392206.1
	NSD-1 - Nuclear Receptor Binding SET Domain Protein 1	*Schmidtea mediterranea*	Chromatin regulator, Methyltransferase, Transferase	Yes - up-regulated comparisons iii and iv	JQ425134.1
	EGFR-1 - Epidermal growth factor receptor 1	*Schmidtea mediterranea*	Developmental protein, Cell differentiation, Kinase, Transferase, Tyrosine-protein kinase	No	HM777018.1
	PIM-1	*Schmidtea mediterranea*	Apoptosis, Cell cycle	Yes - down-regulated comparisons iii and iv	KY849969.1
	not6	*Schmidtea mediterranea*	RNA-mediated gene silencing, Transcription, Transcription regulation, Translation regulation	No	KF781123.1
	TH - Tyrosine hydroxylase	*Schmidtea mediterranea*	Tyrosine 3-monooxygenase activity	No	HM777014.1
	TPH - Tryptophan hydroxylase	*Schmidtea mediterranea*	Serotonin biosynthesis	Yes - down-regulated comparison i; up-regulated comparisons iii and iv	HM777015.1
	602813252F1 NCI_CGAP_Mam4 Mus musculus	*Mus musculus*		No	BG914506.1
	ntn3 - Netrin-3	*Schmidtea mediterranea*	Neurogenesis	No	MK430179.1
	17000600109768 GRN_PRENEU Homo sapiens	*Homo sapiens*		No	CN413076.1
	ntn1 - Netrin 1	*Schmidtea mediterranea*	Neurogenesis	No	AY945304.1
	TUD-1 - Tudor and KH domain containing-1	*Schmidtea mediterranea*	RNA-binding	No	KF781126.1
	17000424334876 GRN_EB Homo sapiens	*Homo sapiens*		No	CN413061.1

NT2NE2	*Homo sapiens*		No	DA687121.1
ntn2 - Netrin 2 mRNA	*Schmidtea mediterranea*	Neurogenesis	Yes - up-regulated comparisons iii and iv	AY945305.1
602815144F1 NCI_CGAP_Mam4 Mus musculus	*Mus musculus*		No	BG916061.1
AGRN - Agrin	*Apis mellifera*	Differentiation	No	XM_026444279.1
17000424180560 GRN_EB Homo sapiens	*Homo sapiens*		No	CN263776.1
not1	*Schmidtea mediterranea*	Transcription, Transcription regulation	No	KF781122.1
17000532555263 GRN_ES Homo sapiens	*Homo sapiens*		No	CN403363.1
EGFR-3 - Epidermal growth factor receptor 3	*Schmidtea mediterranea*	Developmental protein, Cell differentiation, Kinase, Transferase, Tyrosine-protein kinase	No	HM777016.1
DjhistonH2B - Histone H2B	*Dugesia japonica*		No	AB576209.1
17000600023772 GRN_PRENEU Homo sapiens	*Homo sapiens*		No	CN412022.1
Tubb4b.L - Tubulin beta 4B class IVb L homeolog	*Xenopus laevis*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	NM_001087097.2
17000470520581 GRN_EB Homo sapiens	*Homo sapiens*		No	CN257897.1
17000600021690 GRN_PRENEU Homo sapiens	*Homo sapiens*		No	CN394927.1
RCD1 - Cell differentiation protein RCD1 homolog	*Drosophila ficusphila*		No	XM_017200287.1
DjhistonH2B - Histone H2B	*Dugesia japonica*	DNA-binding	No	AB576209.1
Tubb4b.L - Tubulin beta 4B class IVb L homeolog	*Xenopus laevis*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	NM_001087097.2
602813526F1 NCI_CGAP_Mam4 Mus musculus	*Mus musculus*		No	BG914741.1
SUSFLECK Fat Cell	*Sus scrofa*		No	FD623128.1
ntn2 - Netrin 2	*Schmidtea mediterranea*	Neurogenesis	Yes - up-regulated comparisons iii and iv	AY945305.1
TH - Tyrosine hydroxylase	*Schmidtea mediterranea*	Tyrosine 3-monooxygenase activity	No	HM777014.1
DjPTK4 - non receptor tyrosine kinase	*Dugesia japonica*		No	AB014501.1
ntn4 - netrin-4	*Schmidtea mediterranea*	Organ morphogenesis, Axon guidance, Basement membrane assembly, Cell migration, Neuron remodeling, Tissue development	No	MK430172.1
EEF1A1 - Elongation factor 1-alpha	*Mus musculus*	Protein biosynthesis, Transcription regulation	No	BE197615.1
CHD4 - Chromodomain helicase DNA-binding protein 4	*Schmidtea mediterranea*	Transcription, Transcription regulation	No	GU980571.1

	LamC - Lamin-C	*Schmidtea mediterranea*	Chromatin organization, Chromatin silencing, Muscle tissue morphogenesis	No	MK430178.1
	Tubb2a - Tubulin, beta 2A class IIA	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle, Neuron migration	No	NM_009450.2
	Pak3 - p21 (RAC1) activated kinase 3, transcript variant 1	*Mus musculus*	Positive regulation of DNA biosynthetic process, Positive regulation of fibroblast migration, Positive regulation of neuron apoptotic process, Regulation of actin filament polymerization	No	NM_001195046.1
	not7b	*Schmidtea mediterranea*	RNA-mediated gene silencing, Transcription, Transcription regulation, Translation regulation	No	KF781125.1
	NCI_CGAP_Mam4 Mus musculus	*Mus musculus*		No	BQ109050.1
	17000600186324 GRN_PREHEP Homo sapiens	*Homo sapiens*		No	CN279547.1
	Tubb4b.L - Tubulin beta 4B class IVb L homeolog	*Xenopus laevis*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	NM_001087097.2
	Ino80 - INO80 complex subunit	*Mus musculus*	Cell cycle, Cell division, DNA damage, DNA recombination, DNA repair, Mitosis	No	NM_026574.4
	CAPN7 - Calpain-7-like	*Aethina tumida*	Positive regulation of epithelial cell migration	No	XM_020024827.1
	TSA: Musca domestica ALHF_02915.g1743	*Musca domestica*		No	KA645859.1
	EGFR-2 - Epidermal growth factor receptor 2	*Schmidtea mediterranea*	Developmental protein, Cell differentiation, Kinase, Transferase, Tyrosine-protein kinase	No	HM777017.1
	17000470662483 GRN_EB Homo sapiens	*Homo sapiens*		No	CN397903.1
	603305732F1 NCI_CGAP_Mam4 Mus musculus	*Mus musculus*		No	BI661606.1
	17000417760941 GRN_EB Homo sapiens	*Homo sapiens*		No	CN362506.1
	AG4 Human epidermis granular keratinocytes	*Homo sapiens*		No	EL593480.1
	prx1 - Prx1 protein	*Ciona intestinalis*	Cell redox homeostasis, Cellular response to oxidative stress, Regulation of hydrogen peroxide-induced cell death	No	NM_001032511.1
	17000424182159 GRN_EB Homo sapiens	*Homo sapiens*		No	CN339447.1
	603283028F1 NCI_CGAP_Mam4 Mus musculus	*Mus musculus*		No	BI656241.1
	Tubb4b.L - Tubulin beta 4B class IVb L homeolog	*Xenopus laevis*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	NM_001087097.2
	APACD - ATP binding protein associated with cell differentiation	*Schmidtea mediterranea*	Cadherin binding	No	EG417935.1
	DjmkpA - Mitogen-activated protein kinase phosphatase	*Dugesia japonica*	Hydrolase, Kinase, Protein phosphatase	Yes - up-regulated comparison iv	AB576208.1
	Hsp90aa1 - Heat shock protein 90, alpha (cytosolic), class A member 1	*Mus musculus*	Stress response	No	NM_010480.5

	HDAC2 - Histone deacetylase 2	*Gallus gallus*	Biological rhythms, Transcription, Transcription regulation	No	NM_204831.2
	Cpd - Carboxypeptidase D	*Mus musculus*	Protein processing, Carboxypeptidase, Hydrolase, Metalloprotease, Protease	No	NM_007754.2
	Tubb4b.L - Tubulin beta 4B class IVb L homeolog	*Xenopus laevis*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	NM_001087097.2
DNA Repair	*DNAH7* - Dynein heavy chain 7, axonemal-like, transcript variant X1	*Pomacea canaliculata*	Motor protein, Cilium-dependent cell motility	Yes - down-regulated comparisons iii and iv	XM_025252352.1
	KIF16B - Kinesin family member 16B, transcript variant X2	*Haliaeetus leucocephalus*	Transport, Early endosome to late endosome transport, Epidermal growth factor receptor signaling pathway, Fibroblast growth factor receptor signaling pathway	No	XM_010574791.1
	KIF27 - Kinesin family member 27, transcript variant X1	*Pteropus vampyrus*	Motor protein, Cilium biogenesis/degradation, Microtubule-based movement	No	XM_011372048.2
	MCM8 - DNA helicase MCM8-like, transcript variant X1	*Dendronephthya gigantea*	Cell cycle, DNA damage, DNA repair, DNA replication	Yes - up-regulated comparisons iii and iv	XM_028553846.1
	PPFIA1 - PTPRF interacting protein alpha 1, transcript variant X4	*Macaca mulatta*	Cell-matrix adhesion, Signal transduction	No	XM_028833090.1
	BLM - Bloom syndrome protein homolog, transcript variant X1	*Melanaphis sacchari*	DNA damage, DNA repair, DNA replication	Yes - up-regulated comparisons iii and iv	XM_025344998.1
	Cre-ceh-43	*Caenorhabditis remanei*	Regulation of transcription by RNA polymerase II	Yes - up-regulated comparison iv	XM_003106012.1
	TONSL - Tonsoku like, DNA repair protein, transcript variant X1	*Homo sapiens*	DNA damage, DNA repair	Yes - up-regulated comparisons iii and iv	XM_011517048.2
	PIT - probable ATP-dependent RNA helicase pitchoune	*Melanaphis sacchari*	Helicase, Hydrolase, RNA-binding	No	XM_025345418.1
	Tubb5 - Tubulin beta-5 Chain	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	BM877412.1
	DDX17 - Probable ATP-dependent RNA helicase DDX17, transcript variant X1	*Polistes canadensis*	Antiviral defense, Immunity, mRNA processing, mRNA splicing, RNA-mediated gene silencing, rRNA processing, Transcription, Transcription regulation	No	XM_014749697.1
	PCNA - Proliferating cell nuclear antigen	*Schmidtea mediterranea*	DNA damage, DNA repair, DNA replication, Host-virus interaction	Yes - up-regulated comparisons iii and iv	EU856391.1
	me31B - Putative ATP-dependent RNA helicase me31b	*Exaiptasia pallida*	Translation regulation	Yes - up-regulated comparisons iii and iv	XM_021043585.1
	POTEG - POTE ankyrin domain family member G, transcript variant X1	*Macaca mulatta*	Retina homeostasis, substantia nigra development	No	XM_028826296.1
	HELLS - Lymphoid-specific helicase-like	*Parasteatoda tepidariorum*	Cell cycle, Cell division, Mitosis, Transcription, Transcription regulation	No	XM_016057806.1
	CXPD - DNA-repair protein complementing XP-D cells (Xeroderma pigmentosum group D complementing protein)	*Schmidtea mediterranea*	DNA damage, DNA repair	No	EG350329.1
	DDX17 - Probable ATP-dependent RNA helicase DDX17, transcript variant X1	*Polistes canadensis*	Antiviral defense, Immunity, mRNA processing, mRNA splicing, RNA-mediated gene	No	XM_014749697.1

		silencing, rRNA processing, Transcription, Transcription regulation		
MRE11 - Double-strand break repair protein MRE11	*Schmidtea mediterranea*	DNA damage, DNA repair, Host-virus interaction, Meiosis	No	FJ588613.1
HSP90AA1 - Heat shock protein HSP 90-alpha-like	*Callithrix jacchus*	Chaperone, ATPase activity, Cellular response to heat	No	XM_008983535.1
RTEL1 - Regulator of telomere elongation helicase 1	*Schmidtea mediterranea*	DNA damage, DNA repair	No	JQ425144.1
TOP3A - DNA topoisomerase 3-alpha	*Acyrthosiphon pisum*	Chromossome separation, DNA replication, DNA topological change, Meiotic cell cycle, Regulation of signal transduction by p53 class mediator	No	XM_001950492.5
DDX24 - ATP-dependent RNA helicase DDX24, transcript variant X1	*Solenopsis invicta*	Helicase, Hydrolase, RNA metabolic process	No	XM_011162031.2
CycK - Cyclin-K	*Pieris rapae*		No	XM_022268935.1
DDX42 - ATP-dependent RNA helicase DDX42	*Pieris rapae*	Helicase, Hydrolase, RNA- binding, mRNA splicing via spliceosome, Regulation of apoptotic process	No	XM_014629683.1
p68 - RNA helicase p68a isoform I (p68)	*Cynoglossus semilaevis*	RNA helicase activity, Hydrolase	No	KT270442.1
PTIP - PAX transcription activation domain interacting protein	*Schmidtea mediterranea*	Anterior/posterior pattern specification, Embryonic pattern specification, Histone methylation	No	KC262340.1
vas - ATP-dependent RNA helicase vasa, isoform A-like	*Cotesia chilonis*	Differentiation, Oogenesis	Yes - down-regulated comparisons iii and iv	MH366228.1
Gp93 - Glycoprotein 93	*Strongyloides ratti*	Chaperone, Cellular response to heat	No	XM_024644617.1
ALG8 - Probable dolichyl pyrophosphate Glc1Man9GlcNAc2 alpha-1,3-glucosyltransferase	*Strongyloides ratti*	Protein glycosylation, Glycosyltransferase, Transferase	No	XM_024651824.1
Ku70 - ATP-dependent DNA helicase ku70	*Schmidtea mediterranea*	DNA damage, DNA recombination, DNA repair	No	KT375434.1
Putative DNA damage repair protein	*Schmidtea mediterranea*	DNA repair	No	EG353773.1
REV1 - DNA repair protein REV1	*Fopius arisanus*	DNA damage, DNA repair, DNA synthesis	No	XM_011306758.1
SEM-1997 voucher Col2PYCNO RNA helicase	*Nymphon unguiculatum*		No	GQ885642.1
POLG - DNA polymerase gamma, catalytic subunit, transcript variant X1	*Sus scrofa*	DNA replication	Yes - up-regulated comparison iv	XM_001927064.5
PAXIP1 - PAX-interacting protein 1	*Drosophila obscura*	DNA damage, DNA recombination, DNA repair, Transcription, Transcription regulation	No	XM_022361236.1
Ercc2 - General transcription and DNA repair factor IIH helicase subunit XPD-like	*Schistosoma japonicum*	Chromosome partition, DNA damage, DNA repair, Transcription, Transcription regulation	No	BU800920.1
DDX11 - Probable ATP-dependent RNA helicase DDX11	*Polistes dominula*	DNA damage, DNA repair, DNA replication, Host-virus interaction, Transcription, Transcription regulation	No	XM_015324485.1
Mi-2 - Chromodomain-helicase-DNA-binding protein Mi-2 homolog, transcript variant X1	*Polistes dominula*	Transcription, Transcription regulation	No	XM_015326686.1

	WRN - WRN RecQ like helicase	*Homo sapiens*	Helicase, Hydrolase, DNA repair, Cellular response to gamma radiation and UV-C, DNA recombination, Regulation os apoptotic process, Regulation of growth rate	No	NM_000553.6
	DNMT3A - DNA (cytosine-5)-methyltransferase 3A-like, transcript variant X1	*Salmo salar*	DNA methylation, Mitotic cell cycle, Negative regulation of gene expression, epigenetic, Negative regulation of transcription by RNA polymerase II, Neuron differentiation, Positive regulation of cell death	No	XM_014197582.1
	HSP90-2 - Heat shock protein 90-2	*Dugesia japonica*	Stress response, Cellular response to heat, Protein stabilization, Immunity, Innate immunity,	No	MG334594.1
	SNF2 family domain-containing protein	*Loa loa*	Chromatind remodeling	No	XM_020446817.1
	BLM - Bloom syndrome protein homolog, transcript variant X1	*Melanaphis sacchari*	DNA damage, DNA repair, DNA replication	Yes - up-regulated comparisons iii and iv	XM_025344998.1
	ZNFX1 - NFX1-type zinc finger-containing protein 1-like, transcript variant X1	*Centruroides sculpturatus*	Heterochromatin assembly by small RNA	Yes - down-regulated comparison iv	XM_023369544.1
	SpolvlgA - DDX3/PL10 DEAD-box RNA helicase	*Schmidtea polychroa*	Apoptosis, Chromosome partition, Host-virus interaction, Immunity, Innate immunity, Ribosome biogenesis, Transcription, Transcription regulation, Translation regulation	No	FJ599746.1
	Tubb5 - Tubulin beta-5 chain	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	BM877412.1
	UPF-1 - Up-frameshift-1 nonsense mediated mRNA decay factor	*Dugesia japonica*	Helicase, Hydrolase, RNA helicase activity, Cell differentiation	No	GU305865.1
	cenpe - Centromere protein E, transcript variant X10	*Salmo salar*	Cell cycle, Cell division, Mitosis	No	XM_014160204.1
	Uncharacterized LOC105694552	*Orussus abietinus*		No	XM_023434944.1
	PAXIP1 - PAX-interacting protein 1, transcript variant X1	*Cephus cinctus*	DNA damage, DNA recombination, DNA repair, Transcription, Transcription regulation	No	XM_015737251.2
	DDX3X - DEAD-box helicase 3 X-linked, transcript variant X1	*Lagenorhynchus obliquidens*	Apoptosis, Chromosome partition, Host-virus interaction, Immunity, Innate immunity, Ribosome biogenesis, Transcription, Transcription regulation, Translation regulation	No	XM_027091447.1
	RAD18 - RAD18 E3 ubiquitin protein ligase	*Mustela putorius furo*	DNA damage, DNA repair, Ubl conjugation pathway	No	XM_004738431.2
	PES1 - Pescadillo-1	*Schmidtea mediterranea*	Ribosome biogenesis, rRNA processing	No	JX010566.1
	me31B - Putative ATP-dependent RNA helicase me31b	*Rhopalosiphum maidis*	Translation regulation	Yes - up-regulated comparisons iii and iv	XM_026956926.1
	ZNFX1 - NFX1-type zinc finger-containing protein 1-like	*Folsomia candida*	Heterochromatin assembly by small RNA	Yes - down-regulated comparison iv	XM_022097799.1
	vlg3 - vasa-like DEAD-box RNA helicase	*Schistosoma mansoni*	Helicase, Hydrolase	No	JQ619871.1

	CXPC - DNA-repair protein complementing XP-C cells (Xeroderma pigmentosum group C complementing protein)	*Schmidtea mediterranea*	DNA damage, DNA repair, Transcription, Transcription regulation	No	EE668413.1
	MSH6 - DNA mismatch repair protein MSH6	*Schmidtea mediterranea*	DNA damage, DNA repair, Host-virus interaction	No	JF519637.1
	Tubb5 - Tubulin beta-5 chain	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	BM877412.1
	Ino80 - INO80 complex ATPase subunit	*Grammomys surdaster*	Cell cycle, Cell division, DNA damage, DNA recombination, DNA repair, Mitosis	No	XM_028767073.1
	MLH1 - DNA mismatch repair protein MLH1	*Schmidtea mediterranea*	Cell cycle, DNA damage, DNA repair	No	JF511468.1
	PTIP - PAX transcription activation domain interacting protein	*Schmidtea mediterranea*	Anterior/posterior pattern specification, Embryonic pattern specification, Histone methylation	No	KC262340.1
	CXPD - DNA-repair protein complementing XP-D cells (Xeroderma pigmentosum group D complementing protein)	*Schmidtea mediterranea*	DNA damage, DNA repair	No	EG350329.1
	DDX3X - ATP-dependent RNA helicase DDX3X-like, transcript variant X1	*Vulpes vulpes*	Apoptosis, Chromosome partition, Host-virus interaction, Immunity, Innate immunity, Ribosome biogenesis, Transcription, Transcription regulation, Translation regulation	No	XM_025988677.1
	MSH2 - DNA mismatch repair protein MSH2	*Schmidtea mediterranea*	DNA damage, DNA repair	No	JF511467.1
	alkB - Probable DNA repair system specific for alkylated DNA protein	*Schmidtea mediterranea*	DNA damage, DNA repair	No	EE281462.1
	RCK (DDX6 - Probable ATP-dependent RNA helicase DDX6)	*Mus musculus*	Stem cell population maintenance, Negative regulation of neuron differentiation, Spermatid differentiation	No	D50494.1
	rad54l - DNA repair and recombination protein RAD54 like	*Acipenser ruthenus*	DNA damage, DNA repair	No	XM_034029630.1
	DNAH6 - Dynein axonemal heavy chain 6	*Chelonia mydas*	Motor protein, ATP-dependent microtubule motor activity	Yes - down-regulated comparisons iii and iv	XM_007052704.1
	WM6 - ATP-dependent RNA helicase WM6	*Anoplophora glabripennis*	mRNA processing, mRNA splicing, Chromatin remodeling	No	XM_018720653.1
	RAD51 - DNA repair protein RAD51	*Schmidtea mediterranea*	DNA damage, DNA recombination, DNA repair	Yes - down-regulated comparison i; up-regulated comparisons iii and iv	KM487300.1
	Tubb5 - Tubulin beta-5 chain	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	BM877412.1
	HSP90B1 - Endoplasmin homolog	*Stomoxys calcitrans*	Negative regulation of apoptotic process, Post-translational protein modification, Protein folding	No	XM_013260975.1
	XPCA - DNA-repair protein complementing XP-A cells (Xeroderma pigmentosum group A complementing protein homolog)	*Schmidtea mediterranea*	DNA damage, DNA repair	No	EE667522.1
	Tubb5 - Tubulin beta-5 chain	*Mus musculus*	Microtubule cytoskeleton organization, Mitotic cell cycle	No	BM877412.1

	TOP3A - DNA topoisomerase 3-alpha-like	*Acropora digitifera*	Chromossome separation, DNA replication, DNA topological change, Meiotic cell cycle, Regulation of signal transduction by p53 class mediator	No	XM_015912753.1
Cell Cycle Arrest	*CDC42* - Cell division cycle 42, transcript variant 1	*Homo sapiens*	Differentiation, Neurogenesis	Yes - up-regulated comparisons iii and iv	NM_001791.4
	Polr2a - RNA polymerase II subunit A	*Cricetulus griseus*	Host-virus interaction, Transcription	No	NM_001244002.2
	Smarca2 - SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2, transcript variant 1	*Mus musculus*	Transcription, Transcription regulation	No	NM_011416.2
	WRN - WRN RecQ like helicase	*Homo sapiens*	Helicase, Hydrolase, DNA repair, Cellular response to gamma radiation and UV-C, DNA recombination, Regulation os apoptotic process, Regulation of growth rate	No	NM_000553.6
	Polr2a - RNA polymerase II subunit A	*Cricetulus griseus*	Host-virus interaction, Transcription	No	NM_001244002.2
Cell Death	*PPID* - Peptidylprolyl isomerase D	*Homo sapiens*	Apoptosis, Host-virus interaction, Protein transport, Transport	No	NM_005038.3
	ZMAT3 - Zinc finger matrin-type 3, transcript variant 2	*Homo sapiens*	Apoptosis, DNA damage, Growth regulation, Protein transport, Translocation, Transport	Yes - up-regulated comparisons iii and iv	NM_152240.3
	PCD8 - Programmed cell death 8, isoform 1	*Schmidtea mediterranea*	Apoptosis	No	EE671023.1
	18S ribosomal RNA	*Mus musculus*	Ribosome biogenesis, rRNA processing	Yes - up-regulated comparisons iii and iv	AY248756.1
	ACTB - Beta-actin	*Felis catus*	Cell motility and contraction	Yes - up-regulated comparisons iii and iv	AB051104.1
	PDCD2L - Programmed cell death 2-like	*Lepisosteus oculatus*	Apoptosis	No	XM_015367975.1
	PIGK - Putative GPI-anchor transamidase	*Onthophagus taurus*	GPI-anchor biosynthesis	No	XM_023052156.1
	POLR1A - RNA polymerase I subunit A	*Homo sapiens*	Transcription	No	NM_015425.6
	PDCD6 - Programmed cell death protein 6, transcript variant X1	*Galleria mellonella*	Apoptosis	No	XM_026909531.2
	PDCD6 - Programmed cell death protein 6-like	*Xenopus laevis*	Apoptosis	No	XM_018241649.1
	JNK - Stress-activated protein kinase JNK	*Schmidtea mediterranea*	JNK cascade	No	KC879720.1
	PDCD6 - Programmed cell death 6 interacting protein	*Schmidtea mediterranea*	Apoptosis	No	EG405717.1
	ESG017b.E21_C09.3prime ESG01 Drosophila melanogaster	*Drosophila melanogaster*		No	CA807285.1
	ESG0112b.B21_H10.3prime ESG01 Drosophila melanogaster	*Drosophila melanogaster*		No	CA804185.1
	Caspase	*Drosophila willistoni*	Apoptosis	No	XM_015177512.2

	BCL2-1 - Bcl2-like protein	*Schmidtea mediterranea*	Apoptosis, Endocytosis	No	FJ807655.1
	18S ribosomal RNA	*Mus musculus*	Ribosome biogenesis, rRNA processing	Yes - up-regulated comparisons iii and iv	AY248756.1
	CASP8 - Caspase-8	*Schmidtea mediterranea*	Apoptosis, Host-virus interaction	No	KF364484.1
	PDCD5 - Programmed cell death protein 5-like	*Camelus ferus*	Apoptosis	Yes - down-regulated comparison iii	XM_032485437.1
	CASP8 - Caspase-8	*Schmidtea mediterranea*	Apoptosis, Host-virus interaction	No	KF364484.1
	JNK - Stress-activated protein kinase JNK	*Schmidtea mediterranea*	JNK cascade	No	KC879720.1
	ESG012c.E21_D01.3prime ESG01 Drosophila melanogaster	*Drosophila melanogaster*		No	CA806263.1
	PDCD2 - Programmed cell death 2	*Scyliorhinus canicula*	Apoptosis	No	FP885828.1