Analysis of gene expression in stem cells: FOXO transcription factors as master regulators of stemness

Martín Martínez, Irene

Curs 2014-2015

Director: Eusebio Perdiguerò

GRAU EN ENGINYERIA BIOMÈDICA

Treball de Fi de Grau
Analysis of gene expression in stem cells: FOXO transcription factors as master regulators of stemness

Irene Martín Martínez – Biomedical engineering
# Contents

Summary ........................................................................................................................................... 3

Introduction ...................................................................................................................................... 3

Study of the Literature .......................................................................................................................... 8

Methodology ........................................................................................................................................ 12

*In silico analysis* .................................................................................................................................. 12

  - Data collection................................................................................................................................. 12
  - Data analysis..................................................................................................................................... 14

*In vitro analyses* .................................................................................................................................. 18

  - Cell Culture..................................................................................................................................... 18
  - Generation of Knockout cells: The recombination system Cre/LoxP ................................................... 19
  - Quantitative PCR ............................................................................................................................ 19
  - In vitro quiescence assay through infection with BDM ................................................................. 21
  - In vitro differentiation assay ........................................................................................................... 22
  - Fluorescence microscopy images .................................................................................................... 22

Results .................................................................................................................................................. 22

  - Results of in silico analysis .............................................................................................................. 22
  - Results of in vitro models .................................................................................................................. 27

Discussion ........................................................................................................................................... 31

Conclusion ........................................................................................................................................... 33

Bibliography ......................................................................................................................................... 34

Appendix ............................................................................................................................................. 38
Summary

Nowadays, it exist the necessity to understand the intrinsic mechanism that governs stem cells which constitutes a regeneration system. The expression of genes, transcription factors, microRNAs and pathways acting as a regulators enable that stem cells work properly, this elements are the responsible of the particular behavior and as a consequence allow that stem cells present the capacity of self-renewal. It is useful to work and analyze the expression of these elements in order to understand specific mechanisms present in stem cells, for this purpose we implemented a comparative analysis of global gene expression and also it is realized through different programs a analysis of the elements acting behind these genes that are transcription factors, microRNAs and pathways. The results reveals the existence of a common element, which is FOXO3 transcription factor, present in all the types of stem cells and in different species including immortal Hydra, an organism that presents indefinite self-renewal, this fact allow to make the assumption that FOXO3 is a critical regulator that controls the characteristic properties of all the stem cells. We design in vitro quiescence and in vitro differentiation assays with muscular stem cells that confirm the assumption. Specifically, the deletion of FOXO3 in stem cells demonstrates that are unable to self-renew and to re-entry in quiescence, we observe that the cells present a differentiation fate that lead to complete deletion of stem cells. We demonstrate the important role of FOXO 3 transcription factor in maintenance of stem cells; also we define the function of FOXO3 that is the suppression of terminal differentiation allowing to replenish the stem cell pool. In conclusion, we discover a key element that is a master regulator on stemness and it is responsible for existence of stem cells. This particular role suggests that FOXO3 is implicated in longevity and can be a potential marker for regeneration and aging processes.

Introduction

New technologies present today provide to the scientists new tools to investigate in the technical and biological field. We can combine different types of information in order to focus the research in the world of biomedical engineering. Nowadays, we have a set of programs and databases that allow the engineer to propose new questions that lead to investigate aspects that previously could not be investigated. Specifically, the biomedical field presents an intensive activity in the management of information and processing of data. In addition, researchers have been developed programs that allow the calculation of mathematical and statistical parameters. The diffusion of results, creation and development of projects are characteristic features of the research activity. Therefore, I can use this advances to propose a study related with stem cells, the investigation is focused in a global gene analysis that provides new results. Probably, I have found a component which is the key for regeneration, maintenance and regulation of stem cells.
Stem cells are cells in the human body with the capacity to differentiate into other cell types. The potential (multipotency) and self-renewal are the properties that allow us to classify them as a regenerative system. First, it is known that stem cells divide and remain as a population for a period of time; this feature is defined as self-renewal. Secondly, multipotency is based on the fact that stem cells have the ability to generate through differentiation all cell types of the tissue in which they are. These are the characteristics that share any type of stem cell and the term of stemness is used to combine these two concepts that distinguish a stem cell from other cells [1]. The fundamental characteristic of adult stem cells is the ability to produce differentiated cells and maintain the stem cell pool, these abilities allows the regeneration of tissues. Adult stem cells in human body are localized in a niche in a quiescent state, a state of reversible mitotic arrest characterized by an attenuated metabolic activity [2], under certain stimuli such as the necessity of regeneration because the tissue is damaged, a set of factors promotes the activation of stem cells, through a specific procedure this cells proliferates and are converted to progenitor cells, finally differentiate by the fusion in order to form specialized cells, the process consist in the pass from G0 phase characterized by quiescence to S and M phases that are characterized by differentiation. The capacity of self-renewal allow that a part of a cells that escape from the process of differentiation and returns to G0 phase, this cells are reserve stem cells that restore the stem cell compartment and avoid a complete deletion of stem cells in order to assess the regeneration function.

We can classify stem cells in different types depending on the tissue in which they are localized, this fact implicates that every type of stem cell work in a particular form. Germline stem cells are pluripotent stem cells derived from the gonadal ridge of embryos and are the precursors of gametes: the egg and sperm, this type of stem cells are present in a niche and the factors of the niche regulate the balance between self-renewal and differentiation [3]. Embryonic stem cells (ESC) are pluripotent stem cells that can generate each of the tissues belonging to the embryo; these can be isolated from the Inner Cell Mass and maintain them in culture in order to obtain different embryonic lines. The hematopoietic system contain complex levels with cells in different stages of differentiation, the hematopoietic stem cells (HSC) are found in the more immature stage of the whole system and are the responsible to the constant renewal of the blood [4]. The regulation is controlled by a set of genes interacting in order to accomplish the function of stem cells that is regeneration. We can isolate these cells from umbilical cord and bone marrow. The performance is based in the differentiation of stem cells to progenitor cells that can lead to different lineages of blood cells. Neural stem cells (NSC) have been recently discovered and have great potential, this cell type is localized in different sections of the brain and is able to differentiate to progenitor cell and generate new neurons. Skin follicles are structures that present different stages, the skin stem cells (SSC) are responsible to the performance of this stages, consists in a anagen cycle characterized by the growth of hair follicle, a catagen cycle that consist in the destruction and the telogen cycle that is a resting
cycle. The cyclic regeneration and growth of the hair follicle is due to the stem cells of the skin that are in different locations depending on the species [5]. Muscle stem cells (MSC) also referred to as satellite cells work in order to do the regeneration of different parts of the muscle.

It is necessary taken into account that stem cells may be localized in different organism. If we want to study their presence it is important to know that the morphology, and particular mechanisms implicates differences in the performance of stem cells depending on the organism.

The presence of different types of stem cells working in a particular way and the existence of stemness allows us to classify them as a regenerative system that is formed through the development of the body. In multicellular organisms the origin arises from pluripotent stem cells that are located in the Inner Cell Mass (ICM) of the blastocyst. These stem cells divide giving rise to other cells that differentiate into all cell types of the endoderm, ectoderm and mesoderm. The potential of differentiation is restricted as the cells become specialized, the origin begins with totipotent stem cells that can generate any cell of the three germ layers and extraembryonic tissue, the pluripotent inner cell mass cells can generate cells that are observed belong to the three cell layers, when we analyzed adult stem cells we observe multipotent cells that are capable of generating cells of the same germ layer and unipotent cells that are able to differentiate into one type of cells. This type of development allow to the adult organisms to present stem cells that are necessary elements for a principal function that is regeneration.

The capacity of regeneration in stem cells has long attracted attention from scientists. There is present a necessity of understand and know how stem cells possess the properties of self-renewal and multipotency. This fact leads us to investigate what elements are acting behind these cells and how these elements works in order to maintain the identity of stem cells. Any progress in this area provides knowledge of performance and regulation; this it means that we need to find elements acting in the expression of stem cells. Therefore, we need a continuous research in this area that allows Scientists focus in aging cycles and new therapies for diseases. The necessity of investigation in this field let me to present the goal of my work that implicates an in synico and in vitro analysis that demonstrates the presence of a master regulator in stemness.

My project is focused on the analysis of gene expression in stem cells, using statistical tools I can collect, combine and evaluate the results of previous research. It is intended to integrate in a structured way the information from different studies in order to obtain a new result from an investigation that had not been done before. Specifically, I collect experiments in which stem cells are isolated in order to realize a global gene analysis that shows genes that are related with the properties and performance of the stem cells. In order to obtain a robust analysis, the study is realized for different types of stem cells that are germline, embryonic, hematopoietic, neural, skin and muscle stem cells in different organisms that are Drosophila melanogaster, Caenobirditis elegans, Danio rerio, Mus musculus and Homo sapiens.
These data allow to perform a comparative analysis in which a gene list for each cell line is derived, it is obtained a list of genes of the same cellular type that belong to different species. It is realized a data analysis in which these gene lists are uploaded in different programs in order to obtain elements such as transcription factors, microRNA and pathways that act behind specific set of genes, the performance of this elements consist in the regulation and control the expression of the genes. The main goal is to find a common component in all lists of genes that regulate and control all the types of stem cells in all the species that are analyzed, this objective let me to purpose different questions that today still do not have a clear answer. The basic question that arises at the beginning of the project and also is proposed by another study: Do all stem cells express a similar set of “stemness” genes necessary for their unique properties, or do different stem cells express different sets of genes that confer stemness? When it is focused the analysis in the genes, we consider the possibility of the existence of a general component controlling these genes and then this fact let me to the second question: Exist an element that is common to different types of stem cells? If this element exists, it could be the key for the maintenance of stem cells? And finally, in the future we can arrive to regulate this component in order to control processes such as regeneration, cellular aging or the development of different diseases?. The methodology of my project presents results that allow responding these questions.

The analysis shows the existence of common components present in all the types of stem cells that belong to different species, the analysis of significance lead to focus the investigation of one common component that is FOXO3 transcription factor that is a member of the forkhead box (Fox) family of proteins, the family of FOXO contains different FOXOs factors that are present depending on the specie. We compare our data with a data that we have available and belong to hydra, an immortal organism that maintains the longevity through a capacity of indefinite self-renewal. We have found the presence of FOXO3 in Hydra and the literature affirms that FOXO presents the role of maintenance of stem cells, as a consequence is a critical regulator of longevity and indefinite self-renewal in this specie, at this stage we hypothesize that FOXO3 can present the same role in other species.

The next step consists in to prove that the hypothesis obtained through in silico analysis is correct, for this demonstration we design two experiments, with muscular stem cells, that are in vitro quiescence assay and in vitro differentiation assay, this analysis pretends to demonstrate that stem cells need the transcription factor FOXO3 in order to maintain the capacities and properties that are specific for stem cells. Mainly, we compare and study behaviors between wild-type cells that express the normal phenotype and knockout cells that present a deletion of FOXO genes, for this purpose we design a system for the deletion of genes. When these types of cells are subjected to two situations, that are quiescence and differentiation, we can observe different behaviors that allow defining the role of FOXO3 transcription factors. Here we show that FOXO3 is a critical regulator of stemness and is classified as an important contributor in the maintenance of stem cells. Thus, it would be finding a key element that is
probably involved in the activation and proliferation of stem cells and then we can improve our understanding in human cellular aging, regeneration processes and maybe purpose new therapies for different types of diseases or aging processes.

The first part of the project implicates the compilation of different types of information from different databases and this fact implicates that arises complications in every stage of the project. An extensive literature search of different type of data is necessary to make a correct analysis.

Despite all the sources of information available today is hard to find a good base of information and I need an advanced search in order to get results in different phases of the project. First, I need a set of articles in order to learn and understand basic concepts and principles that defines the base of the project. Secondly, we can see that as the project progresses more specific information is needed and then is more difficult to find useful information.

One of the most important limitations is the necessity to combine information from experiments that have been carried out with different protocols. Each study depends on a set of programs used and each analysis is done in a certain way. In order to work with data, get comparisons and obtain robust results we need a consistency and similarity in experiments. It has to conduct a search for models that have been made with similar protocols and this requirement it is not always easy because there are not so many experiments that are similar.

The project also depends on the level of research and studies that have been made. I need to find samples of different cell types in different species but today there are cell types that could not be isolated or simply have not been performed expression studies with certain types of stem cells.

In addition, another difficulty is present and is related with the performance of the collected studies. Although the protocols are similar there is always the problem of the algorithms used by each program, the algorithm influence in the obtained data. On the one hand you can obtain lists of data with restrictive programs that uses complex algorithms that impose their own parameters. On the other hand, the lists can be obtained with low restriction but we obtain irrelevant data. Thus, I need to take into account that exist the possibility in which a comparative analysis comes from two different programs that generate lists with different parameters and different restrictions.

The second part of the project implicates the realization of two in vitro assays; the culture and different conditions that I need to introduce in the experiments lead us to the existence of different limitations. Mainly, is important to work with sterile samples and this fact is complicated due to the facility of contamination of the samples. In addition, we need to infect the cells with different drugs and the infection generates cell death, we have to be able to maintain the viability of the cells.

You can see that this project requires the necessity to integrate data in a logical and coherent way in order to make robust comparative analysis and extract useful information. Also a proper handling of cell cultures is required to obtain successful results.
Study of the Literature

There have been realized different types of investigations related with stem cells. In the meta-analysis of my project is necessary the existence of gene expression studies in different types of stem cells in order to collect integrated information and realize a comparative analysis, as a consequence we need studies that treat the expression of genes and take in to account the existence of analysis that discover common genes between stem cells. Also we need to know what types of studies realize a transcriptional and regulatory analysis. In addition, we revise the existence of in vitro assays that works with transcription factors FOXO in muscle progenitor cells. Therefore it is necessary to explain the state of the art in order to observe that my project presents a study that until now it is not realized.

If the basis of the project is analyzed, we can see that general isolation studies of different cell types are needed. It can be seen that there are different methods available in order to do the isolation and study of stem cells. The label retention assay is a powerful tool for analyzing the properties of cells and identifies whether the sample contains stem cells, thus we can obtain the isolation. This type of study is limited in terms of information and it is needed to use other techniques that allow to obtain a more precise information. In vitro culture is an assay that has more advantages because we can perform a live-imaging highthroughput that allows knowing the behavior of stem cells by manipulating the stemness potential. A technique that replaces this type of test is the transplantation which is considered today the gold standard, this type of test consists in the transplantation of stem cells in a living organism and then we can study the behavior of stem cells in vivo. A powerful tool that is developed in recent years is the in-vivo lineage tracing, with this type of study we can realize experimental and stemness tests [1]. The information that it is used in my anlaysis comes from this type of isolation assays.

Now I will explain studies that are present today related with different stem cell types in different species that I have used in the project. Hydra is an organism that due to its long life draws attention to Scientists and then exist numerous studies reporting the characteristics of his particular genome. Specifically, it has conducted a study with transgenic polyp Hydra, the researchers discover molecular signatures of the three stem cell lineages that reveal to every family a specific set of genes and Transcription factors that have an important role related to the maintenance and control of stemness [7]. We can also find genome analysis in other variants of Hydra such as Hydra magnipapillata or Hydra vulgaris that reveal important features such as the formation, dynamics and structure of the genome, with this information it has been made a comparison of the genome in other animals and it has obtained information on evolution, transcription factors and pluripotent genes[8], [9]. Furthermore, studies have been performed in which it is shown that FOXO is a regulator in the maintenance of stem cells in immortal Hydra, the researchers identifies this transcription factor as a critical driver that regulates and controls the capacity of self-renewal. This fact indicates that FOXO controls the longevity of Hydra and has a great relationship with aging [10].
The cells that belong to the germline type are difficult to isolate, this fact implies the difficulty in order to obtain a model for the study of this cell type. This analysis was conducted in Caenobhirditis elegans and Drosophila melanogaster, the models presents information such as lists of genes that are upregulated and downregulated and studies related with gene ontology, pathways and transcription factors.

In C.elegans species there are published two studies by the same author in which an analysis is performed in the expression of genes to identify germline cells, using mutants that cause a defect in proliferation we obtain germline sets of genes with enriched expression in both sexes, particularly with this study note that the germline genes are not randomly distributed in the genome[11], [12].

In Drosophila species there are expression studies with germline and neuronal stem cells. A method for purification germline stem cells in Drosophila ovary is developed, the researchers found upregulated genes that delay the cytokinesis in premitotic germline cells. In this analysis is suggested the possibility that stem cells use at least two regulatory mechanism common in Drosophila and mammalian [13]. The neurobiological analysis for this specie has developed a unique study that applies a method to isolate numerous differentiated neuroblasts and neurons. Transcriptional profiles are determined by RNA sequenciation and 28 transcription factors are identified [14].

In danio rerio has been isolated only one cell type, these are hematopoietic stem cells. Microarray data obtained from the isolation of the teleost kidney tissue that contains two populations which are side populations and main populations. They can characterize stem cells and progenitor stem cells. The information that they obtain consist in the presence of up-regulated genes involved in cell junction and signal transduction, also there are presence of down regulated genes that participates in DNA replication [4].

The species mus musculus and Homo sapiens present a large number of studies with different types of stem cells. Currently, in these species are defined exhaustively a genetic and transcriptional analysis on different types of stem cells.

In the case of mus musculus we can observe that the Scientist realizes numerous studies with different types of stem cells. There are methods of isolation that studies embryonic stem cells in mus musculus, the identification of genes using cDNA provides information of up-regulated and down-regulated genes [15]. There is evidence of hematopoietic stem cells present in the bone marrow and umbilical cord, the scientific isolate stem cells and progenitor stem cells and realize a study of the expression of genes [16]. In a similar manner, there are present studies that consist in a gene analysis of expression in neural stem cells [17]–[19], skin stem cells [5]and muscle stem cells [20], [21]. Specifically, in the case of neural stem cells that belong to mus musculus exist studies that investigate the role of FOXO3. It has been found that FOXO3 promotes the extension of lifespan and regulates the NSC pool through the induction of genes;
this fact means that the transcription factor prevents the premature depletion of neural stem cells [22].

In the case of the Homo Sapiens there are present specific studies with different cell types. To examine the differences and similarities, it has performed an analysis of expression in different embryonic cell lines belonging to human. Specifically, the Scientists isolate cells that belongs to HSF-1, HSF-6 and H9 lines, thus they can make a comparative analysis and obtain 7385 up regulated genes common to all three cell lines [23]. In this comparative analysis are highlighted genes such as LIN28, OCT4, NANOG, DNMT3B, TGIF, TDGF1, CHEK2, GDF3, GJA1 and FLJ21837. Although many articles have characterized the transcriptome of human hematopoietic cells, only one study has identified genes related with fate decision. Specifically, the transcriptome of umbilical cord blood and bone marrow is compared [24]. The isolation of neural stem cells implies the necessity of tissue that belongs to the brain, there are few studies that have been obtained brain tissue which contain stem cells, for the analysis we have neural stem cells and the application of serum allows the differentiation into astrocytes. Thus the researchers obtain expression profile of genes involved in self-renewal and multipotency that may be the key elements to treat neural diseases [25]–[27]. When we search studies that isolates skin stem cells we can observe that only one assay realizes an exhaustive analysis that characterize skin stem cells, the analysis consist in a global gene expression that studies the bulge cells and identifies surface markers that are unique to characterize the cyclic system carried out by this cell type [28].

Nowadays, we can observe that are present studies with an exhaustive analysis of one cellular type. The analysis of the transcription factors or studies with pathways is done in different cellular types but at this moment we don’t have a study that combines the expression of all the stem cells in all the species. Although the Scientists have conducted comparative studies with different cell types, this comparison belongs to two or three cell types and we need to do a global comparison with all the types of stem cells that belong to different species and this is the goal of my project.

The majority of comparative studies are in mus musculus. Mainly, there are two studies with different protocols that compare the same types of cells that are embryonic stem cells, neural stem cells and hematopoietic stem cells. On the one hand, we can study the analysis in which the researchers demonstrated that the three types of stem cells shared a genetic program. Specifically, there are 216 genes common in the three lines. In addition, there is present the transcription analysis of the genes that belong to each cell type and they discovered similarities between embryonic and neural stem cells [15]. On the other hand, exist another study with the same types of stem cells but includes a comparison between murine and human hematopoietic stem cells that shows common genes and pathways. The overlap result shows a molecular signature that is proper of stem cells [29]. In this area, exist an study in which there is compared the previous analysis in order to check results and the final conclusion is the presence of genes that are common to the analyzed stem cells [6].
We can find a comparative analysis of neural stem cells in homo sapiens and mus musculus, it is obtained a common genetic program that belong to this two species and a comparative analysis is realized in order to find a common set of genes [27].

The last comparative analysis is with hematopoietic stem cells in different species that are danio rerio, mus musculus and human. There are identified evolutionary conserved signatures common to the three species. The analysis presents a set of stem-cell specific genes that consist in 40 up-regulated genes and 40 down-regulated genes that are common to the three species [4].

At this stage we can observe the existence of gene expression studies in different types of stem cells depending on the organism, also there are few studies that treat to discover common genes between different cell lines and realizes an analysis of the elements acting in these common genes. Note that the studies belong to two or three types of stem cells but until now not exist a study that realizes a global comparison in order to found common genes.

The second stage of the project consists in the realization of two in vitro assays. There have been realized quiescence assays with the utilization of BDM drug; the studies are realized with C2C12 myogenic cells [30] or with human myoblasts [31]. The experiments pretend to investigate the effect of BDM drug.

There are present different studies that can be related with self-renewal because presents the analysis of reserve cells, mainly is pretended to demonstrate that stem cells in differentiation produce a population of differentiate and undifferentiated cells, this undifferentiated stem cells are the reserve stem cells [32], [33].

Finally, I need to comment the existence of an experiment in which the Scientifics realize a knockout of Foxo3a in wild-type cells. The results demonstrate that the cells with the deletion lose the capacity to be stem cells, but the observed phenotype is very small then It is needed a consistent result [2]. At this point we can observe that not exist a consistent study that proves the role of FOXO transcription factors and my project tries to prove with a high significance the importance of FOXO.

Therefore, I can say that there are present studies with different types of stem cells and that is demonstrated the role of FOXO in isolated cases but until now not exist an study that consist in the meta-analysis and demonstration of the presence of FOXO in all the types of stem cells and this is part of the goal and results that belongs to my project. As I said, I demonstrate the presence of FOXO through in sylico analysis and I check the importance of this factor through an in vitro analysis that until now the scientist not realize. I use the information that I comment in this section in order to design my project and obtain new results that reveals the existence of a key element controlling stem cells.
Methodology

The project presents two differentiated parts. First of all it is needed an in silico analysis in order to obtain a consistent result. When I obtain the results that belong to the meta-analysis, we can design two experiments in order to demonstrate the general hypothesis that is obtained previously through in silico analysis.

In silico analysis

Data collection

This analysis is the first stage that is required in order to realize the project. First of all, I do a compilation of different studies that presents a list of genes that belong to different types of stem cells in different types of species. As I say before, I need similar experiments in order to obtain coherent results, then is needed an advanced search in order to find adequate analysis.

The experiments that I use are summarized in Table 1, we can observe that there are different authors that realize specific analysis with certain characteristics. This fact leads us to impose a set of requirements when I choose the experiments, it is needed the similarity in the protocols and in the comparisons realized in the analysis, the freshly isolation of the tissue that contain stem cells and a correct purification of this stem cells. In general, the procedure consists in the execution of different steps:

- The isolation of the tissue that contains stem cells constitutes the first step. The Scientists extract a specific part of the tissue; the localization depends on the type of stem cell that we are analyzing.
- When we have the isolated tissue, we need to extract stem cells. In order to purify stem cells we can use two methods that are Fluorescence-activated cell sorting or gradient sorter. Both technologies are based in the sort of the cells and present a separation between stem cells and the rest of the cells that contain the tissue. In this step, it is obtained a population of purified stem cells.
- The possibility to isolate stem cells allows Scientists to do different experiments but my project is focused in a global gene analysis, as a consequence it is needed an analysis of expression profile of this stem cells. Mainly, there are present two technologies that allow studying genes. The typical analysis uses DNA Microarray, we need a sample of mRNA and the conversion to cDNA, this cDNA will be hybridized with a chip that give us the genes that belong to the stem cell analyzed. Recently, there is present a new technology for expression analysis and is RNA Sequencing, it uses abilities of next – generation in order to obtain a direct analysis of genes. The difference between these methods is the experiment, in the first case we need a study of genes that are expressed in different conditions, then we need a comparison. In the second case, the
technology sequences all the genes in the sample and we don’t need different conditions.

Note that this general procedure is followed by different authors that I enumerate in Table 1. If the researchers follow these steps properly we can guarantee good results.

<table>
<thead>
<tr>
<th>Stem Cell Type</th>
<th>Specie</th>
<th>Isolated tissue</th>
<th>Method for purification</th>
<th>Expression analysis</th>
<th>Experiment</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germline</td>
<td>Hydra vulgaris</td>
<td>Interstitial, Ectodermal and endoderm epithelial</td>
<td>FACS</td>
<td>RNA Seq (Roche 454)</td>
<td>Common genes in the three populations</td>
<td>Hemmrich et al 2012</td>
</tr>
<tr>
<td>Germline</td>
<td>Caenorhabditis elegans</td>
<td>Wild-type larval stages</td>
<td>FACS</td>
<td>DNA Microarray</td>
<td>Using mutants that cause defects in germ cell proliferation</td>
<td>Reinke et al 2004</td>
</tr>
<tr>
<td>Germline</td>
<td>Drosophila melanogaster</td>
<td>Wild-type drosophila ovary</td>
<td>FACS</td>
<td>GeneChip Affymetrix</td>
<td>Analysis by overexpression of Dpp or by mutating the bam gene</td>
<td>Kai et al 2005</td>
</tr>
<tr>
<td>ESC</td>
<td>Mus musculus</td>
<td>Lateral ventricle brain and bone marrow main population in embryo</td>
<td>FACS</td>
<td>GeneChip Affymetrix</td>
<td>Comparison of two populations that contains ESC</td>
<td>Ramalho-Santos et al 2002</td>
</tr>
<tr>
<td>ESC</td>
<td>Homo sapiens</td>
<td>Three stem cell lines (H9, HSF-1 and HSF-6) derived from the embryo</td>
<td>FACS</td>
<td>GeneChip Affymetrix</td>
<td>Common genes that belong to the three lines</td>
<td>Abeyta et al 2004</td>
</tr>
<tr>
<td>HSC</td>
<td>Danio rerio</td>
<td>Two populations (SP and MP) from the adult zebrafish kidney</td>
<td>Hoechst fluorescence</td>
<td>GeneChip Affymetrix</td>
<td>Comparison of two populations that contains HSC and HPC</td>
<td>Kobayashi et al 2010</td>
</tr>
<tr>
<td>HSC</td>
<td>Mus musculus</td>
<td>Bone marrow cells</td>
<td>FACS</td>
<td>GeneChip Affymetrix</td>
<td>Comparison between subpopulations that contains HSC</td>
<td>Manson et al 2007</td>
</tr>
<tr>
<td>HSC</td>
<td>Homo sapiens</td>
<td>Rho cells from umbilical cord and bone marrow</td>
<td>FACS</td>
<td>GeneChip Affymetrix</td>
<td>Comparison between Rho(lo) cells that constitute HSC and Rho(hi) cells that are HPC</td>
<td>Eckfeldt et al 2005</td>
</tr>
<tr>
<td>NSC</td>
<td>Drosophila melanogaster</td>
<td>Larval L3 stage.</td>
<td>FACS</td>
<td>GeneChip Affymetrix</td>
<td>Comparison between neuroblast and differentiation cells</td>
<td>Berger et al 2012</td>
</tr>
<tr>
<td>NSC</td>
<td>Mus musculus</td>
<td>Neuronal population</td>
<td>FACS</td>
<td>Illumina Expression Bead-Chip arrays</td>
<td>Comparison between Neurons and NSC</td>
<td>Burney et al 2013</td>
</tr>
</tbody>
</table>
We can observe the existence of a limitation; there are specific stem cells that until now the Scientifics have not isolated or have not realized studies of expression. Thus, I have a collection of gene list that belong to experiments that exist until now.

**Data analysis**

This gene list presents a set of parameters that inform of the regulation and significance of each gene. I need to understand each parameter in order to study and select different genes. The lists often presents a set of informative dates that are the id and gene symbol that is the official name of the gene and a set of bioinformatics statistical dates that are the p value and logarithm Fold Change, these parameters are necessary in order to obtain specific information related with the performance of the genes. The Fold Change consists in a value that indicates how much has changed a value from the beginning to the end, in terms of analysis expression indicates how regulated is a gene in the process of change from stem cell to something more differentiated. The output that is presented is a logarithm of the fold change and can be a positive value which means that the gene is up-regulated in the process or a negative value which means that the gene is down-regulated in the process. The p value inform of the significance in the results, in terms of a list of genes this parameter is useful in order to determine if the resulting genes are really present in the process. Different problems arise when I examine the different list of genes and is needed the utilization of a set of programs. The first problem is that the author’s presents a set of genes in a particular form with different type of parameters and I need the same parameters for all lists, I realize specific calculations in order to
obtain the basic parameters. The second problem is related with the gene symbol, in the majority of the cases the list presents the id of the gene but I need the gene symbol for the comparison. I use DAVID program that is a type of software which converts the id to the official gene symbol. The third problem arises with list of genes that belong to species such as Drosophila melanogaster, Caenorhabditis elegans and Danio Rerio. I can't compare this type of genes because are specific to the specie, I need to use DIOPT Software that compute the orthologs of the respective genes. If I upload the list of genes in DIOPT, I obtain an output of genes that are comparable with Mus musculus and Homo sapiens. The last problem arises from the particularities and restrictions imposed by the researches in the experiment, sometimes the list of genes are presented as a supplementary table that arises from specific parameters and then there are irrelevant genes or insufficient genes for my comparative analysis. It exist an option to reanalyze the dates of a specific experiment with GEO2r that is a tool that allows to create groups with the samples and realize a particular comparison with the samples of the experiment, this fact allows to obtain a list of genes more extensive with a specific comparison. When I resolve these problems I obtain different lists that present the same parameters and inform us about the existence of different genes and their regulation. At this stage the lists are prepared for the comparisons.

The second step is to implement the design of my analysis, we can observe a general scheme in Figure 1 that shows the methodology which I will explain below. It is realized a comparative analysis with my collection of gene lists. I design two strategies in order to do the comparisons. The first strategy consist in the collection of genes that belong to the same cell type but are present in different species, then I obtain lists of common genes of the same cellular type belonging to different species. The design of the second strategy depends on the species; this type of analysis allows the study of the different types of stem cells belonging to certain species. In this case I want to study the genes that control and regulate the expression of stem cells in a positive way, and then the comparisons are made with up-regulated genes. I follow two requirements for the selection of genes, for each strategy I do comparisons with all the up-regulated genes and only with up-regulated genes that presents a fold change greater or equal to 1.5.

The software that I used to compute the common genes between the lists is Venn diagram program, it is a platform that belongs to Bioinformatics Evolutionary Genomics and calculates the intersection between lists of elements. The output consists in a Venn diagram and a table which indicates the elements that belong to an intersection or the elements that are unique to a specific list. When two strategies are executed, I obtain a set of lists and it is needed the analysis in order to select the best strategy.

The third stage consists in the realization of data analysis. Specifically, I make two types of studies. The first step is a basic study of the data in order to obtain a functional annotation that classify the genes in different categories and inform of the specific functions in the genes. The
second step consist in a complex analysis that implicates the use of different programs, I try to find elements acting behind this genes and elements that change the properties of this genes. These elements are transcription factors, microRNAs and pathways that regulate and control the genes, also exist a program for the calculation of chemical and genetic perturbations of the genes. The functional annotation is extracted with DAVID program, for the calculation of the rest of elements I use different programs. The transcription factors are calculated with GSEA and Opossum Softwares, the microRNAs are calculated with GSEA, the pathways are calculated with DAVID and GSEA Softwares and the chemical and genetic perturbations are calculated with GSEA Software.

David Software is a program that provides different analytical tools with the capability of execute algorithms and obtain different type of information. One of the characteristics is the availability of around 1.5 million genes from more than 65,000 species, this fact allow computing the functional annotation that consist in a functional classification of the genes. The system works with a novel algorithm that search relationships of the updated genes with the available databases and present an output with genes aggregated in categories, through this output it is possible to search the existence of possible pathways acting in these genes. The algorithm for the calculation of pathways search the overlap between upload genes and the KEEG Database that contains a large set of genes aggregated in networks.

Gene Set Enrichment Analysis (GSEA) is a software that presents a computational method which allow to introduce a set of genes and obtain different type of information. The program has available the Molecular Signatures Database (MSigDB) that contains different collections of genes. Mainly the algorithm used for any type of calculation is focused in the search of commonalities between the uploaded genes and their own database. I use this program to obtain microRNAs that shares the 3'UTR with the uploaded genes, transcription factors that are common with TRANSFAC Database which is a sub database belonging to MSigDB, chemical and genetic perturbation that shows elements acting like a perturbation and canonical pathways are also calculated through MSigDB.

Opossum calculates the transcription factors that bind with the uploaded genes. The software uses two algorithms, the single site method look at each binding site and the anchored combination site method look through clusters of binding sites. The transcription factors that I obtain are calculated with the single site method. The output shows a table with transcription factors and the genes that are regulated by this factor and there are two important parameters. The Z-Score compare the rate of occurrence of transcription factors in my list of genes with the expected rate that is calculated previously. The Fisher compares the expression of genes that contains a transcription factor binding with the expected proportion of genes that contains this particular site. The parameters inform of the significance, Z-score is more restrictive than the Fisher Score and then we obtain different set of genes depending on the parameter.
Note that sometimes the same information is calculated with different programs due to every program uses his own algorithm.

In this stage I have every list of genes with all the information about elements that regulates these genes. The next step is to do a second comparative analysis with the goal to obtain a common transcription factor in all the list of genes. When I generate this analysis I obtain common transcription factors and it is needed the examination of the genes that are controlled by the different transcription factors in order to decide if exist a key element controlling stem cells.

It is realized an analysis of significance that depends on Fischer score and Z-Score, the results of this type of analysis informs of significance and allows focusing in one transcription factor.

The results of this part of the analysis arise in the different stages of the process and the final conclusion, obtained in the last stage, constitutes an important result that is related with the discovery of a key element that regulates the maintenance of stem cells.

Figure 1. Scheme representing the procedure that it is followed in data analyses of in syilco methodology
In vitro analyses

In this part of the project it is decided to design two specific experiments that are in vitro quiescence and differentiation assays. The analyses are realized in the Cell Biology lab (DCEXS). This study allows proving the hypothesis obtained through in silico analyses that consist in the demonstration of the importance of FOXO3 in the regulation and maintenance of Stem Cells. It is pretended to do a comparison and determine the differences between two types of muscular stem cells. Specifically I determine the function of FOXO family that is an important regulator in stem cells.

Cell Culture

The cells are derived from the muscle of Mus musculus; these cells are purified by fluorescence activated cell sorting (FACS). As a consequence muscle stem cells are obtained. We have worked with two types of stem cells, the first type are wild-type muscle stem cells which express the typical phenotype without modifications, the second type are knockout muscle stem cells that are cells which expression contains the deletion of one or more elements, in this case are cells with the deletion of FOXO1, FOXO3, and FOXO4 genes. Initially, we have available wild-type cells and then it is needed the creation of knock-out cells through the implementation of the system Cre/LoxP that I would explain later. Specifically, at this stage we have available wild-type cells that belong to three lines:

- AQM 312 and AQM 313: Pax7 – CRE – ER- FOXO 1,3,4 FLOX – YFP
- APO 46: Pax7-CRE-ER-FOXO1,3,4 FLOX

Wild-type cells are thawed at 37ºC. It is needed the mechanical cell separation from the medium and then I centrifuged the samples in order to obtain the cell pellet that corresponds to the muscle stem cells. The cells are deposited in three cell culture dishes (plates) that previously are coated with collagen which allows a correct adhesion of the cells, this cells are cultivated with growth medium (GM) and basic fibroblast growth factor, the exactly composition it is specified in the appendix. At this stage we need the maintenance in a proliferation culture until we have sufficient cells for the experiments, for this purpose we need to control the confluence of the cells every day in order to avoid the accumulation of many cells that have the risk of fusion and differentiate. When it is observed a high percentage of confluence it is realized this specific protocol:

- Aspire the medium, introduce Phosphate buffered saline (PBS) and aspire PBS.
- Add 1 ml of Trypsin that is an enzyme which breaks the peptide binds allowing that the cells can float in the plate. The plate is introduced in the incubator around 4 minutes in order to obtain the desired effect.
- When the cells in suspension are obtained I can recollect and insert in the Neubauer Chamber in order to do a calculation of how many cells I dispose.
- It is decided an adequate dilution and introduce the cells in a plate with growth medium and basic.

We divide the cells in different plates and we obtain two plates for every line. The proliferation state it is maintained during 11 days, at this stage we have sufficient cells for the specific assays.

*Generation of Knockout cells: The recombination system Cre/LoxP*

The assays require knock-out cells that present the deletion of FOXO1, FOXO3 and FOXO4 genes. It is decided to delete the FOXO genes that are responsible the generation of FOXO transcription factors; this fact ensures that FOXO factors are not present. Note that we work with the deletion of three genes in order to observe a strong phenotype because only with the deletion of FOXO3 we generate a weak phenotype. There is a technique that generates a knock-out of wild-type cells. At the moment that I have an adequate quantity of cells proliferating in a plate we can infect the cells with an adenovirus that express the recombinase Cre, a genetic recombination of tyrosine enzyme. This concept arises from the recombination system Cre/LoxP that is characterized by a specific performance. LoxP sites are introduced and positioned in the sequence of DNA that I want to delete, the recombinase Cre recognize this sites and binds to this specific positions in order to cut the sequence defined by LoxP and paste the extremes. This type of cells are called floxed because they come from a mouse that presents LoxP sites positioned, the introduction of this sites not triggers any change because it is necessary the recombinasa Cre for the deletion. For the implementation of this deletion system we infect one plate of cells of each line with 6.5 µl of adenoCMV-Cre that express the recombinase Cre, this cells constitutes knockout cells. It is necessary to realize a correct comparison of behaviors, for this purpose we need to infect the remaining plates with 6.5 µl of adeno-CMVGFP that allows to visualize the cells in green color, the infection with GFP not generates any change and this cells constitutes wild-type cells. At this stage we generate Cre and Gfp cells for each line.

*Quantitative PCR*

Quantitative polymerase chain reaction is realized in order to check that the deletion with Cre/LoxP system is successful. The technique uses a method to amplify and detect the quantity of DNA present in a sample. In the case of my project, it is obtained a small sample which contains a few quantity of cells, as a consequence we decide to use RNeasy MicroKit (QIAGEN) that is the adequate for small samples.
The MicroKit presents a specific protocol that allows obtaining RNA from the samples:

- We introduce different types of buffers and the sample is centrifuged after every addition. The purpose is to lyse the membranes of the cells and the buffers allow the extraction of RNA.

At this stage we obtain RNA and it is realized the quantification through a nanodrop spectrophotometer that only needs a drop of the sample. The parameters present in table 2 are obtained through a specific program with the spectrophotometer; the most important data is the concentration of RNA and the parameter 260/280 that is a value which informs about the purity of the sample, the correct values are between 1.8 and 2.1. We realize a data analysis in order to calculate the concentration of RNA that we will use; it is calculated 225 ng from every sample.

The next step is the introduction of the concentration in the PCR device. It is needed the introduction of cDNA, as a consequence is necessary manipulate the specific concentrations of RNA in order to obtain cDNA. For this purpose we need to prepare the retro transcription reaction that consists in the pass of RNA to cDNA. It is retrotranscribed 250 ng of RNA.

- We prepare 0.5 µl of random primers that are necessary for cDNA synthesis and 1 µl of deoxynucleoside triphosphates (dNTP) that build blocks of DNA, the mix is introduced with the RNA at 65°C during five minutes. The final volume at this process is 13 µl.
- We prepare 4 µl 5X First-Strand Buffer, 1 µl 0.1 M DTT, 1 µl RNeaseOUT and 1 µl of superscrit . The mix is introduced at 25°C during five minutes. The final volume at this process is 7 µl.
- These two preparations are mixed and collocated in a eppendorf that constituters the final sample and contains a final volume of 20 µl.
The final sample is composed of cDNA, the forward and reverse oligos that are different depending on the analysis of FOXO1, 3,4 or myogenin and SYBR green that is a fluorophore that intercalates the DNA and allows the visualization of the gene expression.

The sample is introduced in the thermocycler that realizes the quantitative real time polymerase chain reaction (qRT-PCR). The samples are collocated in the device in different chambers in order to obtain gene expression results for WT and KO cells, we need the introduction of a housekeeping gen that is L7 which presents the same expression in any situation and it is used for the normalization. The protocol consists in the introduction of particular parameters on the PCR device at 50°C during 60 minutes and 70°C during 15 minutes in order to inactivate the reaction by heating. The performance of the thermocycler is based in the realization of 30 cycles with amplifications of DNA in every cycle. Finally, the device shows a specific graphic that represents the cycles versus the concentration of fluorophore.

*In vitro quiescence assay through infection with BDM*

This type of assay pretends to simulate a quiescence situation in a population of muscle stem cells, the induction of quiescence it will be accomplished with 2,3, butanedionemonoxime (BDM). We deposited wild-type and knockout cells from the lines APO 46, AQM 312 and AQM 313 in 6 microplates with 8 chambers, 4 chambers contain knockout cells and the other 4 contain wildtype cells.

- Three microplates are infected with BDM drug

We prepare 0.496 grams of BDM that is diluted with 0.496 µl of Dimethyl sulfoxide (DMSO); the dilution is filtered in order to avoid the introduction of particles. For the infection it is necessary the mixture of BDM with growth medium medium and basic fibroblast growth factor, taking into account the relation 2 ml: 6.06 µl it is calculated the volume that we need to introduce in the chambers. It is prepared 7 ml of GM with 21,21 µl of BDM and 0.7 µl of basic fibroblast growth factor, we introduce in every chamber 250 µl of the mixture. The BDM acts during 48 hours and at this point we fix the cells to the plate with paraformaldehyde (PFA) in order to detain quiescence. For this purpose we introduce 250 µl of paraformaldehyde 4% that acts during 10 minutes generating covalent unions, we remove paraformaldehyde and introduce phosphate buffered saline in order to eliminate a possible rest. Finally, we use a 500 µl solution of phosphate buffered saline with sodium acida that is added to every chamber in order to conserve the cells.

- Three microplates fixed with paraformaldehyde (PFA)

These microplates are the control of the experiment. We need to use a chemical fixative in order to fix the cells to the plate and avoid proliferation. For this purpose we
introduce 250 µl of paraformaldehyde 4% that acts during 10 minutes, we remove paraformaldehyde and introduce phosphate buffered saline in order to eliminate a possible rest. Finally, we use a 500 µl solution of phosphate buffered saline with Sodium acida that is added to every chamber in order to conserve the cells.

*In vitro differentiation assay*

In this assay it is pretended to study the self-renewal property. Mainly, the experiment consists in the introduction of cells in two plates that contains chambers with growth medium. We have three chambers for each type of cell. The second step is to remove growth medium and introduce differentiation medium (DM), the exactly composition is specified in the appendix. The cells are conserved in differentiation medium during 4 days in order to obtain a differentiation state, we aspire and introduce new medium every day. An Important point is that this type of assay not requires a control. After 4 days we need to use a chemical fixative in order to fix the cells to the plate and avoid differentiation. For this purpose we introduce 300 µl of formaldehyde that acts during 10 minutes, we remove formaldehyde and introduce phosphate buffered saline in order to eliminate a possible rest. Finally, we use a 500 µl solution of phosphate buffered saline with Sodium acida that is added to every chamber in order to conserve the cells.

*Fluorescence microscopy images*

We obtain images of the cells of every assay with fluorescence microscopy that belongs to PRBB.

**Results**

**Results of in sylico analysis**

The logic and the existence of consistent results is demonstrated when the analysis is done by cell type and not by species.

It is integrated a study of gene expression in stem cells and it is realized a comparative analysis. For this purpose we design two strategies in order to found commonalities between stem cells, we want to prove which strategy is the most correct and presents significant biological results. The first strategy is designed depending on the cell type of different species and the final result consists in different data that informs about the performance of particular types of stem cells. The second strategy is designed depending on the organism; we obtain results that belong to different types of stem cells in the same organism. In this case the results inform about the performance of stem cells in particular species. Figure 1 may help to understand the analysis realized in each strategy. In general, the realization of a comparative analysis of this type allows obtaining information about the genes that regulate stem cells. As a consequence when I observe the data the first results that I can comment is the existence of a
logical output that defines a base for the regulation and control of stemness. This logical output arises from the analysis that belongs to the first strategy, when the data is analyzed we can make a comparison and observe that the results define a consistent profile of behavior because exist a common set of genes that controls every type of stem cell in each specie. In change the second strategy present results that show a few set of common genes and this fact suggests the absence of a consistent set of genes controlling every type of stem cell present in a organism. Therefore, we obtain consistent results with first strategy that define a gene expression profile characterizing the behavior of particular stem cells. For this reason we decide to analyze and work with this strategy that provides coherent results and that we can observe in the following sections.

It is confirmed the existence of genes, pathways, microRNAs and transcription factors common to certain types of Stem cells

We obtain different results in the comparative analysis of genes that belong to particular stem cells. We need to work with this data in order to found elements that act behind these genes; these elements are pathways, microRNAs and transcription factors.

The analysis of the data indicates the existence of a common set of genes in particular types of stem cells that belong to different species, we can observe this fact in figure 2 that allows to respond the question that I present in the introduction: Do all stem cells express a similar set of “stemness” genes necessary for their unique properties, or do different stem cells express different sets of genes that confer stemness?[6]. It is demonstrated that not exist a common set of genes controlling stemness in all the types of stem cells, I can observe in the results of my comparative analysis that there are a different set of genes that confer stemness. The results show that the genes are unique to each stem cell. Supplementary tables show specific genes that belong to each overlap in figure 2. Although, it is decided that the analysis which depends on the types of stem cells present in particular species not presents significant biological results, there are supplementary images of corresponding Venn Diagrams and the data that belongs to the comparative analysis.

![Figure 2. Venn Diagrams showing common genes between different species in the same stem cell type](image-url)
As we can see, the comparative analysis with genes not reveals a common gene present in all the types of stem cells and is needed the analysis of elements controlling these genes in order to prove the existence of a key element controlling the stemness.

The analysis of the elements that are acting behind these genes confirms the existence of key elements common to all types of stem cells in every organism. The results show the existence of microRNAs, pathways and transcription factors controlling the behavior of particular types of stem cells. I select significant results that are present in table 3 and 4. The complete list of elements is present in supplementary material.

<table>
<thead>
<tr>
<th>Stem Cell Type</th>
<th>Total</th>
<th>microRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESC</td>
<td>3</td>
<td>MR-1, MR-206 and MR-506</td>
</tr>
<tr>
<td>MNC</td>
<td>1</td>
<td>MR-630A</td>
</tr>
<tr>
<td>Germline</td>
<td>1</td>
<td>MR-101</td>
</tr>
<tr>
<td>ESC</td>
<td>3</td>
<td>MR-510C, MR-510G and MR-518A</td>
</tr>
<tr>
<td>SSC</td>
<td>3</td>
<td>MR-200B, MR-200C and MR-429</td>
</tr>
<tr>
<td>ESC</td>
<td>1</td>
<td>MR-124A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stem Cell Types</th>
<th>Total</th>
<th>Pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESC</td>
<td>4</td>
<td>Cell cycle, mitotic DNA replication, HIV infection</td>
</tr>
<tr>
<td>Germline</td>
<td>2</td>
<td>mRNA processing, Processing of capped RNA containing pri-mRNA</td>
</tr>
<tr>
<td>NSC</td>
<td>1</td>
<td>Mitotic M01 phases</td>
</tr>
<tr>
<td>ESC</td>
<td>1</td>
<td>Metabolism of RNA</td>
</tr>
<tr>
<td>SSC</td>
<td>1</td>
<td>Transmembrane transport of small molecules</td>
</tr>
<tr>
<td>MNC</td>
<td>4</td>
<td>Extracellular matrix glycoproteins, Core Extracellular matrix, Focal adhesion</td>
</tr>
</tbody>
</table>

We can observe that microRNAs and pathways not are present in all the types of stem cells and are common to two or three types. In change, we can visualize in figure 3 the existence of 116 transcription factors common to all the types of stem cells, it is realized an analysis of significance that shows the presence of 15 common transcription factors in the area of significance and we visualize the presence of FOXO3. Supplementary tables show every set of transcription factors and also I generate a table with the enumeration of the 116 transcription factors that is supplementary table 1 in appendix.
In addition it is analyzed the functional annotation of every type of stem cell and I detect that the genes usually belongs to classifications such as nuclear lumen, nucleolus, intracellular organelle lumen, organelle lumen, membrane-enclosed lumen and nucleoplasm.

We can affirm that exist different elements that are common to different types of stem cells or to all the types of stem cells. Specifically, it is proved the presence of common genes, microRNAs and pathways but the elements that are present in all the types of stem cells are transcription factors. The information that it is obtained allows to define a profile of behavior, also it is clarified the performance of specific elements acting as regulators in stem cells.
In silico analysis demonstrate the existence of a key element that controls and regulates the maintenance of Stem Cells

As I say in the previous section in silico analysis demonstrate the existence of different elements controlling and regulating stem cells. Although the presence of different elements that are interesting, we decide to focus my project in the transcription factor FOXO3 that is present in every type of stem cell and in every organism. It is needed the analysis of different parameters in order to decide that FOXO3 is a relevant transcription factor that can provide significant results. It are elaborated different graphics that represents the values of z-score that depends on the number of genes that a transcription factor regulates, the area of significance is defined by the fact that the factors regulating a big set of genes are the significant factors. We can visualize the graphics in figure 5 in which we can observe that FOXO3 is localized in the area of significance and regulates an important set of genes in every situation that corresponds to every type of stem cell. As a consequence I can affirm that FOXO3 regulates an important set of genes.
The results of the graphic, the presence of FOXO3 in all the types of stem cells that belong to different species, the existence of literature that affirms that FOXO3 act as a regulator of stemness in particular stem cells and his implication on cellular aging lead me to decide to focus my project in FOXO3.

At this point I can respond to the question presented in the introduction: Exist a component that is common to different types of stem cells? Different studies and the analysis of significance predicts that FOXO3 is a consistent element and acts behind different set of genes that belong to different types of stem cells in different species.

Results of in vitro models

In vitro analysis demonstrate the hypothesis obtained through in sylico analysis

In sylico analysis conclude with the discovery of a key element that is present in all the types of stem cells in different species, this key element is FOXO3 transcription factor and act controlling and regulating a set of genes expressed in stem cells. When this component it is found, an important question arises: If exist a common component would be the key for maintenance of stem cells? Different studies and the analysis of significance predicts that FOXO3 would be essential for the maintenance of the stemness and these fact constitutes an hypothesis but it is needed a consistent investigation on the role of FOXO3 in order to prove that the hypothesis is correct, for this purpose I design two in vitro assays that demonstrate the necessity of this factor for the maintenance of stem cells, these assays work with two types of cells that are wild-type cells and knockout cells in which a deletion of FOXO 1, 3 and 4 genes, we delete more genes in order to ensure significant results with a more strong phenotype. We decide to delete FOXO genes that create FOXO transcription factors in order to ensure that a creation of FOXO factor is not possible. Importantly, we can verify that in vitro analysis demonstrate the hypothesis.
obtained through in silico analysis. The specific results of the deletion and the in vitro models are clarified in the following sections.

**Quantitative Polymerase Chain Reaction confirms the deletion of FOXO 1, 3 and 4 genes**

A quantitative polymerase chain reaction is realized in order to demonstrate that the deletion of genes with the system Cre/LoxP has been performed in a successful way. The analysis is realized for the gene expression of FOXO1, 3, 4 and myogenin. The results are present in figure 6 A, B and C that shows gene expression levels for different types of cells that are GFP cells (Wild-type cells) and CRE cells (Knockout cells). As we can observe the gene expression of FOXO1, 3 and 4 is reduced for Knockout cells, in change wild-type cells express elevated gene expression of FOXO genes. The results present in figure demonstrate that it is successfully generated the deletion of FOXO genes in Knockout cells.

![Figure 6. Results obtained from PCR](image)

(A), (B) and (C) Graphics showing particular gene expression for different FOXO genes.

(D) Graphic showing the expression of myogenin gene.

In addition, we decide to analyze the expression of myogenin in different situations because it is hypothesized that FOXO genes inhibits myogenin, a master gen of muscular differentiation. The scientists are working with this hypothesis in order to demonstrate this fact. In figure 6 D we can visualize that Knockout cells express elevated expression of myogenin due to its deletion of FOXO genes and wild-type cells present a decreased expression of myogenin, this fact may corroborate the hypothesized situation and provide new results.
In vitro assay of quiescence suggests that Knockout cells lose the capacity to be stem cells

To investigate the consequences of disrupt FOXO genes on stem cells, it is generated an induction of a quiescent state that is needed in order to study the capacity to produce the G0 arrest in wild-type and knockout cells. The G0 arrest consist in the detention of the proliferation status which is accomplished with 2,3, butanedionemonoxime (BDM) drug, a pharmacological inhibitor of the cytoskeletal that avoids differentiation and induces quiescence state.

First of all, it is demonstrated that the infection with BDM drug works properly. When we compare the control experiment with the BDM experiment, we can observe that the control experiment not presents any change and the cells show the same behavior that consist in a proliferation status. In change we note that BDM experiment presents a detention on proliferation.

As a result, it is observed that wild-type cells present the capacity to return to quiescent state with the application of BDM drug, wild-type cells presents the properties of stem cells and then are able to realize a G0 arrest and remain as cells in quiescent state, we can visualize this particular behavior in figure 7 in which the cells presents a round shape that suggest the presence of quiescence. The case of knockout cells suggest different results, this type of cells present a deletion in FOXO genes, when BDM drug is introduced we can observe that the cells are unable to realize a G0 arrest and as a consequence are not able to remain in a quiescent state, it is observed that this cells initiates a process of differentiation in order to generate a specialized cell type, the behavior is observed in figure in which the cells presents a elongation in the shape that suggest differentiation.

**Figure 7. Results of quiescence assay in wild-type and knockout cells with BDM drug**

*First column: Images from day 1 and 2 of wild-type cells
Second column: images from day 1 and 2 of knockout*
This behavior indicates the fact that knock-out cells have lost the properties of stemness and as a consequence all the stem cells become differentiated cells. Therefore, it is demonstrated the role of FOXO 1, 3 and 4 transcription factors in the maintenance of stem cells. Specifically, it is proved that FOXO factors are essential for the suppression of terminal differentiation allowing to reserve a percentage of stem cells that are able to re-entry in quiescence state, this percentage is necessary in order to have stem cells.

**In vitro differentiation assay suggests that knockout cells loose the property of self-renewal.**

Differentiation assay also it is realized to investigate the consequences of disrupting FOXO genes on stem cells. In this type of assay it is analyzed the capacity to maintain the self-renewal property in wild-type and knockout cells. For this purpose we induce a differentiation state with the introduction of differentiation medium and we compare different behaviors. Specifically, we demonstrate the necessity of FOXO factors in the expression of a self-renewal property.

![Differentiation Assay](image)

**Figure 8. Results on differentiation assay with different types of stem cells**

*First column represents two situations of the chambers that contain wild-type cells*

*Second column represent two situations of the chambers that contains knockout cells*

It is observed that wild-type cells in differentiation state present a percentage of cells that remains undifferentiated; this percentage constitutes the reserve stem cells that are cells that maintain the capacity of self-renewal because present the ability to escape from the process of
differentiation. As a consequence, we observe a heterogeneity population with the presence of differentiated cells characterized by elongated shape and reserve stem cells characterized by round shape, this population is represented in figure 8. In the case of knockout cells we can observe a population with all the cells in a process of differentiation and there is not presence of reserve stem cells. As a consequence we can observe a homogeneous population with only differentiated cells, this particular behavior is represented in figure 8. This fact leads us to say that the cells with a deletion in FOXO genes are unable to present reserve stem cells because all the population differentiates to a specialized cell, as a consequence not maintains the property of self-renewal. Therefore, this assay also demonstrates that FOXO factors are needed for the maintenance of stem cells.

**In sylico and in vitro assays suggest that FOXO is implicated in longevity and can be a potential marker in aging processes.**

The results that we obtain through in sylico and in vitro analysis provides information related with the role of FOXO in the maintenance of stem cells. Specifically, we have observed that FOXO factors allows the production of reserve stem cells that are responsible to maintain self-renewal and also this factors allows that stem cells entry in a quiescent state. As we can observe in previous results the presence of FOXO factors allows the existence and maintenance of stem cells. As a consequence it is logic affirm that FOXO is implicated in longevity. If we could maintain stem cells indefinitely we can maintain a continuous ability to self-renewal and it is demonstrated that the maintenance is possible thanks to the presence of FOXO factors. The continuous self-renewal lead to immortality and this type of behaviour is present in Hydra. We can affirm that FOXO3 is a potential marker as a solution of aging processes.

Finally, I can realize a question related with the applications of this discovery: In the future, the researchers can focus an investigation with FOXO directed to processes such as regeneration, cellular aging or different diseases? The information that I obtained is essential in order to investigate processes such as longevity in humans. Also the results can be useful for possible investigations in different therapies for different regenerative diseases.

**Discussion**

The results show the requirement of FOXO 1, 3 and 4 in the induction of stem cells quiescence and the property of self-renewal. Importantly, we can observe the ability of stem cells to realize a G0 arrest and as a consequence the re-entry in a quiescence state, also it is demonstrated the self-renewal property with the existence of a heterogeneous population that presents differentiated and reserve stem cells that are a subset of progeny that returns to quiescent state in order to guarantee the maintenance of stem cells. It is demonstrated through these specific assays that FOXO regulates the behaviour of stem cells by a mechanism that avoid a terminal differentiation and allows to control stemness, as a consequence a down-regulation of this
factor led to a reduction in the population growth rate [10]. An earlier study realized by Gopinath et al and other authors reports that a knock-out of FOXO3 gen in muscular stem cells leads to terminal differentiation characterized by a loose of stemness, they suggest that FOXO3 is a critical regulator of quiescence and self-renewal, this study only realize a deletion of FOXO3 [2], whereas my study consist in the deletion of FOXO1, 3 and 4 that establishes a more strong phenotype and as a consequence more consistent results. Although, the particularities of the study are different in comparison with my project, the results related with the role of FOXO are similar and both demonstrate the implication in the regulation of the particular properties of stem cells. A part of my experiment consist in the deletion of FOXO genes, this deletion is realized only in muscular stem cells, in change exist another study (Hu et al 2008) that generates the deletion of FOXO3 in every tissue of the mice, this fact involves that the results not shows sufficient accuracy, the correct protocol is the deletion when we have available only muscular stem cells. Specifically, self-renewal assay is consistent with other analyses of isolated GFP muscular stem cells in human and animal models that present an heterogeneity population with reserve stem cells and differentiated cells (day2010 Baroffio 1995,Quinn1985, rouger2004, yablonka-reuveni 1987, shefer 2006). A complementary study presents the same results working with C2C12 myoblast [32]. The results of quiescence assay are in agreement with Dhawan et al [30] that works with a G0 arrest in C2C12 myoblast, the basic observation is the behaviour of the cells with this arrest that consist in the ability to entry in a quiescent state, also the author demonstrate that the introduction of BDM generates a non-adherence between cells that allows to simulate a quiescence model, this fact is observed in my project. My experiments demonstrate that muscular stem cells were able to enter in a quiescent state depending on the infection with BDM drug. However, exist another method that induces quiescence, exist a reversible model that generate a G0 arrest in human myoblast through an alteration of culture condition [31], the results are similar results to my project.

The existence of different studies demonstrate that the transcription factor FOXO is involucrate in the regulation and control of different types of stem cells [2], [34], [35], this fact allow to observe that the results of my project are in agreement with different authors that affirms the existence of FOXO in particular types of stem cells. It is demonstrated that the disruption of FOXO in neuronal stem cells leads to an increase of neurogenesis and a depletion of stem cells, this fact suggest that FOXO decreases the ability to re-entry quiescence and maintain self-renewal [22], [35]. Another study works with the same hypothesis in hematopoietic stem cells and demonstrates that the depletion of FOXO leads to severe effects in hematopoietic stem cells influencing the behaviour [34]. The results of these studies are in according with my project in which the in sylloco methodology allows the discovery of a master regulator present in different types of stem cells in different species. In addition, the results of these studies suggest that FOXO is a critical regulator, as a consequence we can observe that in vitro results of my project presents consistency because exist different authors that affirm the same fact in other type of stem cells. In agreement with F.Anton-Erkleben et al [10] it is demonstrated that the project presents logical and consistent results, the author of the study concludes with the
identification of FOXO as a critical driver of the continuous self-renewal property in Hydra, this fact lead to think that the results that I obtain are correct.

Relative to in sylico analysis, this project provides new results that establish a base of regulation in stem cells, it is demonstrated the presence of microRNAs, pathways and transcription factor common between stem cells. This type of analysis is not realized until now and as a consequence I don’t have material to do the discussion of this part of the project, we can observe the presence of studies that works with microRNAs and pathways present in stem cells but these studies are involucrate in different disorders. Relative to transcription factors, exist different studies that works with transcription factors and it is known that FOXO is present in lineages of Hydra, in neuronal, hematopoietic and muscular stem cells [2], [10], [35]–[37] but until now not exist an study that works with a collection of studies in order to found FOXO as a common regulator in all the type of stem cells in different species.

Recently, it is observed an increase of the research in these areas and as a consequence is useful any type of information that reveals specific features that belong to stem cells and my project presents useful information.

The final results suggest the association of FOXO with longevity, is logical to affirm that FOXO is involved in the regulation of aging processes because exist literature that concludes with the same result [10], [22], [36]–[39]. FOXO is the responsible of continuous self-renewal in Hydra and also is implicated in longevity in other stem cells.

In summary our findings demonstrate a master regulator of stemness that is FOXO3 which is present in different types of stem cells and different species, in vitro assays shows a functional requirement for FOXO1, 3 and 4 in the maintenance of muscular stem cells. Specifically, knock-out cells present a loose of the ability to maintain stemness, as a consequence it is affirmed the role of FOXO in maintenance and regulation.

**Conclusion**

In this project is implemented an integrated analysis of different information that proceeds from different articles, the data analysis provide new results of a common element present in different types of stem cells that belong to different species, the key element is FOXO3 transcription factor. At this moment, the researchers found that FOXO is present in some types of stem cells but in this project it is demonstrated the presence of FOXO in all the types of stem cells. The existence of different parameters allows to hypothesize that this element is a master regulator of stemness.

Additionally, we decide to demonstrate the validity of the hypothesis through in vitro analysis with muscular stem cells. We describe the generation of in vitro models that demonstrate the important role of FOXO1, 3 and 4. We generate a successful deletion of FOXO genes that
allows studying the behaviour of stem cells when we induce the population to different situations based in the models that we design. Specifically, the assays permit the analysis of the mechanism involved in inducing and maintaining quiescence and self-renewal of muscle stem cells. It is observed the loss of stemness in knock-out cells; as a consequence we can affirm that FOXO factors are essential in maintenance and regulation of stem cells.

In adult skeletal muscle, the genetic mechanisms that regulate the behaviour of stem cells have only just begun to be explored and it is needed any kind of information that allows to understand processes that are acting on stem cells and my project presents an extensive study of the factors regulating muscular stem cells.

We know and we can affirm that FOXO controls longevity in Hydra, human and mice[22], [36], [38], [39]. Specifically, in muscular stem cells we observe that FOXO factors allow maintaining the properties of stem cells, as a consequence is regulating regeneration processes, this fact lead to affirm the existence of clear implications on cellular aging.

As a future perspective, we can observe that the information present on this project can be useful in order to generate solutions in aging processes or in different diseases that need the implications of stem cells. If the researchers work with FOXO factors and try to regulate or modify these factors, they can obtain a possible solutions or therapies in affections such as neurodegenerative diseases or muscular dystrophies among others.

We conclude with the discovering of a key element regulating stemness, it is affirmed that FOXO factors are critical drivers that characterizes the behaviour of stem cells. Importantly the project provides information for future perspectives involved in the discovery of possible solutions for aging and different diseases.

**Bibliography**


Appendix

This section contains the explanation of medium compositions and is present supplementary images and tables of the project. There are supplementary materials that cannot be presented in this section because they are very long excel tables.

Growth medium (GM) composition

HAM’S F10 (500 ml) that contains glucose and proteins, it is added 20% of fetal bovine serum (FBS) and 1% of penicillin / streptomycin (Antibiotics), the mix generates a growth medium that allows a correct growth and proliferation of cells. The increased concentration of serum allows proliferation.

Differentiation medium (DM) composition

DMEM (50 ml) that contains a high percentage of glucose, it is added 2% of horse serum and 1% penicillin / streptomycin (Antibiotics), the mix generates a differentiation medium that allows the differentiation of the cells due to the decreased concentration of serum.

Phosphate buffered saline (PBS)

Water salt solution that contains sodium phosphate and sodium chloride

Supplementary figure 1: Venn Diagram defining the list of genes that I use, the extended list that I use contains all the overlaps between two situations.
Supplementary table 1: 116 Transcription factors common to every type of stem cell in every specie

<table>
<thead>
<tr>
<th>Stem Cell Types</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESC, Germline, HSC, MSC, NSC, SSC</td>
<td>116</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transcription Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tcfcp21, INS1M, ESF2, Foxd3, EWSR1-FL1, YY1, ELK1, RORA_1, PLAG1, PPARG, RXRA, Esrrb, SP1B, RXR, RAR, RARB, RORA_2, REST, Nkx2-5, PPARG, TAL1::TCF3, NFIL3, HIF1A::ARNT, HNF4A, CTCF, Sox5, FOXF2, Foxq1, NR1H2::RXRA, Foxa2, AP1, RELA, ARID3A, Sox9, FEV, Stat3, MIZF, znf143, TP53, Gfi, Myb, SRF, STAT1, REL, HLF, FOXO3, Sox17, Amt::Anr, Gata1, GABPA, RUNX1, HNF1a, Egr1, Sp11, TBPA, TPaX1, Evi1, ZEB1, MEF2A, Mycn, ZNF354C, HOXA5, RFXRA::VDR, NFATC2, Nkx3-2, RREB1, ELK4, NFKB1, Spz1, HNF1B, Sox2, MYC::MAX, NFYA, NZF1-5-13, Nrf2, Pmz2, SP1, FOX11, CEBPA, Hand1::Tcfst2a, IRF1, PBX1, Pou5f1, NHLH1::IRF2, Klf4, USF1, Pax4, Lhx3, MAX, EBF1, Amt, MZF1-1-4, E2F1, NRI1, Zfx, Ddit3::Cebpa, Pax5, Pax6, Nkx3-1, NFIE2, Zfp423, TEAD1, Ar No box, NR2F1, FOXD1, NR3C1, Tal1::Gata1, NF-kappaB, FOXA1, ELF5, SRY, ESR1, Myc, CREB1, Nyc, TLX1::NFIC</td>
</tr>
</tbody>
</table>