uc3m | Universidad Carlos III de Madrid

University Degree in Biomedical Engineering
2017-2018

Bachelor Thesis

"Bioinformatic interpretation of microRNA role in three phenotypically related genodermatoses"

Andrea Santos Cortés

Tutor: Carlos León Canseco

Leganés, September 2018

ABSTRACT

The well-known regulatory function of microRNAs seems to play an important role in disease mechanism. Recent hypotheses have positioned them as a promising option for the study of genetic disorders. In this context, the microRNA profile of three phenotypically related rare genodermatoses, namely Recessive Dystrophic Epidermolysis Bullosa, Kindler Syndrome and Xeroderma Pigmentosum type C, is going to be analyzed taking as a reference healthy controls within the frame of network medicine concepts. Bioinformatics tools have proven to be essential to keep track of the alteration of the dysregulated microRNAs all over the organism. This thesis provides a direct correlation between the real and observable symptoms of the three conditions and the distortion at molecular level caused by pathogenic pathways. From the network analysis and functional enrichment analysis, a small selection of microRNAs and target genes are proposed as potential candidates for further research. Thus, with this project, it becomes evident the need of powerful predictive research tools previous to laboratory validation. In the considerably new field of microRNAs, the design and elaboration of more precise treatments as well as the discovery of biomarkers for early detection prevail among the possible different applications.

Keywords: Bioinformatics, network science, microRNAs, Recessive Dystrophic Epidermolysis Bullosa, Kindler Syndrome, Xeroderma Pigmentosum type C

ACKNOWLEDGEMENTS

A mi familia y amigos por quererme tanto y tan bien y ser un apoyo incondicional en todo lo que hago. Gracias porque no concibo una vida sin vosotros.

A Carlos por ser el mejor tutor que se puede pedir. Gracias por tu entusiasmo, tus consejos y tus ánimos. Me has enseñado mucho más de lo que te imaginas.

Y, con permiso de todos los anteriores, le dedico con especial cariño este Trabajo de Fin de Grado a la persona que más orgullosa se sentiría de verme terminar la carrera:

mi abuelo Moisés.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1. Medical background

The human body is constituted by thousands of cellular components that interact among them giving raise to specific functions. All these interactive pathways are collectively known as the human interactome [1].

Interactome network-based dependencies comprise undoubtedly a massive amount of data to be considered as complex systems. Network science arose at the beginning of this century to face the challenge of studying and analyzing whole interconnected systems, viewed as integrated networks. The spotlight is now on understanding the collective, holistic behavior instead of the reductionist analysis of the individual system's components. A firm knowledge of the assembly principles present in any network appears to be the key for a correct interpretation of such an intricate entanglement [2].

The recent revolution concerning computational analysis has motivated the emergence of network maps and bioinformatics tools have resulted crucial to deal with the processing of data obtained from them. In the big data actual scenario, driven by omic technologies in the biomedical field, computing algorithms have undoubtedly surpassed human abilities in terms of speed of analysis and storage capacity [2].

1.2. Network medicine

Medical research is focusing its efforts on the elucidation of genotype-phenotype relationship. Genetic disorders have been associated to abnormalities on particular genes. However, the link between the mutation and the resultant phenotype is not straightforward as the Beadle and Tatum's hypothesis "one gene/one enzyme" suggested. Even identical genotypes under the same environmental conditions may result in different phenotypes. Further investigations towards these lines would provide decisive clues to reveal the great enigmatic mechanism behind human diseases [3][4].

The complex interactome network may explain why the phenotypic signs of a disease are not uniquely the result of a damaged gene but the breakdown of the interrelation of that particular gene with all its surroundings molecules - direct interactors. A perturbation on the interactome implies that multiple processes functionally associated become altered. Consequently, diseases appear externally as a clinical frame of diverse pathological signs and symptoms usually influencing different systems of the organism [1].

Until now, the diagnosis of illnesses has been based on the hopefully fortunate recognition of a known classified group of clinical manifestations by physicians. The posterior treatment is then focused on palliating the symptoms. In this context, a new concept of medicine is emerging, the so-called network medicine, that draws the attention to the identification of disease modules**.** In the clinical practice, it supposes a complete twist of the screw, seeking for the root of the defect instead of the consequences [1][4].

Understanding a disease from a network-minded perspective implies to go back to the genetic origin and trace the biological routes that lead to the external perceptible symptoms.

1.3. Network principles applied to human disease

It might be thought that the first approximation to find the origin of a disease would be to identify the genes involved in the condition. It is not an easy task indeed if we consider that only 10% of the total human genes have a recognized disease-association [1]. Furthermore, even in the cases when the causal genes are known, there are intrincate and complex gene-phenotype relationships that must be understood in order to develop specific treatments for each disease.

In order to approach this problem, the *modus operandi* of network medicine consists on localizing neighborhoods of interconnected nodes, responsible for a determined cellular function in the organism, and whose rupture defines the phenotypic characteristics of a particular disease. Only with a general overview of the interactions that give answers to the observable phenotype, it is possible to localize the “wire” (cause) that is not working

when an uncommon external manifestation from the inner malfunctioning is perceived (effect) [1].

In network theory, nodes that connect to many others and interact functionally with them are referred as "hubs" (Figure 1.1). In this sense, hub proteins (those proteins that are highly connected to others) seem to play a fundamental role in the propagation of a disease. The reason is that a great number of connections implies a higher impact on multiple functional associations. In fact, the encoding genes of hub proteins are believed to correspond mainly to essential genes. Mutations in them would impede early development and provoke the spontaneous death of the embryo. However, in some cases, the mutation is tolerated, the embryo survives but shows more phenotypic changes compared to the model organism, resulting in a heritable disease [1][3].

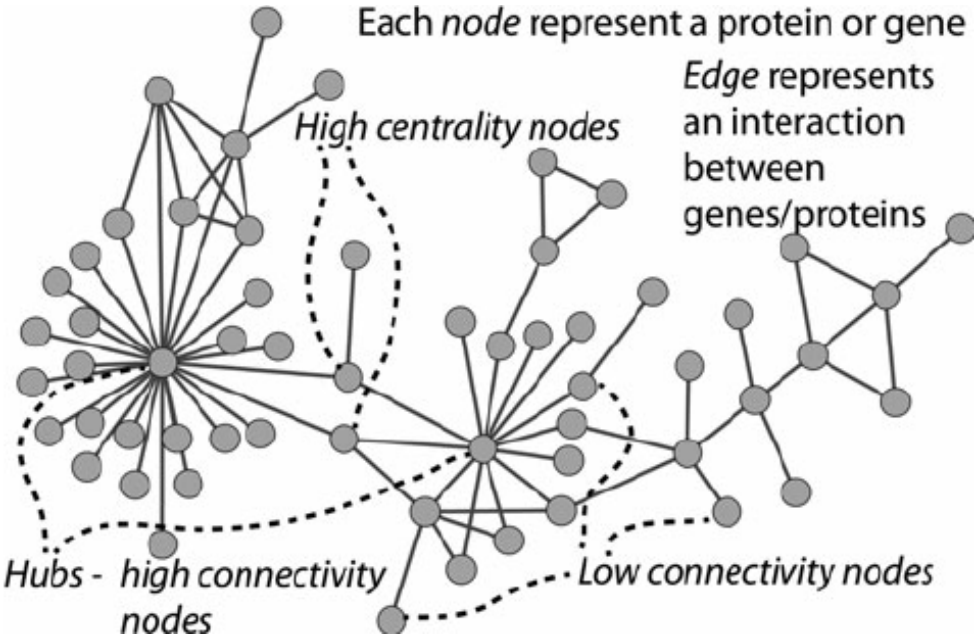


Figure 1.1. Illustration of network concepts. Source: [5]

Recent studies have proven that diseases with similar phenotypes share, to a greater or lesser extent, the connectivity patterns of their disease modules. As time passes by and research goes on, it has been assumed that diseases cannot be considered as independent from a molecular point of view. The overlap of disease modules explains comorbidity and phenotype similarity among certain conditions [1].

Global disease network maps have emerged to represent the dependency of diseases with known common molecular mechanisms or symptoms[3]. In the human diseasome,

two diseases are linked if they have in common at least one genetic component [6]. The establishment of such connections supposes a significant step in the search of genotype-phenotype associations. In fact, it has been proven that not only does a correlation exist between disorders with similar symptoms and common genes but also the proteins encoded on those genes tend to interact more among them [4][7]. Therefore, as a direct consequence, a great majority of the same pathways will be altered in these cases.

1.4. Genodermatoses under study

In human beings, genetic information is contained in 23 pairs of chromosomes. The last pair is the one that determines the sex of the individual. The rest of pairs receive the name of autosomes. Genes are small fragments that occupy a determined location within each chromosome. Genetic disorders arise from mutations on specific genes that are inherited from parents to children. Depending on the inheritance pattern, inherited diseases are classified as dominant, if the mutation only affects one copy of the pair of chromosomes or recessive, if both copies present the gene mutated. Besides that, a disease is considered rare when is suffered by less than 5 people out of every 10,000 inhabitants. There are estimated to be around 7,000 rare diseases affecting the 7% of the world population. According to Federación Española de Enfermedades Raras (FEDER), the number of people affected in Spain exceeds 3 millions [8].

Herein, three rare autosomal recessive diseases are presented. The reason why they have been grouped for this study is that, despite their different genetic background, they share similar phenotypic characteristics; among them, it is worth mentioning a strong tendency to develop cell carcinoma at the squamous cells of the epidermis.

1.4.1. Recessive Dystrophic Epidermolysis Bullosa (RDEB)

Epidermolysis Bullosa (EB) is an inherited disorder mainly characterized by the appearance of a blistering pattern in the skin. Given its genetic origin, EB is caused by mutations on specific genes that impede the correct expression of groups of proteins responsible for the adherence of epidermis to dermis (Figure 1.2) [9].

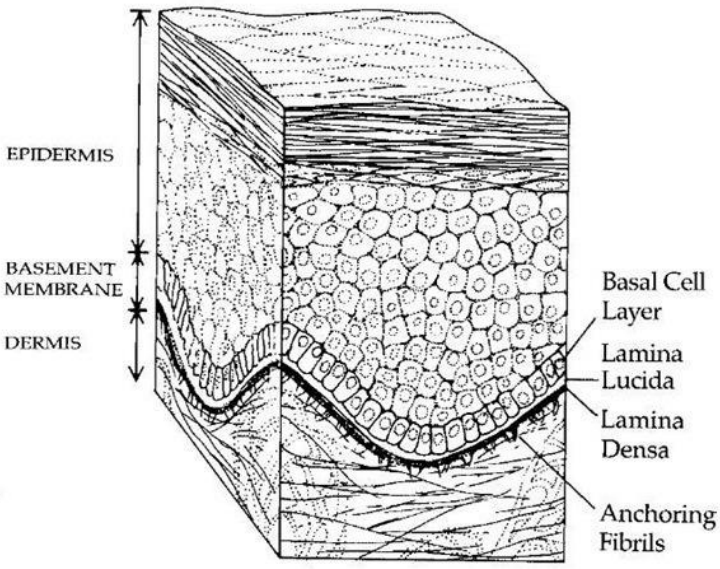


Figure 1.2. Representation of skin layers. Source. [10]

Attending to the type of proteins lacking, EB is classified in four main types (Table 1.1):

TABLE 1.1. TYPES OF EPIDERMOLYSIS BULLOSA. Source:[11] **.**

EB type	Lacking proteins	Level of detachment
EB simplex (EBS)	Keratins 5 and 14 Plectin	In the epidermis
Junctional EB (JEB)	Laminin 5 (now 332) Collagen XVII Integrin $\alpha_6\beta_4$	In the lamina lucida
Dystrophic EB (DEB)	Collagen type VII	In the sublamina densa
Kindler Syndrome (KS)	Kindlin 1	Any layer

For the diagnosis, immunofluorescence mapping by antigen staining is a technique widely used to classify the type of EB from a skin biopsy of the blistering area (Figure 1.3) [12]. The last known figures concerning EB estimate that there are between 50.000 and 80.000 cases in Europe and around 2.000 in Spain [13]. In order to have an official approximation of the prevalence of DEB, it is necessary to go back to the conference of Spain RDR in December 2014. By that date, a 70% of the 69 patients included in the EB Registry in Spain had been diagnosed with the dystrophic type of the disease [14].

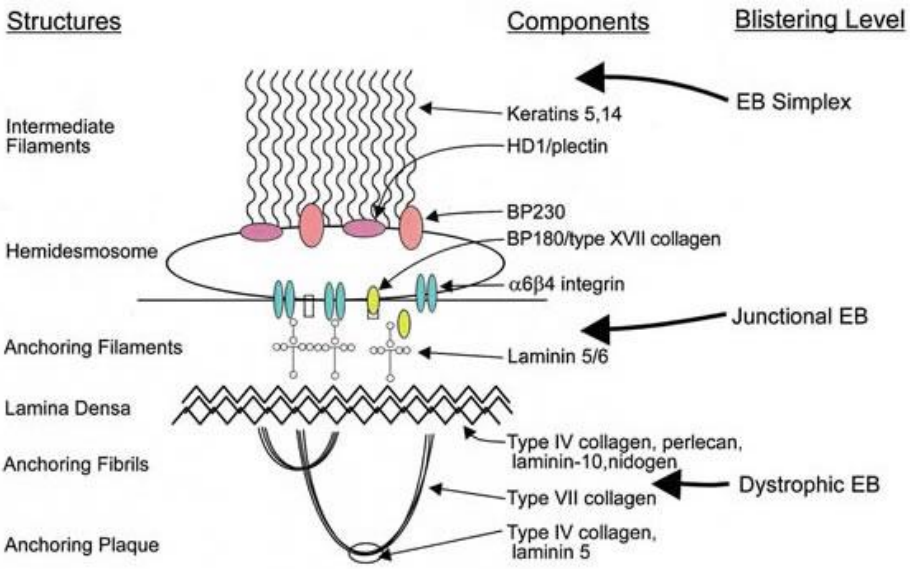


Figure 1.3. Scheme of the level of damage of the different types of EB. Source:[15]

Two patterns of inheritance can be distinguished in the Dystrophic EB: dominant (DDEB) and recessive (RDEB). RDEB is associated to a mutation on the gene COL7A1, responsible for the production of collagen type 7. This protein serves as the major constituent of the anchoring fibrils that joins the sublamina densa of the basement membrane to the dermis [11][12].

People suffering from RDEB show an increased fragility of the skin and suffer from complications in the mucous membranes, especially affecting the eyelids or oral cavity as well as the respiratory, esophageal and intestinal tracts. In the most severe cases, the disease may lead to an early death (before the age of 30) usually due to the outbreak of renal problems or the high risk to develop carcinoma [12][16].

1.4.2. Kindler Syndrome (KS)

Kindler Syndrome (KS) is the most singular and unusual type of Epidermolysis Bullosa. In the same way as the previous mentioned RDEB, patients with KS also present skin blisters. It is characterized by the loss of the structural protein Kindlin-1, caused by the mutated form of the FERMT1 encoding gene [17]. The distribution statistics presented at the conference of Spain RDR in December 2014 stated a 1% of KS in the 69 registered patients with EB in Spain [14].

From its definition as syndrome by Theresa Kindler in 1954, numerous papers and articles have been published reporting cases of people affected all over the world [18][19][20]. Affected individuals coincide in a high photosensitivity and fragility of

the same damaged tissues in RDEB. Proof of this is that some publications refer to the skin of these patients as cigarette paper- like skin. The symptom picture is quite similar to the previous genodermatosis described.

However, the aggressive phenotype that the disease presents cannot be uniquely explained by a genetic mutation, especially the clinical features not related to direct skin damage such as photoaging or the higher propensity to develop cancer. In a recent study, Zapatero-Solana et al., suggested oxidative stress to be a major factor in the general derangement of the organism in KS. Loss of function in kindlin-1 seems to disrupt several signal transduction pathways in which integrins and focal adhesions take part [21].

A more abundant generation of reactive oxygen species (ROS) imbalance the redox status and harm the genetic material. As a defense, the mechanisms responsible for the detoxification are activated to produce antioxidant molecules, such as glutathione, in an attempt to recover the equilibrium. Keratinocytes from KS patients show aberrant-shaped mitochondrias and a clear pro-oxidant state: a greater ratio between oxidized and reduced glutathione (GSSG/GSH), the downregulated expression levels of gamma-glutamyl cysteine ligase (GCLC and GCLM) which accelerates the synthesis of GSH and an increased amount of malondialdehyde (MDA), a known degradation product from ROS reactions (Figure 1.4) [21].

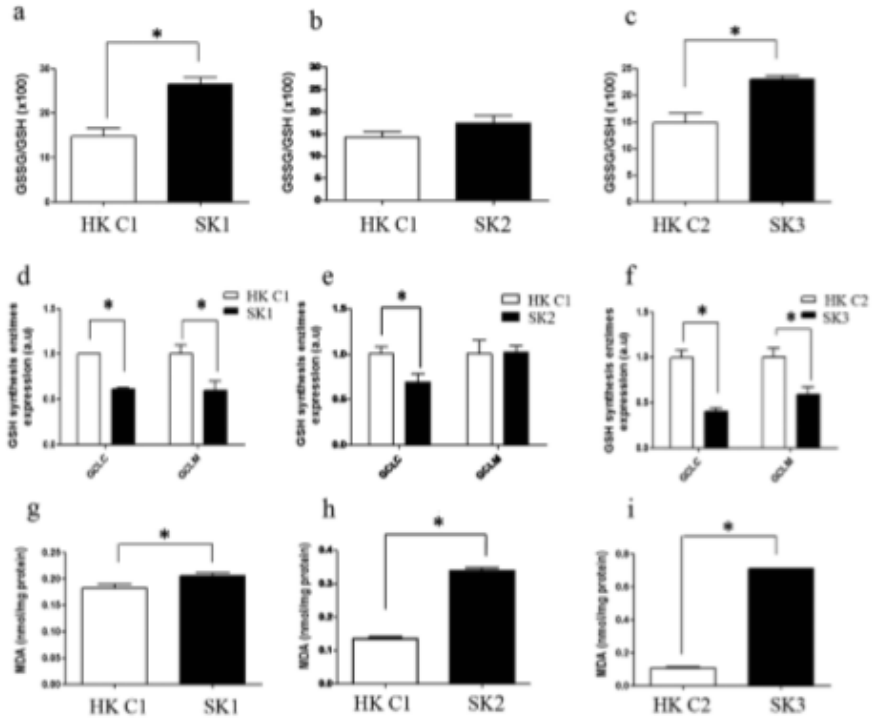


Figure 1.4. Results obtained in an experiment with KS keratinocytes.
Source: [21].

1.4.3. Xeroderma Pigmentosum complementation group C (XPC)

The malignancy of Xeroderma Pigmentosum (XP) disorder is attributed to mutations on eight genes related to the nucleotide excision repair (NER) pathway that serves as a redress mechanism for the damage caused by UV light on DNA (Figure 1.5). The different types of the disease are named according to the gene affected. In the case of XPC, the gene mutated is XPC involved in the recognition of photoproducts in DNA [22]. The prevalence of XP in Europe have been established in 1 person per 250.000 inhabitants [23].

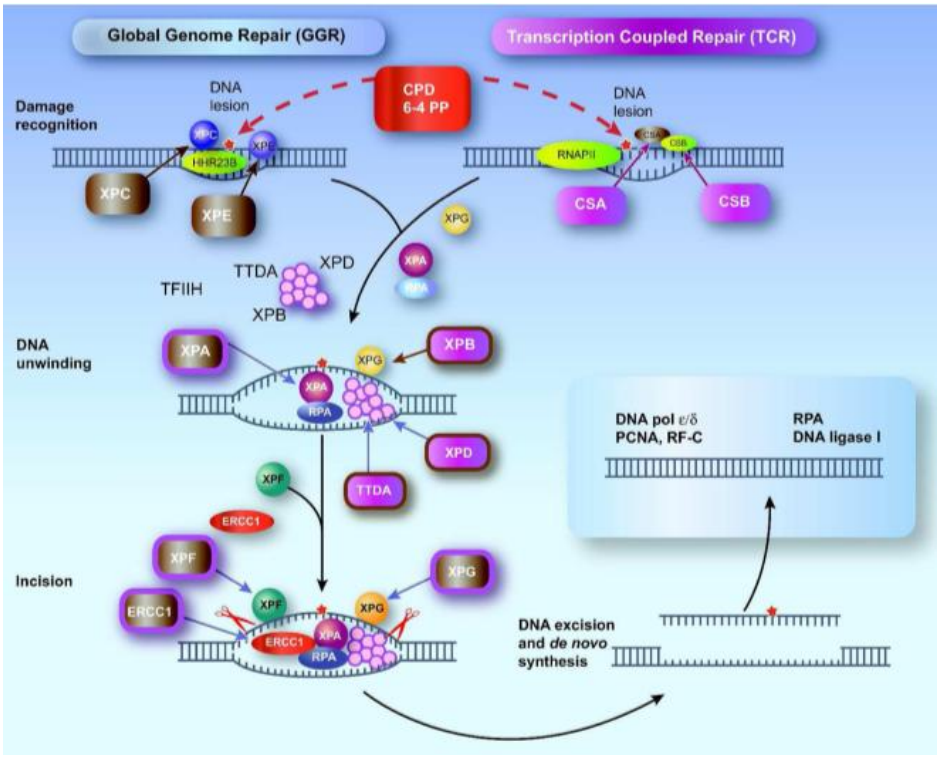


Figure 1.5. Scheme of the NER pathway. Source: [24].

In more than half of the cases, even a brief exposure to sunlight causes severe sunburns, giving rise to a characteristic pattern on the skin with different level-of-pigment patches and blisters. The remaining cases progressively develop anomalous excessive freckle-like pigmented lesions (lentiginosis) on sun-exposed areas. Only a quarter of the diagnosed XP patients in US incur in neurological complications that seriously aggravate their quality of life. Symptoms usually appear since the early infancy and, as the other genodermatoses, XP patients have an elevated tendency to develop carcinogenesis [24]. In fact, according to FEDER, XP patients have a 1,000 times increased risk of suffering skin cancer than the rest of the population [25].

1.5. MicroRNAs

On the basis of the above, the new "think globally – act locally" perspective to handle with the cure of a disease is being directed towards comprehending in deep the mechanisms of diseases so that a satisfactory treatment can be implemented. Nevertheless, as drugs need to be targeted to a specific point, key intermediary-pathways players must be accurately localized. In this context, microRNAs have emerged as a promising option.

The recently discovered microRNAs (miRNAs) molecules have groundbreakingly adopted a privileged place in biomedical research. These small non-coding RNA (ncRNA) molecules of 18-25 nucleotides modulate gene expression post-transcriptionally and are present in almost every metabolic pathway in the organism. Due to their regulatory activity, they have attracted considerable interest in the study of diseases. They might be the triggering agents for the disease cascade in the human interactome, and thus, the ultimate responsible for the pathogenesis of diseases [26][27].

Although this thesis will be focused entirely in miRNAs, the reader should also know that there exists a different variety of ncRNA molecules including: transfer RNA, nucleolar RNA, nuclear RNA, phage and viral RNA, small interfering RNA (siRNA) or PIWI-interacting RNAs (piRNAs) among others [28].

1.5.1. Discovery of miRNAs

In 1993, *lin-4* was identified at Dartmouth Medical School as the first miRNA in *Caenorhabditis elegans*, a nematode species. The original RNA was not translated into proteins as expected but instead it was divided in small transcripts [26][29].

The very same year another research group from Harvard Medical School shed light on the functionality of these short fragments. They confirmed the antisense base-pairing between the *lin-4* miRNA and *lin-14* messenger RNA (mRNA). This mRNA, declared afterwards the first target gene of a miRNA, was surprisingly repressed due to the complementary binding [26][29].

However, the spotlight started to be focused on this area when in 2000, the second miRNA discovered, *let-7*, was proven to have been perfectly conserved throughout evolution during hundreds of millions of years. Since this moment, there was a spectacular progress in the search of new miRNAs [26][29].

1.5.2. Biogenesis of miRNAs

The central dogma of biology states the directionality of genetic information from DNA material (Figure 1.6). However, it is being questioned nowadays due to the non-coding RNAs (ncRNAs) molecules. The predecessor DNA fragments of ncRNAs undergo transcription but do not fulfill the translation step. Therefore, the final function of these types of RNAs is not structural (to conform a protein) but regulatory [30].

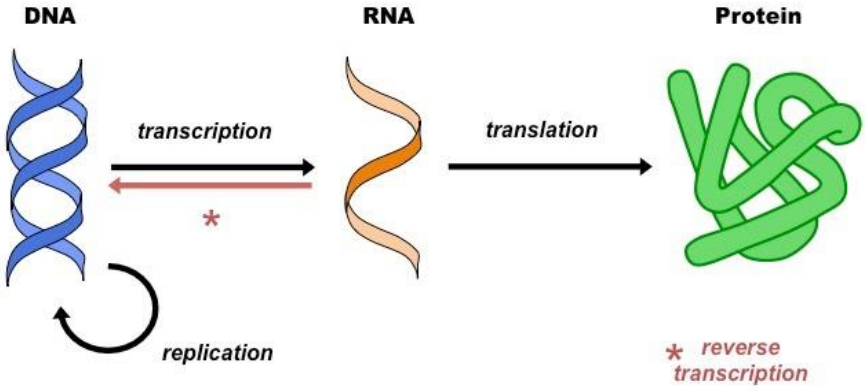


Figure 1.6. Representation of central dogma of biology. Messenger RNA is created after transcription, which is then translated to proteins. Source:[30]

The exact mechanism of action of miRNAs is not yet fully understood and many aspects remain still unclear [7]. Nevertheless, until now, it is believed to happen as follows (Figure 1.7) [28][31][32]:

1. Transcription of the genome generating a long precursor RNA molecule, primary miRNA (pri-miRNA) and normally mediated by the enzyme RNA polymerase II.

   This pri-mirna is characterized by a hairpin, a cap and a poly A tail.
2. Recognition of pri-miRNA structure by the nuclear RNAse III enzyme Drosha and its cofactor DGCR8/Pasha (all of them form a complex known as Microprocessor) with the subsequent cleavage of pri-miRNA that gives rise to the precursor miRNA (pre-miRNA) of 70 nucleotides.

3. Translocation of pre-miRNA from nucleus to cytoplasm via Exportin-5
4. Processing (second cleavage) by another RNase III enzyme, Dicer, to form mature miRNA duplexes (double-stranded miRNA/miRNA*) of approximately 22 nucleotide length
5. Loading of the mature strand of the miRNA duplex into the RNA-induced silencing complex (RISC) facilitated by Argonaute proteins. Finally, they jointly target specific mRNAs, mainly at the 3' untranslated (UTR) region, to produce translational repression or mRNA degradation, depending on the level of complementarity between bases.

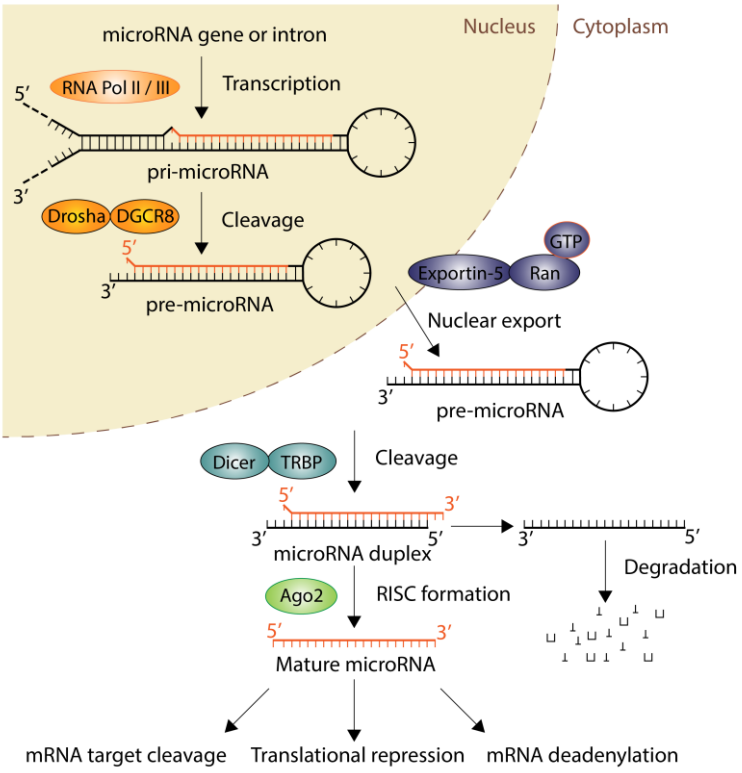


Figure 1.7. Scheme of biogenesis of miRNAs. Source: [61]

1.5.3. Role of miRNA in diseases

The mechanism of action of miRNAs remains still in question. At the beginning, it was said that miRNAs were involved in the post-transcriptional regulation of mRNAs at the cytoplasm. However, recent studies open up the possibility of their implication at the nuclear level leading to alternative splicing [31].

Be that as it may, there is strong evidence that dysregulation, either up or downregulation, of some miRNAs is behind human malignancies, and in particular,

behind cancer development. Either by overexpression, repression or deletion of the RNA targets, tumorigenesis is caused in one or other way [33].

Although, the identification of miRNA-mRNA target interactions is still going on, there is great hope in employing miRNAs in the theranostics field. They have been proposed as potential drug targets to reverse gene expression of diseases, or even, synthetic miRNAs that could replace the malfunctioning ones [26].

Along this thesis, the miRNA-RNA interaction network modules overlapping in three genodermatoses is going to be analyzed.

2. HYPOTHESES AND GOALS

The similarity in certain phenotypical characteristics of Recessive Dystrophic Epidermolysis Bullosa (RDEB), Kindler Syndrome (KS) and Xeroderma Pigmentosum type C (XPC) raise the curiosity of finding shared altered regulatory mechanisms. Network medicine has appeared as a tool to relate the biological compounds in disease modules with successful targeted treatments. The ambitious objective of finding the cure to a genetic disease must accept the fact that mutations themselves do not give a complete explanation to all the phenotypical features observed.

Until the present day, no publications have been issued covering the phenotypical resemblance of these three diseases. Subsequently, any research towards this direction enters a completely unexplored area. However, given the overlap in symptomatology, there are huge expectations on the discoveries that could be made from a molecular-level network analysis.

A pioneering study [34] carried out by UC3M Bioengineering Department along with Instituto de Investigación Sanitaria de la Fundación Jiménez Díaz (IISFJD), Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER) have provided source data to start with this thesis. The RNA sequencing (RNA-Seq) analysis performed on fibroblasts from RDEB, KS, XPC patients and healthy donors, provided raw data of miRNA to be analyzed. Fortunately, remarkable conclusions were achieved at RNA level: the three disorders shared 227 transcripts and 42 signaling circuits altered. Certainly, once common molecular pathways were proven to exist, a promising horizon in miRNA analysis started to be glimpsed. The starting point in this thesis is, therefore, the treatment of the available miRNA data to throw light on disease common mechanisms, and with the following goals in mind:

- Extract the common dysregulated microRNAs to the three conditions taking as reference healthy controls.

- Find enriched functions in which the miRNAs of the previous point are implicated.
- Study the RNA target molecules through an interaction network.
- Select miRNAs candidates potentially useful as therapeutic targets.

3. MATERIALS AND METHODS

3.1. Data collection

Next generation Sequencing (NGS) is related to all the novel high-throughput platforms that are able to sequence nucleotide chains (both DNA and RNA). Nowadays, for RNA profiling, the most widely–used technique is RNA-Sequencing (RNA-Seq), a NGS method that is able to detect both coding and non-coding RNAs. Although there are three main NGS platforms that perform the sequencing step, Illumina HiSeq, Ion Torrent and SOLiD system, the general protocol is the same for all three. First, RNA is isolated from the tissue of interest and a quality control is carried out to correct possible defects from the signal acquisition procedure. Then, RNA samples are fragmented in small pieces and they undergo reverse transcription to form double-stranded complementary DNA (cDNA). During this library preparation step, adapters are also added to be ligated to both ends of the fragments and PCR amplification is implemented. Depending on the platform used, the characteristics of the adapters vary as well as the mode of reading the sequence. Anyway, the number of reads that aligns to a transcript (counts) is returned as output, and in RNA-Seq is then related to the amount of expression (quantity of reads) of the RNA. A statistical analysis is usually made afterwards to provide further information such as the level of expression compared to control samples or the p-value for evaluating the significance of the results [35].

Our bioinformatic analysis started with the miRNA counts coming from fibroblasts of eighteen subjects: nine RDEB patients, three KS patients and three XPC patients, taking as reference three healthy donors [34].

3.2. Data treatment

The real challenge of omic technologies is to know how to interpret the massive amount of information obtained from high-throughput techniques. An exhaustive bioinformatic data processing is necessary to translate numbers, variables and parameters to an understandable biological meaning. Furthermore, the validity of any research findings is closely related with the way the analysis of the information is geared. The action plan for this thesis has been carefully designed to arrive to conclusions that can guide future lines of research.

3.2.1. Understanding source data

For each one of the genodermatoses, raw data consisted on the list of miRNAs expressed (almost a thousand) along with information relative to its comparison with the control group:

- **miRNA name**: it corresponds to the identifier given by *miRBase* [36], at the moment of its incorporation to the database. More details are given hereunder.
- **logFC:** means log fold-change and gives a clue about how much absolute change exists in a specific variable between two measurements. In this case, the fold change is calculated by dividing the average read counts of the disease condition by the average read counts of the healthy control. Then, the sign tells us if there is overexpression (when positive) or underexpression (when negative).
- **logCPM:** stands for log counts per million and it also provides information to compare expression levels between groups.
- **p-value**: expresses the level of certainty of an existenting difference. By convention, a result is considered statistically significant when the p-value is lower than 0.05.
- **FDR:** stands for false discovery rate. The FDR approach serves to reduce the number of false positives**.** It can be used as a corrected p-value that takes into consideration the great number of variables with respect to the number of samples analyzed. FDR is going to be used in this analysis given the high amount of miRNAs in comparison with the small number of subjects.

miRBase (Figure 3.1) [36] is the world reference database where all miRNAs are registered. Not only the sequence but also annotations associated to each molecule can be found.

Figure 3.1. *miRBase* [36] interface.

miRBase [36] has established the official nomenclature to be followed for the introduction of any miRNA in the repository. The format consists on these four elements separated by hyphens:

- Three initial letters that refer to the organism. In our case, these will always be "hsa" as we are going to focus just on the Homo Sapiens species.
- Three following letters, which can be either "mir" regarding the miRNA gene or "miR" for the mature miRNA and in some exceptional cases, "let" and "lin" due to historical reasons.
- A number that is assigned sequentially according to the order of discovery.
- Posterior lettered suffixes are added in some cases to distinguish similar sequences with different precursors or from opposite arms.

The deposition of new miRNAs in *miRBase* [36] has grown exponentially in the last years, achieving a total of 38589 entries on March, 2018 [36].

3.2.2. Comparison of data groups

Given the volume of data, it is essential to narrow the field of possibilities just to the most relevant ones ignoring non-meaningful data. Following these guidelines, the first obvious step is to know which significant miRNAs are shared by the three genodermatoses under study. To accomplish this task, *Venny* (Figure 3.2) [37] has consolidated as an exceptional option.

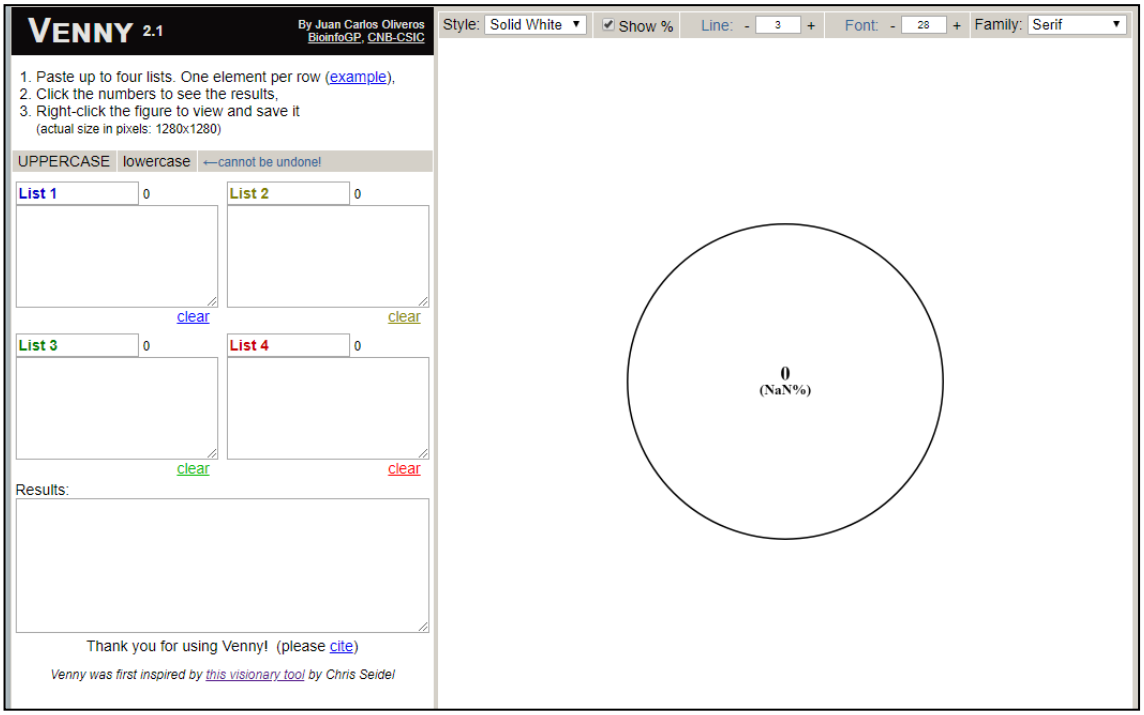


Figure 3.2. *Venny* [37] interface.

This program offers a clear graphical comparison of different group categories. It receives as input a maximum of four lists of elements. As a result, a figure with overlapping circular diagrams is drawn on the screen, commonly known as Venn diagram, in which each circle represents one list. Percentages at the intersections can be shown to assess the level of similarity between circles. Attributes such as style, format or color can be changed depending on one's preferences.

3.2.3. Analysis of interactions

The study of the interactions of miRNAs results crucial to understand their role in the organism. The interface of *miRNet* (Figure 3.3) [38] platform enables the integration of data in networks to visualize the relationships among known-sequence molecules and its targets.

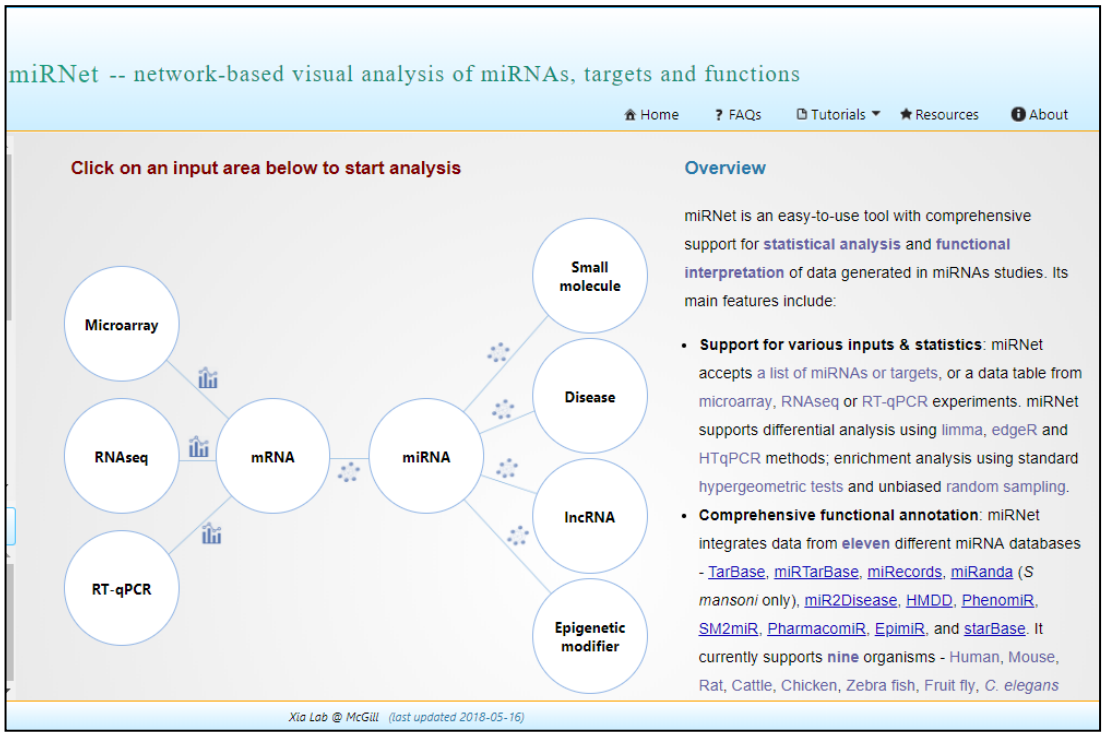


Figure 3.3. *miRNet* [38] interface.

A network can first inspire respect because of its complexity. However, it is important to keep in mind the simple definition of networks: they are just nodes connected by links. Taking advantage of the defined structure any grand network is based on, there are a set of useful tools that allow to explore them and find its most influential elements:

- The **degree of a node** will indicate which nodes have a greater number of edges or, in other words, the number of connections to other nodes. Hubs (nodes with a high degree) seem to be key participants in diffusion processes. Thus, they are also considered important in this study where diseases are seen as spreading events affecting a variety of functions in the subject.
- The **betweeness of a node** has a lot to do with connectedness. Those nodes that are present in most of the connecting paths between two nodes, will show a high value for betweeness. If these interconnective nodes are somehow altered, the connections between modules in the network will fail or at least not be produced in the same way as in normal conditions. Betweeness should also be an important parameter of study.
- **Clustering** could also attract attention in the analysis of diseases. It is not an individual property of a node but instead it looks at the neighbors of the node. Clusters could explain related mechanisms (modules) to be altered in a particular disease.

The *miRNet* [38] program works in three main steps:

i. **Data upload and processing**: It gives the possibility of entering diverse type of data from lists of miRNAs, transcripts or small molecules to tables from RNA-Seq or microarrays.

ii. **Network creation and refinement**: An interaction network is generated. By default, input molecules appear as red circular nodes and target molecules as light blue squares, all of them linked by white edges over a black background, although sizes and colors can be easily modified. Several functionalities are available to transform huge networks in simpler ones depending on the aim of the research. Deletion of nodes/edges or filtering by degree, betweeness or shortest path are some of the options that help in the controlled reduction of the network.

iii. **Network analysis**: *miRNet* [38] also has a functional analysis panel to perform a functional enrichment analysis (see below). This allows a further in-deep analysis of the biological pathways associated to the network selected. Different databases such as Gene Ontology (GO) [39], Kyoto Encyclopedia of Genes and

Genomes (KEGG) [40] or REACTOME [41] are consulted to seek for the functions affected.

3.2.4. Functional enrichment analysis

It is widely believed that genetic disorders are intrinsically linked to an overexpression or underexpression of certain genes in comparison with normal controls. These are called differentially expressed (DE) genes and stand as highly probable candidates of being involved in the disease (either its generation or the phenotypical outcome).

In the same way, miRNAs will be considered determinant for the disease, if they are differentially expressed (with respect to healthy controls). The relationship between DE miRNAs and DE genes arise a curious phenomena that could help us in the prioritization of miRNAs. The concept is that if the miRNA is overexpressed, and considering that they usually negatively regulate the expression of coding RNA, it would mean higher inhibition of the genes than in normal conditions. Therefore, regulated genes in this case may appear underexpressed. And the opposite may happen if the miRNA is underexpressed.

However, more than just knowing the nomenclature of the genes, the real interest ultimately lies on discovering the functional role of the genes affected. Once again, bioinformatics facilitates this task by making an automatic search of annotations (functions or pathways) related to a particular gene and enclosing under the same concept genes that share similar annotations. A functional enrichment analysis therefore provides those functions or pathways that are statistically relevant (and consequently have a small p-value) in a list of genes, that can come from a list of miRNA targets. These functions can be established in terms of the place where the genes are active (cellular component), the activity or specific function in which the genes are implicated (molecular function) or the biological pathway to which the genes contribute (biological process), and are annotated in the Gene Ontology (GO) Database [39], or others such as KEGG [40] or Reactome [41].

From all the enriched functions, the interest focuses normally on those that include a greater number of DE genes, or those with smaller p-value, due to the again difficult task of extracting meaningful information from large lists of data. *REViGO* (Figure 3.4) [42] defines itself as the tool to summarize an extended enumeration of Gene Ontology (GO) terms. When a great amount of hits are returned, it is quite useful to just rescue the

most significant outcomes and ignore the rest, or group them together into relevant groups. GO terms are approved wording expressions to design specific features occurring in biological systems. *REViGO* [42] analyzes a list of GO terms with assigned p-values and gives as a result the highlighted pathways or elements for each category: molecular function, cellular component and biological process.

Figure 3.4. *REViGO* [42] interface.

3.3. Data selection

After all the steps, the analysis should conclude with the most remarkable miRNAs. The idea is to make a selection of a small group of miRNAs that could serve as biomarkers for diagnosis or targets for treatment. Hopefully, it is possible to delimit the initial 3.000 miRNAs field to just some miRNAs that exhibit evidence or at least clear objective clues to be partially responsible of the common disease mechanism hidden behind RDEB, KS and XP.

4. RESULTS

4.1. Analysis of the dysregulated microRNAs in diseases vs. controls

As it has been exposed above, the expressed miRNAs from RDEB, KS and XP patients had been identified by the RNA-Seq experiment. The data consisted on the list of miRNAs names in Excel format with the statistical parameters of logFC, logCPM, p-value and FDR. Healthy controls were taken as reference.

The original list gave a total of 1252/964/891 miRNA outcomes in RDEB vs. Healthy, KS vs. Healthy and XPC vs. Healthy respectively and was necessarily reduced by discarding miRNAs with a FDR higher than 0.05. The new list had 27/99/148 components (Annex), figures quite more reasonable to work with. Table 4.1 summarizes this reduction.

TABLE 4.1. LISTS OF MIRNAS

	#miRNAs RDEB vs. Healthy	**#miRNAs KS vs. Healthy**	**#miRNAs XPC vs. Healthy**
Total miRNAs detected	1252	964	891
Significant miRNAs (FDR<0.05)	27	99	148

These three short lists were first introduced in *Venny* [37] to obtain common and specific miRNAs for each condition (versus healthy controls). The results obtained can be seen in Figure 4.1. A considerable overlap among the three can be appreciated with a total of 18 significant DE miRNAs in common.

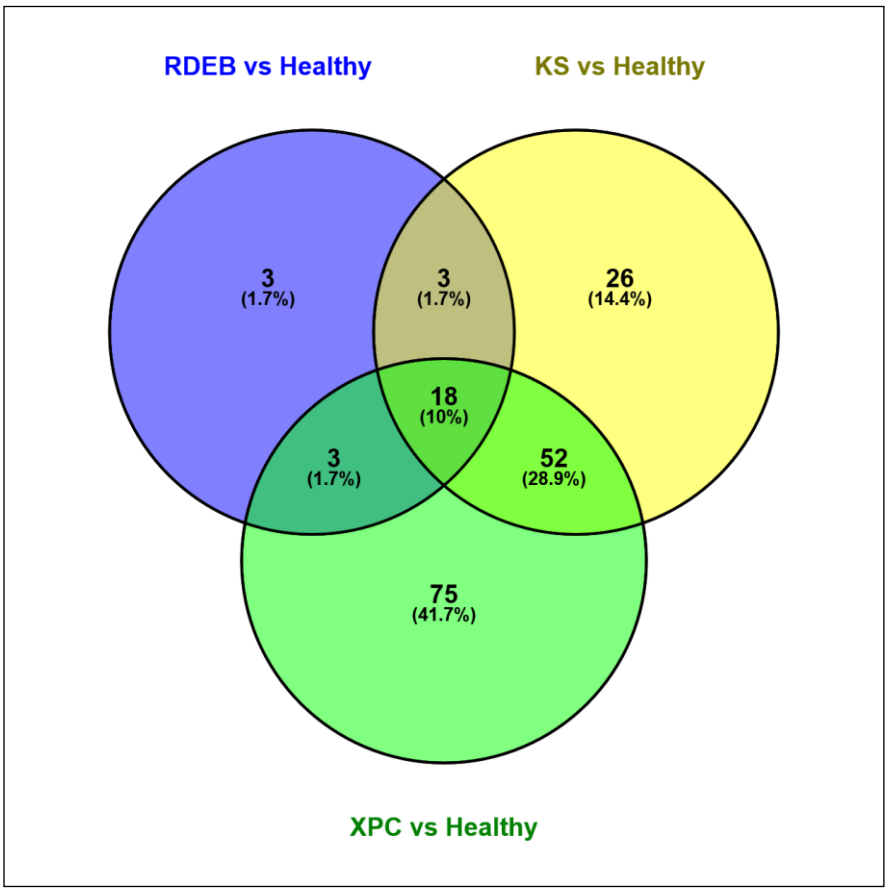


Figure 4.1. Venn diagram of the significant miRNAs.

The hypothesis at this point is the following: shared miRNAs might justify the similarities between the diseases, while specific miRNAs in each condition would account for the differences. In this thesis, the analysis has been made primarily with the miRNAs shared by the three conditions, in order to try to find an explanation to the phenotypical resemblance.

The eighteen miRNA names are listed in Table 4.2, with an arrow next to it indicating the change of expression compared with control. The pointing-up arrow represents overexpression and in the opposite direction, it means underexpression. Surprisingly, it was possible to establish an up or down arrow for each miRNA because in the three comparisons the direction of change in expression is the same.

TABLE 4.2. LEVEL OF EXPRESSION OF THE SHARED DYSREGULATED MIRNAS.

miRNAs	logFC(RDEB vs Healthy)	logFC(KS vs Healthy)	logFC(XPC vs Healthy)
hsa-miR-10a-5p (↑)	8.82	10.50	10.44
hsa-miR-10a-3p (↑)	8.59	9.54	9.48
hsa-miR-556-5p (↓)	-4.54	-6.28	-4.17
hsa-miR-6507-5p(↓)	-4.78	-5.37	-6.47
hsa-miR-6842-3p(↓)	-2.42	-2.77	-3.31
hsa-miR-195-3p (↑)	2.90	2.01	2.36
hsa-miR-129-5p (↑)	3.60	3.58	4.00
hsa-miR-615-3p (↑)	2.97	3.24	3.64
hsa-miR-129b-2-3p (↑)	-2.15	-2.10	-1.71
hsa-miR-29b-2-5p (↓)	4.38	5.14	5.15
hsa-miR-10b-5p (↑)	4.18	4.93	4.99
hsa-miR-10b-3p (↑)	-3.01	-4.68	-5.67
hsa-miR-1295a (↓)	3.92	3.25	3.47
hsa-miR-148a-5p (↑)	2.11	2.01	1.53
hsa-miR-195-5p (↑)	2.55	2.93	3.15
hsa-miR-629-3p (↑)	1.97	1.89	1.51
hsa-miR-148a-3p (↑)	2.07	2.28	2.19
hsa-miR-1468-5p(↓)	-2.01	-3.59	-3.25

Furthermore, it is worth mentioning that the absolute difference of the change almost coincides numerically in the three. The scatter plot of Figure 4.2 represents in Cartesian coordinates the relationship among the log FCs. Each axis corresponds to the data from one genodermatosis vs. control. The diagram shows a clear positive correlation for each pair of diseases. The pattern of blue dots approximately adjusts to the identity line so it can be confirmed that the three genodermatoses change following the same trend in growth.

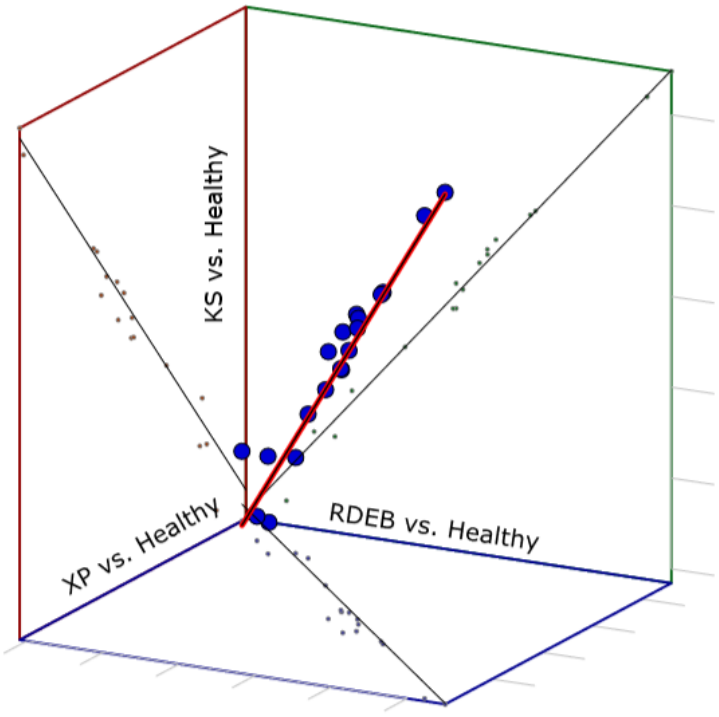


Figure 4.2. Scatter plot of the logFC of the 18 shared miRNAs.

4.2. Analysis of the generated regulatory networks

To really give sense to the arrows, the effect of the alteration must be verified by looking at the target genes that are being regulated by the 18 miRNAs. The interaction network was then implemented with the aid of *miRNet* [38].

The list with the eighteen elements was uploaded choosing the corresponding fields: as organism, H. Sapiens (human); as ID type, *miRBase* ID and as target type, genes. This is because the purpose is to investigate the genes regulated by the differentially expressed miRNAs and the functions of the organism that are consequently affected. One of the main problems faced when using databases is their incompleteness. Even more in the case of rare diseases, it is common not to find related terms. That was exactly the reason why the tissue field in *miRNet* [38] had to be left blank. When "skin" was selected, a reduction to almost null results was obtained due to the fact that only 2 of the 18 miRNAs are reported so far in the *miRNet* [38] database for the skin category. Nevertheless, it was fortunate not to have closed the search to any particular tissue because interesting target genes present in other tissues may have not appeared.

The network builder returned as overview details: 18 queries, 3295 nodes (miRNAs: 18, Targets: 3277) and 4200 edges. In the next step, it was possible to visualize the network

(Figure 4.3), a tangle of interconnected genes and miRNA, a scenario not especially accessible to deal with. However, predicting this, the program offers filter options to transform the network in an easy-to-use format to work with.

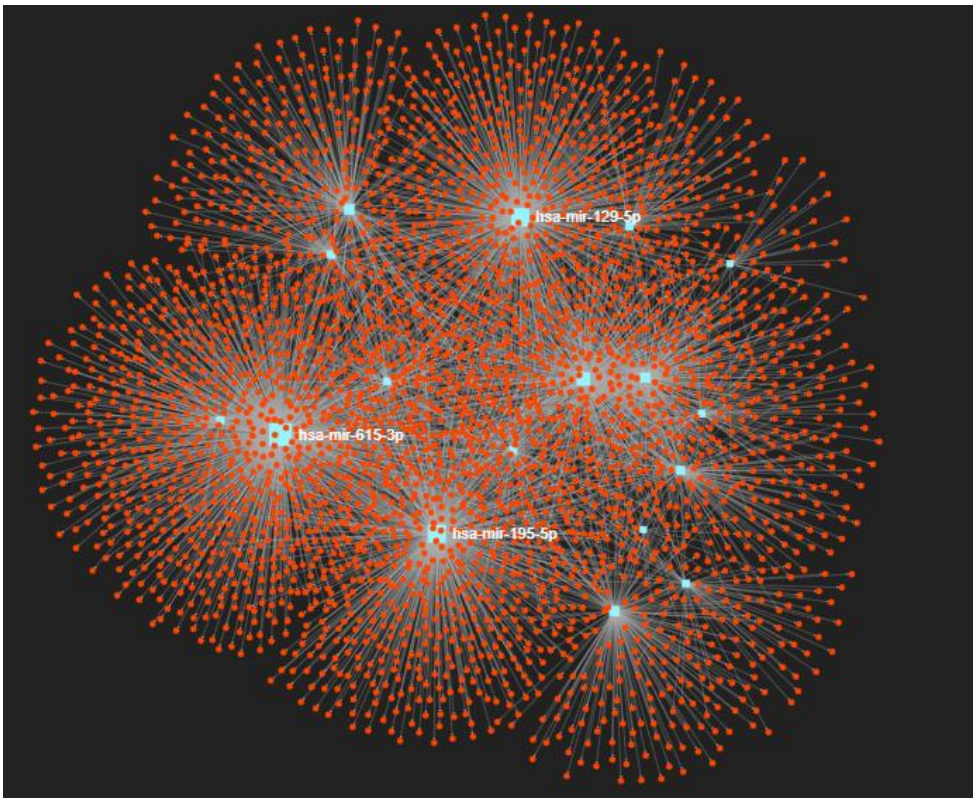


Figure 4.3. Interaction network built from the eighteen predominant miRNA

The most important information that can be extracted from this network, in terms of degree of the nodes, is summarized in Table 4.3 and 4.4. The top-5 miRNAs (in terms of degree, Table 4.3) announce to be the main players in the network. They are regulating a great amount of transcripts so there is more probability for them to be involved in the disease cascade. On other side, the genes being regulated by a greater number of altered miRNAs (Table 4.4) might also attract our interest. Even if their regulating miRNAs show slight changes of expression, together may converge in a substantial difference in the target gene.

TABLE 4.3. MIRNAS WITH THE HIGHEST DEGREE IN THE GLOBAL NETWORK.

Pos	miRNAs	Degree
1	hsa-mir-615-3p	893
2	hsa-mir-195-5p	640
3	hsa-mir-129-5p	625
4	hsa-mir-10a-5p	463
5	hsa-mir-10b-5p	323

TABLE 4.4. GENES WITH THE HIGHEST DEGREE IN THE GLOBAL NETWORK.

Pos	Genes	Degree
1	MTX3	5
2	NR2C2	5
3	CDK6	5
4	ARL6IP1	5
5	NPTX1	5

4.2.1. Application of degree filter

The degree of a node is defined as the number of edges it has to connect with other nodes. A hub is a node with a great number of connections compared to the rest. As it has been previously mentioned, hubs are liable to be related with the alteration of several and diverse functions in a diseased subject.

In the list of results, ordered from highest to lowest degree, miRNAs occupy the first positions with their degrees oscillating between 893 and 17, and afterwards comes the genes. In fact, it is an obvious order of appearance if we take in consideration that the miRNAs have been forced to constitute the basis of the network, as it is a miRNA-target RNA network. In addition to this, dysregulated miRNAs seem to contribute notably to the disease signature and they have been proven to correspond to nodes with higher degree in a miRNA-genes network [43].

From the 3277 target genes,

- 0.15% show degree 5
- 0.73% have degree 4
- 99.12% corresponds to genes with degree 3 and less.

Low degree genes show fewer interactions which implies a weaker association to dysregulated miRNAs as well as a reduced susceptibility to be altered. Hence, a degree filter is applied to eliminate from the map those genes with degree less than 4. The change in the network after the degree filter is quite illustrative (Figure 4.4).

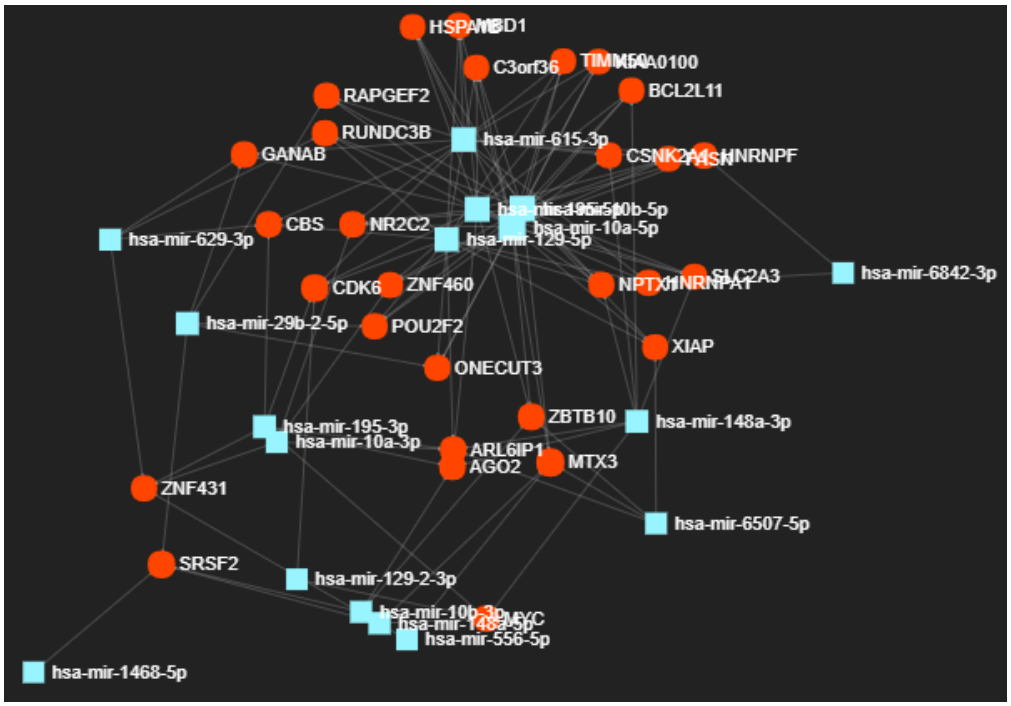


Figure 4.4. Interaction network after the application of the degree filter.

As a result, the network overview details showed lower values than before: 46 nodes (17 miRNAs and 29 targets) and 121 edges. Furthermore, the degree of all the nodes was obviously reduced. Now, the highest degree is 20. The miRNAs with higher degree, Table 4.5, are the same of Table 4.3 but in a different order. The genes with highest degree keep unchanged as expected.

TABLE 4.5. MIRNAS WITH HIGHEST DEGREE IN THE NETWORK ONCE THE DEGREE FILTER HAS BEEN APPLIED.

Pos	miRNAs	Degree
1	hsa-mir-10a-5p	20
2	hsa-mir-10b-5p	18
3	hsa-mir-195-5p	14
4	hsa-mir-615-3p	14
5	hsa-mir-129-5p	12

miRNet [38] offers the option of selecting the nodes of interest and extracting from the complete network the interactions affecting to those particular nodes. Figure 4.5 illustrates the subnetwork obtained when hsa-mir10a-5p and hsa-mir-10b-5p, the highest degree miRNAs after the degree filter, were chosen. Eighteen genes are commonly regulated by these first two miRNAs. Following the degree order, when the third miRNA is added to the subnetwork, the number of common regulated genes is halved. If the fourth miRNA is introduced only 4 genes out of the initial 8 remain interconnected (Table 4.6).

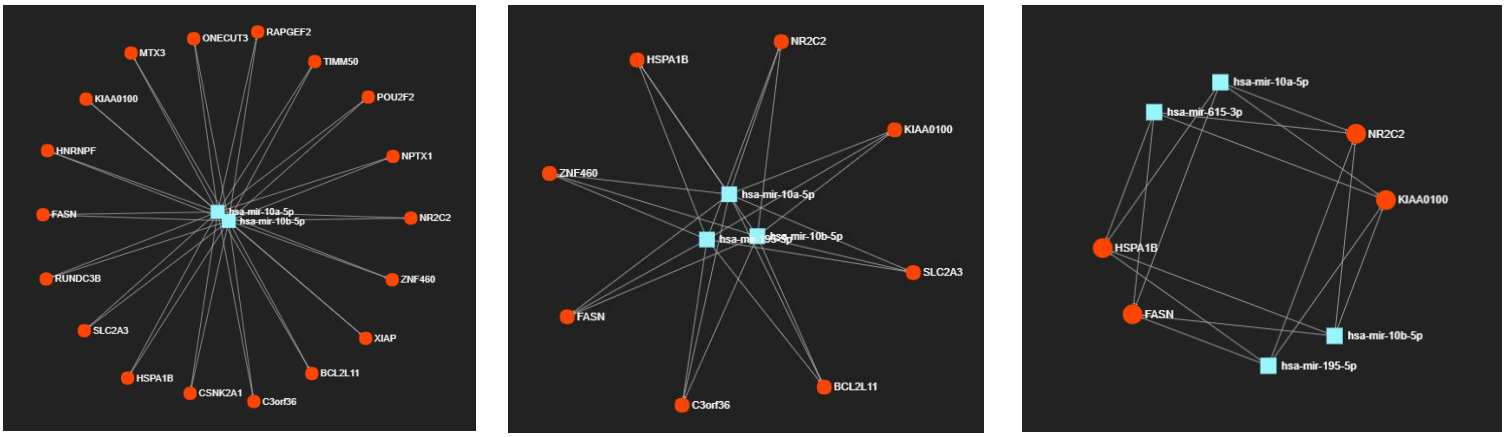


Figure 4.5. Subnetworks extracted from the filtered network when selecting specific miRNAs.

TABLE 4.6. TARGET GENES SHARED BY THE MIRNAS WITH HIGHER DEGREE.

GENE	1st two miRNAs	1st three miRNAs	1st four miRNAs
POU2F2	X		

XIAP	X		
NPTX1	X		
TIMM50	X		
RUNDC3B	X		
ONECUT3	X		
C3orf36	X	X	
CSNK2A1	X		
MTX3	X		
BCL2L11	X	X	
FASN	X	X	X
HNRNPF	X		
HSPA1B	X	X	X
NR2C2	X	X	X
RAPGEF2	X		
KIA0010	X	X	X
ZNF460	X	X	
SLC2A3	X	X	

If the four miRNAs with higher degree are effectively the most dysregulated, as presumable, the greater impact will be first noticed in their direct interactors. In this sense, FASN, HSPA1B, NR2C2 and KIAPP10 deserve to be highlighted for further research.

On the other way around, different combinations of the genes with the higher degree lead to the miRNAs present in the top-5 after the application of the filter (Figure 4.6). The extent of these miRNAs is confirmed to be somehow distributed all around the network and in consequence, they stand as susceptible candidates for biological validation.

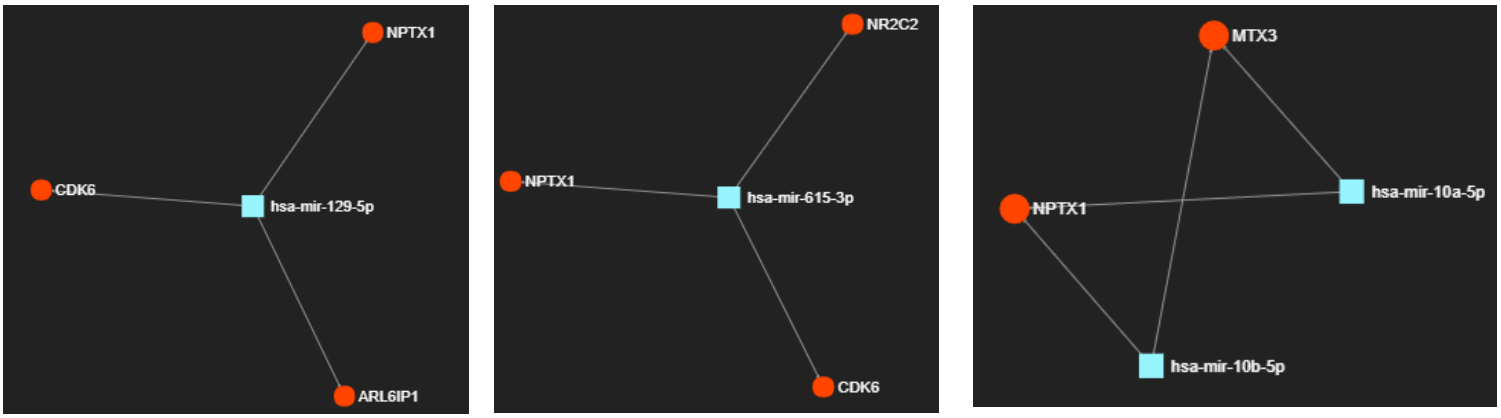
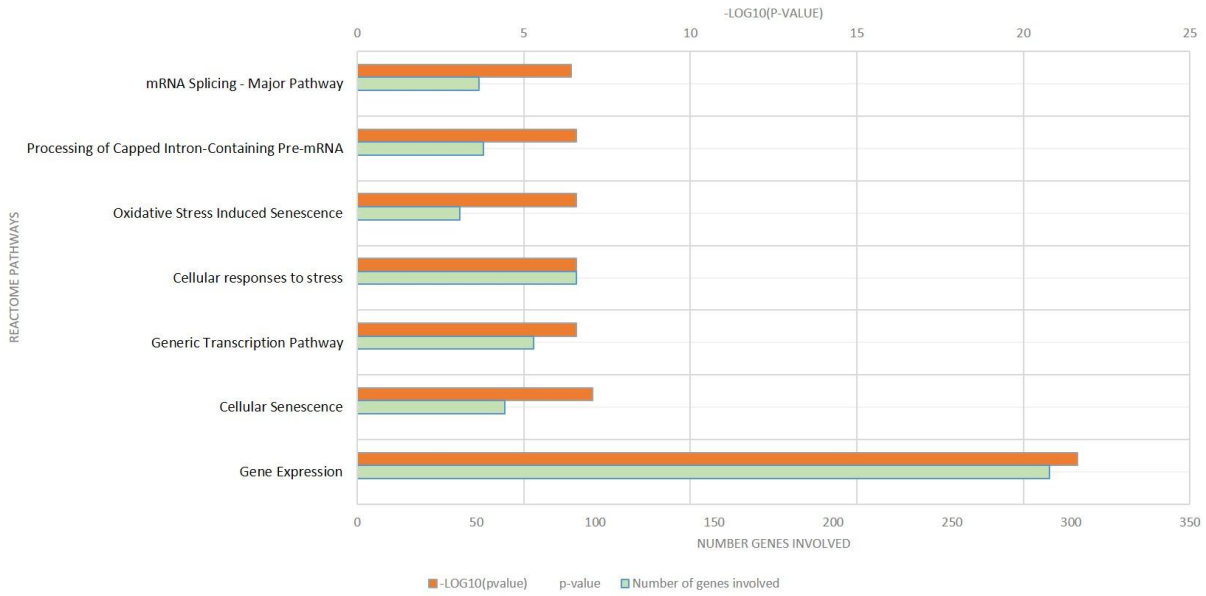


Figure 4.6. Subnetworks extracted from the filtered network when selecting specific miRNAs.

4.2.2. Functional enrichment analysis

miRNet [38] also provides the opportunity to evaluate if any function is significantly enriched. The sample space is required to be sufficiently large to get significant results. That is why the functional enrichment analysis was only performed within the frame of the original network.

The statistical method employed was a hypergeometric test and *Reactome* [41], the database where it looked for outcomes. Functional modules are validated through this algorithm by looking for common GO annotations on the gene network selected. As a result, 223 processes with an interesting biological meaning were shown, among which, it is worth highlighting the first hits (those with lower p-value) (Figure 4.7).



	Gene Expression	Cellular Senescence	Generic Transcription Pathway	Cellular responses to stress	Oxidative Stress Induced Senescence	Processing of Capped Intron-Containing Pre-mRNA	mRNA Splicing - Major Pathway
Numer of genes involved	291	62	74	92	43	53	51
p-value	2.31e-22	8.71e-8	2.65e-7	2.65e-7	2.65e-7	2.65e-7	3.73e-7

Figure 4.7. Main significant functionally enriched *Reactome* pathways.

It can be stated that the results from the functional analysis are not random but keep a close relationship with concepts and ideas that have been exposed along the thesis and that are ultimately associated with the three genodermatoses of the study. In order to assess the dependency of the highlighted pathways, the genes involved for each process are compared to each other. Figure 4.8 is another network that serves us to explain the complementarity of the functions enriched. *Venny* [37] facilitated the counting as it returns the exact number of shared components in the overlapping sections of the diagram. The figures obtained are the ones that appear above the arrows of Figure 4.8 connecting the related terms.

It is possible to distinguish three main clusters around the main one (Gene Expression):

- One cluster associated to cellular state, that englobes oxidative stress induced senescence, cellular senescence and cellular response to stress
- Another one related to the transcription procedure (generic transcription pathway)
- And, the last one related to the synthesis of RNA where mRNA splicing and processing of capped-intron containing pre-mRNA

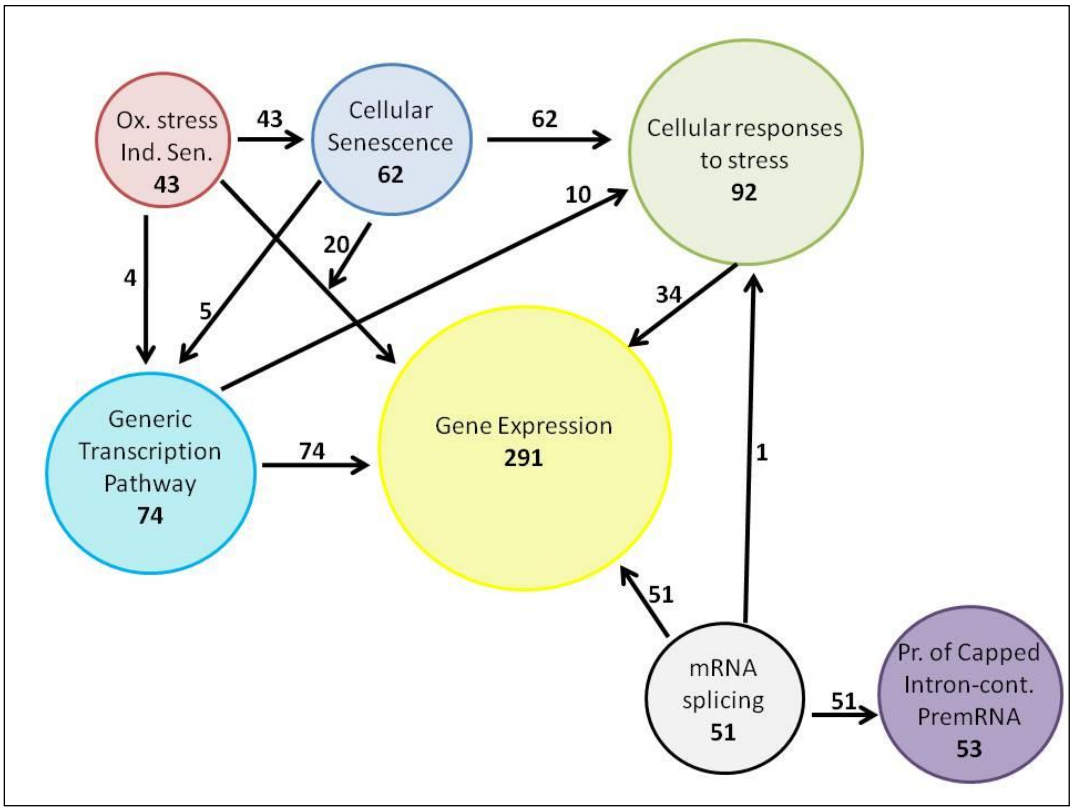


Figure 4.8. Network with the most significant functionally enriched pathways.

Although the functional enrichment analysis is performed on the target genes, it is necessary to keep the link with miRNAs. In this sense, the targeted genes associated with each of the emphasized pathways were introduced as input in *miRNet* [38]. To limit the network, a manual batch filter was applied with the five miRNAs from Table 4.3.

Table 4.7 depicts the miRNA degree for each RNA-miRNA network. From the information of this table, it can confirmed that the five miRNAs regulate many of the genes implicated in the observable phenotypes of the diseases.

TABLE 4.7. DEGREE OF THE FIVE SELECTED MIRNAS ASSOCIATED TO EACH ENRICHED PATHWAY.

Genes involved in enriched pathways	miRNAs	Degree
43 genes Oxidative Stress Induced Senescence	hsa-mir-615-3p	14
	hsa-mir-195-5p	13
	hsa-mir-129-5p	10
	hsa-mir-10b-5p	6
	hsa-mir-10a-5p	5

62 genes Cellular senescence	hsa-mir-615-3p	20
	hsa-mir-195-5p	17
	hsa-mir-129-5p	12
	hsa-mir-10b-5p	8
	hsa-mir-10a-5p	6
92 genes Cellular responses to stress	hsa-mir-615-3p	34
	hsa-mir-195-5p	24
	hsa-mir-129-5p	16
	hsa-mir-10a-5p	16
	hsa-mir-10b-5p	11
291 genes Gene Expression	hsa-mir-615-3p	124
	hsa-mir-195-5p	58
	hsa-mir-10a-5p	49
	hsa-mir-129-5p	42
	hsa-mir-10b-5p	31
74 genes Generic Transcription Pathways	hsa-mir-615-3p	19
	hsa-mir-195-5p	18
	hsa-mir-10a-5p	14
	hsa-mir-10b-5p	13
	hsa-mir-129-5p	11
53 genes Processing of Capped Intron-Containing Pre-mRNA	hsa-mir-615-3p	36
	hsa-mir-195-5p	10
	hsa-mir-10a-5p	9
	hsa-mir-129-5p	7
	hsa-mir-10b-5p	4
51 genes mRNA splicing	hsa-mir-615-3p	35
	hsa-mir-195-5p	10
	hsa-mir-10a-5p	9
	hsa-mir-129-5p	6
	hsa-mir-10b-5p	4

4.2.3. Grouping of the functionally enriched pathways

Even delimiting the study to the significant results (p-value < 0.05) that appeared in the functional enrichment analysis of *miRNet* [38], 223 hits continue to be an unmanageable quantity to draw conclusions from. Therefore, the program *REViGO* [42] was useful to classify and group the most relevant GO terms.

miRNet [38] functional analysis returns the name of the pathways and the p-values. Once the hits with 95% of reliability ($p<0.05$) were differentiated and they were assigned its corresponding GO code, they were submitted to *REViGO* [42]. The outcome is then classified attending to the three categories of GO. The spotlight will be made on the category of biological processes (Figures 4.9 and 4.10). The greatest part of the resultant figure is referred to cellular senescence, followed by a still large area highlighting the oxidative single-stranded RNA demethylation and the last big group that forms part of the most representative pathways is gene expression. These are quite interesting terms that coincide with the literature of the three genodermatoses as will be thoroughly discussed in the next section.

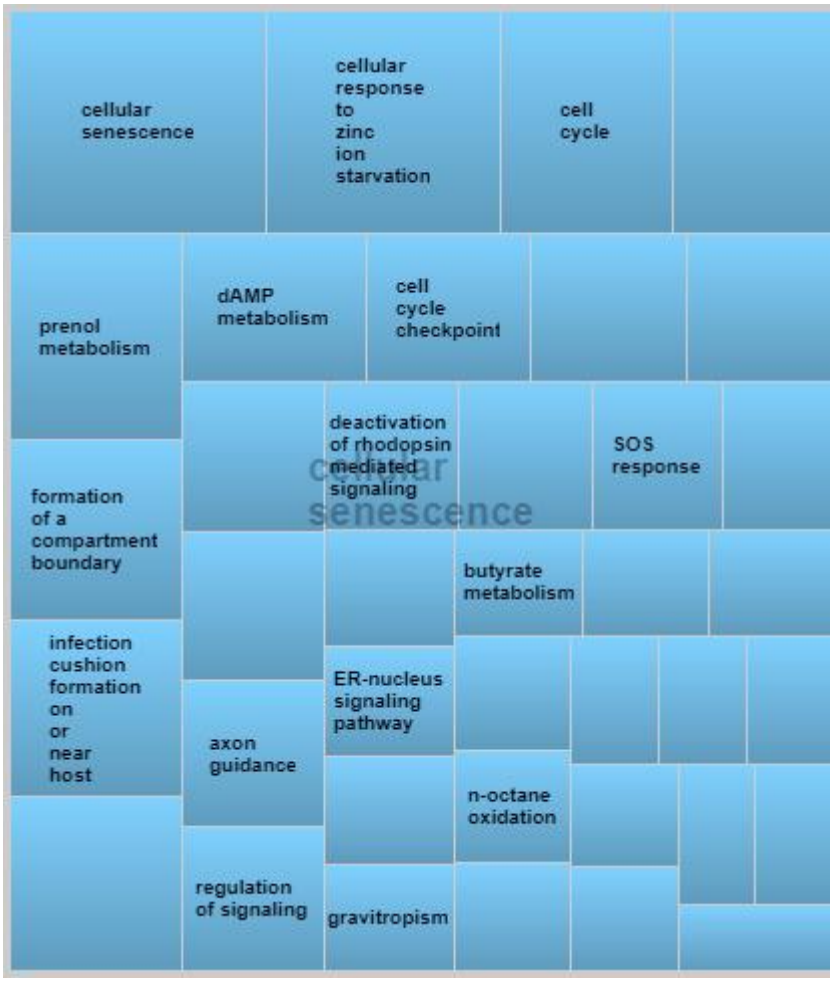


Figure 4.9. Grouping of the most relevant biological process GO terms highlighted by *REViGO*.

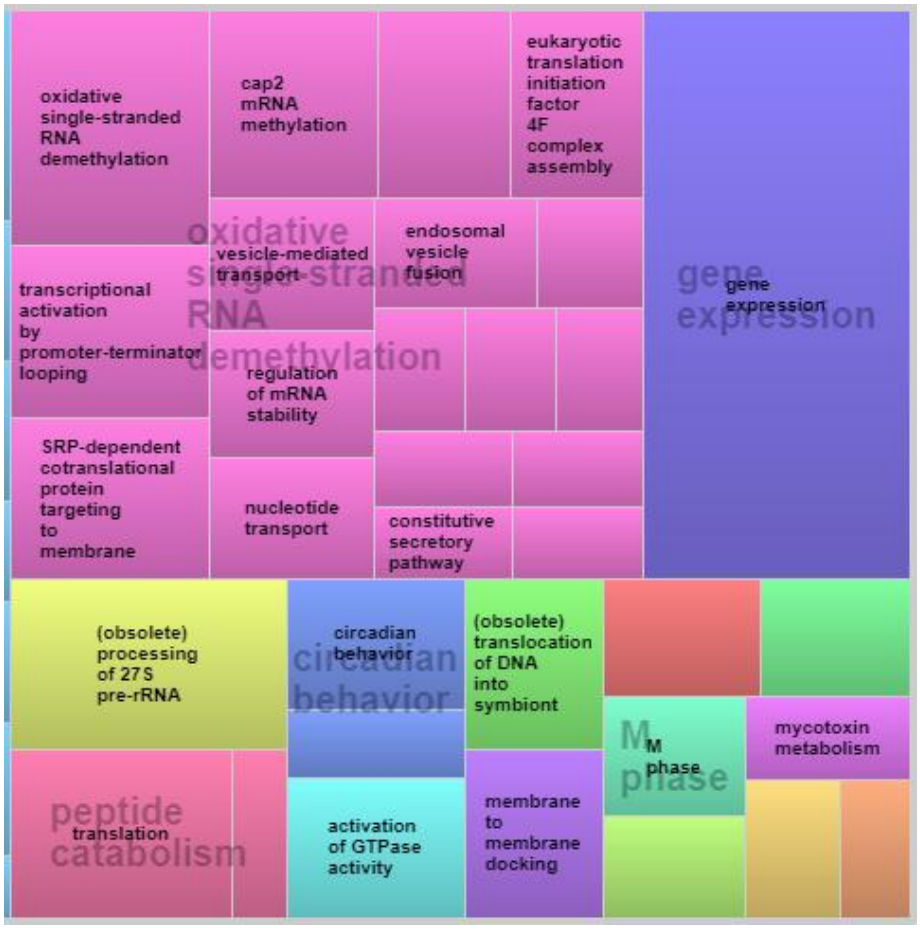


Figure 4.10. Grouping of the most relevant biological process GO terms highlighted by *REViGO*.

4.3. Comparison with previous results

This work was especially intended to continue the lines of research initiated by the RNA study of RDEB, KS and XP. Since the miRNA analysis was not started from scratch, it would be nonsense not to relate both studies in order to try to search common features and add additional findings to make progress. For these reasons, another approach consisted on seeking for the correlation among the 227 genes highlighted in the previous study and the 18 miRNAs of the actual one. A manual batch filter was applied to the original network so that target genes were inserted manually (Figure 4.11). The new network connects 13 miRNAs with 36 targets (Figure 4.12).

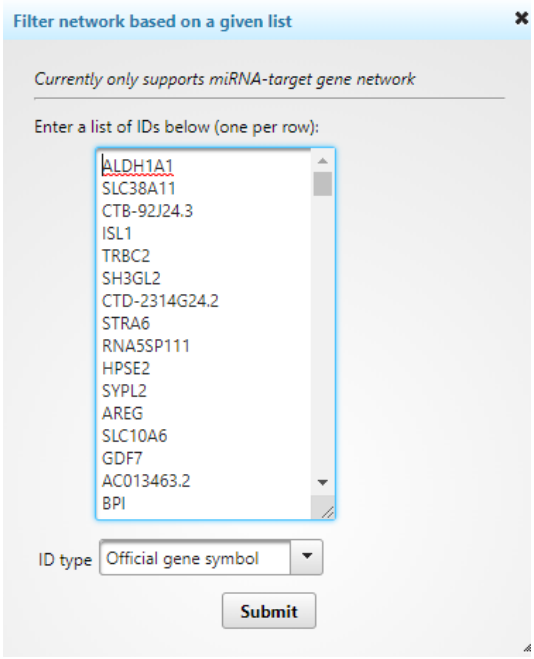


Figure 4.11. Introduction of miRNA-target genes with the manual batch filter.

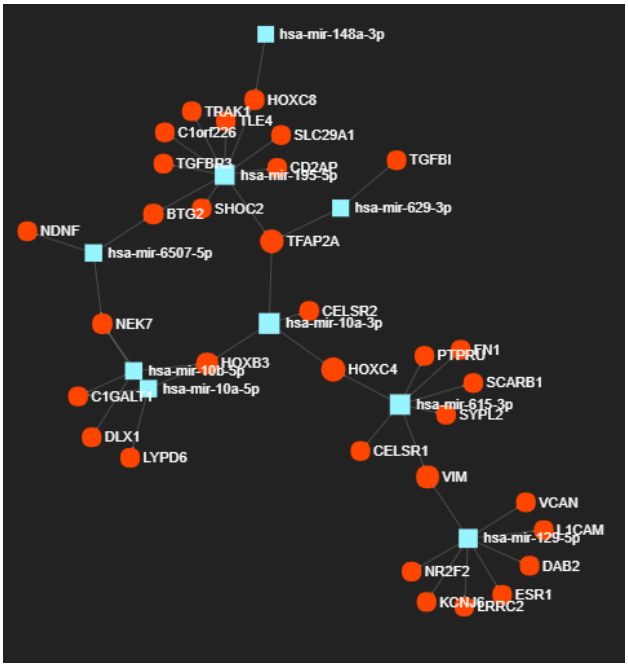


Figure 4.12. Interaction network after the application of the degree filter.

A pretty good point on the network generated is the presence of almost independent clusters. So in this graph, the significance of the nodes resides mainly in the betweeness (Table 4.8).

TABLE 4.8. MOST RELEVANT NODES IN THE FILTERED NETWORK.

Pos	Nodes	Betweeness
1	hsa-mir-10a-3p	462.33
2	hsa-mir-615-3p	419
3	HOXC4	375
4	hsa-mir-195-5p	354.33
5	TFAP2A	321.33
6	VIM	256
7	hsa-mir-129-5p	252
8	HOXB3	175.33

Furthermore, this comparison makes possible a deeper analysis in the relationship between altered miRNAs and altered RNA targets through their expression profile. When the RNA targets were unknown and *miRNet* [38] examined in its database the possible connections, the targets obtained lacked fold-change values. However, data of RNA targets was available from the previous study including logFC values.

MiRNA-RNA interactions keep, in theory, an inversely proportional relationship [44]. Therefore, activation of a miRNA will imply loss-of-function in the genes connected to it and affected by such alteration. Therefore, genes whose expression levels do not respond to the theory exposed above imply that their different state of expression is not due to that particular altered miRNA.

Table 4.9 represents the expression analysis performed. In order to establish a unique fold change value attributable to the three diseases, the median of the log fold change is calculated per gene. Yellow cells correspond to genes whose median log FC values tell us that the possible cause of their up/down-regulation may be generated by the miRNA they are associated to in the table.

TABLE 4.9.ANALYSIS OF THE EXPRESSION PROFILE DEPENDENCY BETWEEN MIRNAS AND RNA TARGET GENES.

miRNA	Genes	log FC (MEDIAN)	RDEB	KS	XPC
hsa-mir-195-5p + 2.93	BTG2	-2.254	-1.616	-2.254	-2.277
	TGFBR3	1.948	1.528	1.948	2.600
	TLE4	-1.691	-2.117	-1.691	-1.393
	TRAK1	1.363	1.213	1.363	1.401
	HOXC8	6.613	7.013	6.613	6.512
	SHOC2	1.809	1.433	1.809	1.821
	C1orf226	-4.820	-5.239	-4.820	-4.449
	CD2AP	1.817	1.653	1.817	1.866
	SLC29A1	-2.416	-2.416	-2.402	-2.615
	TFAP2A	-3.635	-3.635	-3.410	-4.636
hsa-mir-129-5p	VIM	1.357	1.157	1.357	1.559

+3.58	VCAN	4.194	3.995	4.194	4.411
	KCNJ6	3.763	4.631	3.250	3.763
	L1CAM	-5.724	-5.078	-5.724	-7.364
	ESR1	-3.656	-4.419	-3.656	-3.429
	LRRC2	3.445	3.609	3.445	3.411
	NR2F2	-1.753	-1.825	-1.753	-1.699
	DAB2	1.294	1.070	1.837	1.294
hsa-mir-615-3p +3.47	HOXC4	4.168	4.168	3.642	4.168
	SCARB1	-2.175	-1.960	-2.175	-2.471
	FN1	2.653	1.403	2.653	3.201
	VIM	1.357	1.157	1.357	1.559
	PTPRU	-2.069	-1.657	-2.701	-2.069
	SYPL2	-6.553	-6.553	-3.808	-8.256
	CELSR1	-3.827	-3.521	-3.827	-4.999
hsa-mir-10b-5p +5.14	C1GALT1	2.159	2.047	2.159	2.336
	DLX1	4.285	4.081	4.401	4.285
	NEK7	1.837	1.782	2.511	1.837
	HOXB3	6.238	6.068	6.238	6.787
hsa-mir-10a-3p +9.48	TFAP2A	-3.635	-3.635	-3.410	-4.636
	HOXC4	4.168	4.168	3.642	4.168
	CELSR2	-3.229	-3.157	-3.266	-3.229
	HOXB3	6.238	6.068	6.238	6.787
hsa-mir-10a-5p +10.44	LYPD6	2.526	3.097	2.526	2.240
	DLX1	4.285	4.081	4.401	4.285
	C1GALT1	2.159	2.047	2.159	2.336
	NEK7	1.837	1.782	2.511	1.837
hsa-mir-6507-5p -5.37	NDNF	4.047	4.047	3.505	4.078
	BTG2	-2.254	-1.616	-2.254	-2.277
	NEK7	1.837	1.782	2.511	1.837
hsa-mir-629-3p +1.89	TGFBI	2.715	2.369	2.947	2.715
	TFAP2A	-3.635	-3.635	-3.410	-4.636
hsa-mir-148a-3p +2.19	HOXC8	6.613	7.013	6.613	6.512

hsa-mir-556-5p -4.54	PTGER3	4.518	4.945	4.518	3.938
hsa-mir-148a-5p +2.01	AFF2	5.905	5.999	5.360	5.905
hsa-mir-195-3p +2.36	AR	-3.316,	-3.347	-3.316	-2.691
hsa-mir-6842-3p -2.77	SLC2A1	-2.239	-2.140	-2.803	-2.239

Once again, it can be stated that there is a similar correlation in the numerical values in the logFC for the three conditions. This correlation can be appreciated in the scatter plot of Figure 4.13.

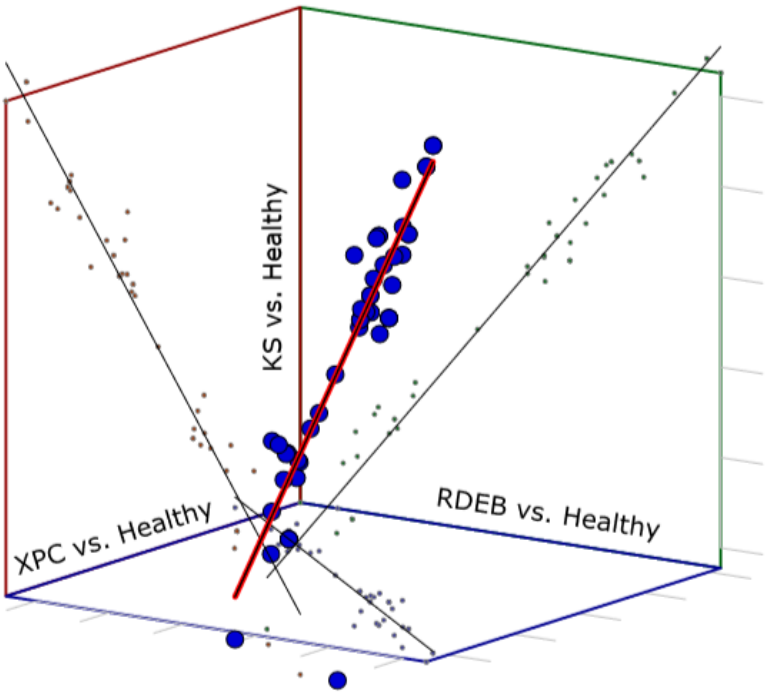


Figure 4.13. Scatter plot of the logFC of the genes from TABLE 4.9.

In line with the way of proceeding with previous results, it is convenient to underline a small selection that due to scientific reasons should be dug deeper. The five miRNAs that were highlighted on Table 4.3 and Table 4.5, are included in these 13 miRNAs. And out of this five, just three of them show a correlation in alteration of the genes

affected. It then would be interesting to look for the function and presence of those genes in the organism (Table 4.10).

A curious remark is that the two miRNAs that are left out in this last filter belong to the mir10 family. This family seems to regulate a large amount of pathways [45] so the fact that the expression of the RNA target genes (out of the 227) is not inhibited in this case may suggest that their overexpression may not be related to the disease mechanism of the three genodermatoses. They might have appeared as important due to their great number of connections in the network but they could end up being out of study. Nevertheless, it would be a good idea to validate this hypothesis.

TABLE 4.10. SELECTION OF MIRNAS-GENES BASED ON PREVIOUS RESULTS.

miRNAs	Target genes
hsa-mir-195-5p	BTG2, TLE4, C1orf226, SLC29A1, TFAP2A
hsa-mir-615-3p	SCARB1, PTPRU, SYPL2, CELSR1
hsa-mir-129-5p	L1CAM, ESR1, NR2F2

5. DISCUSSION

It is becoming more and more evident that the richness of emerging data lies in the information researchers can extract from it and apply into scientific advances. This, in a large extent, also depends on the storage and organization of the data so as to be accessible and manageable.

A new interdisciplinary science was born due to the necessity of handling the inconceivable amount of biological data that started to be generated, especially since the Human Genome Project and subsequent genomic sequence experiments [46]. Bioinformatics works as a powerful predictive tool in the phase before laboratory testing and validation. It takes into account a great diversity of parameters to try to recreate and model the reality as precisely as possible, and come up with a selective group of elements that has shown statistical significance of being involved in a particular process. Prioritization plays a fundamental role in directing a study to a more focused area, mainly at the beginning when a wide range of possibilities is open.

5.1. Strategies employed in the prioritization step

Different parameters can be taken into account when looking for preferential candidates for further research. Depending on the purpose of the study, greater weight will be given to one or other variable. Herein, the most common parameters that can be used to rank the miRNAs of biological interest are discussed:

a) **p-value:** It is a value widely used in scientific experiments to discriminate data that does not satisfy the significant level imposed. Most times, the cut-off is made at p-value<0.05 so that the results with a p-value greater than 0.05 are ignored. This means that there exists a 95% of confidence that the given results are true. Nevertheless, there would still be a 5% of doubt that could give rise to false positives and false negatives. Accordingly, the corrected p-value, FDR, can be used with the same level of significance. The main difference is that FDR adjusts better to data in which many variables appear from few subjects. Of course this parameter provides an accurate filter but it could not be employed alone.

Furthermore, p-value and FDR depend a lot in the statistical power, this is, the number of samples vs. the number of variables and the effect of the causal relationship. The greater the sample size, the smaller the range of uncertainty with respect to the consistency of the data. One of the main drawbacks when studying rare diseases is the small number of subjects recruited. In this case, the sampling space was large enough to get valuable results. But, for example, the functional analysis could not be performed with the filtered network as the majority of pathways enriched appeared with a p-value=1 and therefore, they could not be considered significant.

b) **Fold change:** It provides the biological meaning. The change is quantified in comparison with a benchmark. First, it must be assured that controls, in this case corresponding to healthy patients serve as correct and precise references. False negatives in this step would have baneful effects, even if the strategy is the most accurate and precise. In this case, the fold change could be a double-edged sword. A gene may be regulated by several miRNAs in such a way that, the contribution of up-regulated and down-regulated miRNAs at the same time could cancel each other and, as a result, find no significant differences with the expression profile of the healthy controls. Fold change would be then reliable in a one-to-one interaction but in a multiple contribution procedure, it may not be possible to detect easily the altered genes. Even with several hidden false negatives, the advantage of this parameter is the fact that the validity of positives is almost unquestionable. MiRNAs that show a considerable fold change are truly affecting the state of genes connected to them and that show the same level of difference from controls.

c) **Relevance in a miRNA-target genes network:** Their position and connection within a network can say a lot about the influence of a node on other nodes or in the network in general. Degree and betweeness are useful for filtering and clustering and for giving explanation to the diversity of mismatches that diseases cause in an organism. Attending to this criterion, relevant nodes are more probable to be associated to disease but typically, a lot of nodes may pass these filters. Then, other tools will be needed for a more precise classification.

d) **Target genes involved in a certain ontology/pathway of interest:** Functional enrichment analysis serves to validate the participation of a specific component, function or process for which there is a still unclear evidence of its contribution. It is usually performed at the end when different proposals and hypotheses have been made. Therefore, the enrichment analysis is considered fruitful whenever results are expected. Databases that perform this type of selection usually return large amount of information. Hence, it is a risky option to go hypothesis-free through this analysis.

e) **Common results to other condtions:** In this particular case, in which three diseases are being compared, a variable focused on similarity must be added. Any filtering made according to this parameter will be considered to be based on serious grounds. Indeed, the conclusions are desired to be certain for the three conditions, so the fact of appearing in just one of them is a qualifying criterion.

In the end, we should be able to demonstrate the level of confidence in that our predicted results will adjust to reality. Traditionally, a combination of all parameters is performed [57-59].

5.2. Selection of relevant miRNA candidates for further research

After all the aforesaid, miRNAs that have demonstrated to potentially influence on the development of RDEB, KS and XP will be summarized in this section, along with the supporting justifications.

At the beginning, there were 18 miRNAs (section 4.1) that appeared altered in the three disorders and with a FDR < 0.05. Both reasons made them susceptible to be highly related with disease mechanisms and they are therefore the first conclusive candidates for further research. Nevertheless, as further research implies investment, the field of possibilities should be reduced to a more selected group. In this procedure, the other parameters will come into play.

After the building of the network of miRNAs and target genes, 5 out of the previous 18 miRNAs were highlighted due to their high degree: hsa-mir-615-3p, hsa-mir-195-5p, hsa-mir-129-5p, hsa-mir-10a-5p and hsa-mir-10b-5p (Table 4.1). A condition they maintained even after the application of the degree filter (Table 4.3). Furthermore, they

were also regulating the target genes with the highest degree (Table 4.2). All of this leads to the conclusion that these five miRNAs are the nodes that mainly sustain and control the entire network. If dysregulated, their impact will be notorious on almost all nodes and consequently it can explain the distorting phenotype of the genodermatoses.

The five miRNAs are also behind the targeted genes that show the most significant enriched pathways. They are listed in order of degree in Table 4.5. The first position in all of them is occupied by hsa-mir-615-3p, with a considerable difference with respect to the second position. This miRNA also appears in the selection made according to the log fold change relationship with transcripts from the previous study. Therefore, it can be confirmed that hsa-mir-615-3p is the miRNA that shows greatest evidence of being responsible for controlling the disease module in RDEB, KS and XP.

Regarding the other two highlighted miRNAs in relation to logFC, hsa-mir-129-5p has been related to squamous cell carcinoma in laryngeal tissue. MingHua *et al.* proved that has-mir-129-5p appeared significantly overexpressed in these type of carcinomas and its downregulation successfully reduced the progression and growth of the tumor [47]. On other side, hsa-mir-195-5p has been proposed as being involved in the inhibition of WNT signaling. And, as a result, responsible for causing loss-of-function in dermal papilla cells [48]. In the same way, these miRNAs may be behind the skin squamous cell carcinoma and loss of ability of skin components, two facts specially representative of the three conditions.

5.3. Selection of relevant RNA candidates for further research

Even though miRNAs are in the focus of the research, target genes that appear to be significant in the network also deserve to capture attention for future investigations. More than the ones with higher degree, a selection that spurs special interest corresponds to targeted genes that share connections with the outstanding miRNAs just mentioned above. Removing hsa-mir-129-5p from the group of five, the rest share interactions with FASN, HSPA1B, NR2C2, and KIAA0100.

- FASN, whose official name is fatty acid synthase. It is the gene encoding a lipogenic enzyme that takes part in fatty oxidation pathways [49]. In cancer, this

pathway is remarkable. The reason is that cancerous cells are known to excessively consume glucose in aerobic conditions through glycolysis and generates precursors to activate the lipogenic switch [50]. However, in this case, the four miRNAs are upregulated so that the gene should be repressed.

- HSPA1B, refers to a heat shock protein from the Hsp70 family. It mediates in the synthesis of proteins [49]. An analysis at protein level may give more information in this aspect.

- NR2C2 encodes a nuclear receptor protein that acts as a transcription factor and it also have a cellular protective function against oxidative stress. Therefore, if it is repressed the protective function could be weakened and oxidative stress could have a stronger impact in the organism. A characteristic present in the three genodermatoses that could be justified in part by this particular gene.

- KIAA0100: It has been found in breast carcinoma and seems to be related with miR-195 [49].

On other side, the functional enrichment analysis was performed on all genes of the global network. The results obtained in *miRNet* [38] along with the ones in *REViGO* [42] reveal clear variations in cellular senescence and oxidative imbalance, two processes that, to a large extent, justify the phenotypic profile of the three genodermatoses.

Several studies have brought to the fore the similarities between senescent fibroblasts and cells from aged tissues[50] [51] .Cells start to slow down the division process until they are not able to grow and proliferate anymore. However, they still continue being active especially through the senescence-associated secretory phenotype, known as SASP, that produce modifications in the extracellular matrix. In association with this, detrimental effects such as tumorigenesis are originated (Figure 5.1) [52].

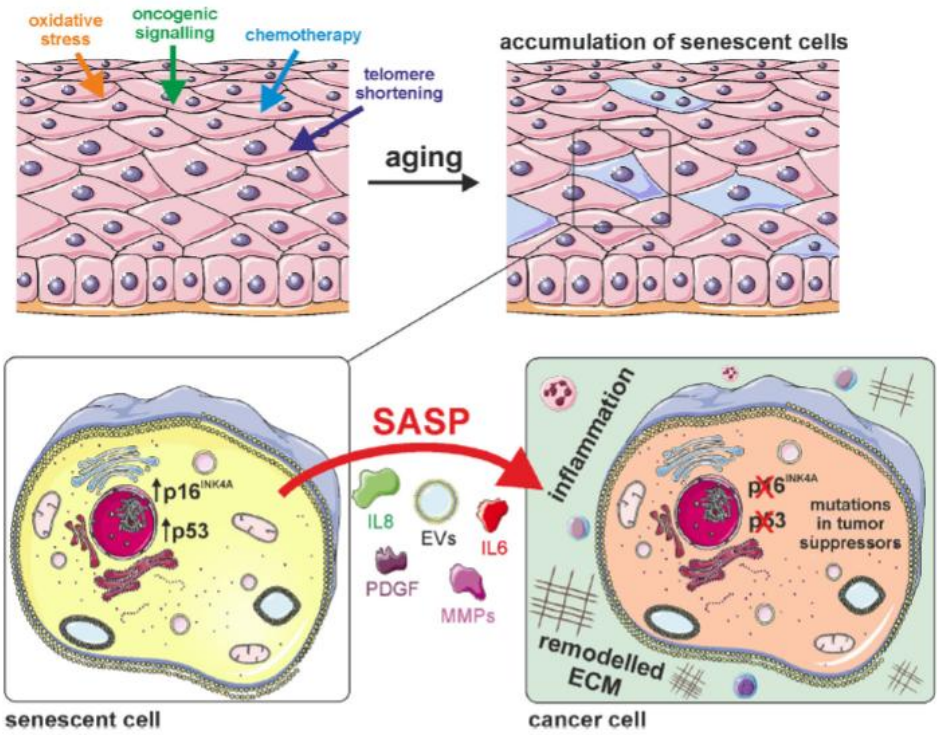


Figure 5.1. Representation of the pro-tumorigenic microenvironment created from SASP of senescent cells. Source: [53]

Literature also endorses the oxidative stress in dysfunctional skin [21][54]. Oxidative stress can be defined as the imbalance between reactive oxygen species (ROS) and the antioxidant cellular components. ROS are normally generated in the cell respiration process mainly at complex III in the electron transport chain (ETC) in mitochondria. As it happens in elderly people, a reduction in the rate of ETC involves an overproduction of ROS and pathways such as glycolysis are preferred instead [55]. Furthermore, the excessive build-up of ROS has been reported to be the cause for promoting cellular senescence [56]. Genes appearing as responsible for the enrichment of both processes are interesting to be thoroughly examined.

Finally, the other group of genes that could be interesting to highlight are those that show an inverse correlation in the log fold change with the miRNA fold change (Table 4.8). In order to make a relevant selection, all the genes that respond in opposite way to the dysregulation of their directed miRNAs may be explored for future lines. Among

them, it is worth underlining two downregulated genes linked to the upregulation of hsa-mir-615-3p:

- PTPRU, it belongs to the protein tyrosine phosphatase (PTP) family. It acts as a signaling molecule in a quite variety of processes, particularly cell-to-cell interactions [49].
- CELSR1 is in charge of producing cadherin, an epidermal growth factor. Cellular growth and migration are modified in terms of its expression [49].

In summary, with the miRNAs obtained from RDEB, KS, XPC and healthy patients it has been possible to perform a bioinformatic analysis directed to highlight specific molecules and pathways that are predicted to be altered in the three genodermatoses. For the prioritization, it has been tried to make a balance of five parameters that are understood to be essential for the selection process. These are: p-value/FDR, fold change, relevance in networks, involvement in certain ontology/pathway and similarity. Finally, five miRNAs have been chosen to continue with research for demonstrating to be key molecules in the analysis. The final miRNA candidates for further validation are:

- hsa-mir-615-3p, hsa-mir-195-5p and hsa-mir-129-5p with great expectations in justifying the phenotype of RDEB, KS and XPC.
- hsa-mir-10a-5p and hsa-mir-10b-5p with doubts in their possible participation in the disease mechanism of RDEB, KS and XPC.

With reference to the genes, FASN, HSPA1B, NR2C2, KIAA0100, PTPRU and CELSR1 have been concluded to take part, in some way, in the development of the diseases. But, maybe, the most relevant point of this thesis to the research is the fact that most miRNAs and genes, direct or indirectly, contribute to the oxidative stress, cellular senescence and tumorigenesis. Future research might be focused in clarifying the wiring diagram behind these outstanding pathways.

6. SOCIO-ECONOMIC IMPACT

From a social point of view, this project can suppose an incredible advance on treating rare diseases. With the excuse that it affects a small percentage of population, sometimes they fall into neglect. But the truth is that hospitals and clinical sites are full of cases like the three genodermatoses under study. The only way to address with unknown diseases is biomedical research. Research projects are focused on giving answer to the two essential questions: what is going wrong in the organism and what can be done to correct it. And the response to these questions is highly related with the final aim of increasing as far as possible people's quality of life.

On the other hand, the economic impact could also be notorious. Given the predictive nature of bioinformatic analysis, costs and time can be both reduced. If the number of molecules to be validated decreases, it is evident that logistics and resources will be better economized. However, it is parallely necessary to improve the storage in databases and develop more precise algorithms so that the probability of success is considerably higher than the margin of error. The estimated budget to carry out the project in this thesis is summarized in Table 6.1.

TABLE 6.1. ESTIMATED BUDGET FOR THE ACCOMPLISHMENT OF THE THESIS PROJECT

FUNGIBLE	miRNA seq	Illumina sequencing	Sample preparation	1.330,07 €
			Sequencing	1.063,83 €
		Bioinformatics	Data handling and delivery	132,56 €
			Data analysis	868,89 €
EQUIPMENT	Bioinformatic analysis	Hardware	Computer	700,00 €
		Software	Databases, programs, computing tools, papers	0,00 €
PERSONNEL	Research staff	Research coordinator	250 hours	3.750,00 €
		Research assistant	550 hours	6.600,00 €
			TOTAL :	14.445,35 €

7. CONCLUSIONS AND FUTURE DIRECTIONS

Networks have clearly demonstrated to be useful in the interpretation of huge amounts of data thanks to its simplification in manageable and understandable ordered structures. With many biological processes still to be unveiled, knowledge on the general functioning of the organism can be at least described by defining their connections to other pathways. The emergence of a wide variety of network sciences has shifted public interest to a general insight of system biology. These are exactly the reasons why we have focused our efforts on knowing the impact of the eighteen common dysregulated miRNAs on a large scale. Although, afterwards, it was necessary to narrow down to specific nodes for retaining a small outstanding selection.

The approach for prioritization of miRNAs and target genes made along this thesis has been based on repetitive presence when attending to different criteria. The greater the presence, the larger the probability of being key pieces in the control of disease modules. Nevertheless, with a deeper knowledge on statistics, a more precise ranking of miRNAs and target genes could have been achieved. The desired objective would be to create an automatic score formula in which all participating variables are taken into account and depending on the circumstances, they received a higher or lower weight. Several attempts have been made oriented to develop techniques and scores for prioritizing genes and miRNAs, trying to integrate statistical information without losing the biological meaning [57][58][59]. However, there is still a long way to achieve the appropriate score formula.

As a conclusion, we can state that there is correlation in the results obtained with the phenotypic profile of the disease. It is possible to justify, in broad terms, the carcinogenic predisposition of RDEB, KS and XP through the cellular senescence and oxidative stress altered pathways. Nevertheless, it is important to recall that the performance of all the databases used along the analysis is based on the data that has been initially and progressively uploaded to create it. This means that networks are generated by algorithms that integrate the information we give with what the database already has. Hence, reliability of the results is not assured and depends greatly on the

completeness of the database itself. Hypotheses and premises should be sought previously in the literature to have the ability to interpret the quality of the networks.

Framed in the translational research, the bioinformatics analysis performed has a potentially important therapeutic use through two immediately direct clinical applications:

- The first one consists on the recognition of biomarkers that appear as pathogenic signatures of the disease. The diagnosis step determines in great part the success or failure of any treatment. Personalized medicine is the new trend that fights to be implemented in daily medical practice understanding the uniqueness of each patient to undergo its own personalized treatment according to the genome differences.

  In theranostics, molecular biomarkers serve to classify patients so that they receive the most effective treatment according to their condition.

- The second one is focused on finding target molecules to which effective drug treatments could be directed. Surely, the disease mechanism behind the three genodermatoses is originated by a group of events happening at specific conditions, hence, a multiple origin. In this context, bioinformatics try to predict which are the key molecules that provoke the distortive cascade. By targeting those specific compounds, symptoms are expected to be largely palliated.

  Drug repurposing is a considerably recent trend that chases the use of already approved drugs for new targets. Both cost and time are substantially reduced given the enormous difficulties to launch new drugs into the market. In fact, it is estimated that less than 15% of chemical substances that initiate the drug development process are finally approved [60].In this sense, studies should try to look for drugs that can be reused to fight against the excessive oxidative stress. The key is to find a previous pathway that contributes to the oxidative stress and to which there is a known successful targeted treatment.

To sum up, miRNAs have just opened a new door for the discovery of disease mechanisms and bioinformatics has proven to be an indispensable requirement for researchers to cross through it.

8. BIBLIOGRAPHY

[1] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network Medicine: A Network-based Approach to Human Disease," *Nat Rev Genet*, vol. 12, no. 1, pp. 56–68, 2011.

[2] A.-L. Barabási, "Network science book," no. [Online] Available at http://networksciencebook.com/.

[3] M. Vidal and M. Cusick, "Interactome Networks and Human Disease," *Cell*, vol. 144, no. 6, pp. 986–998, 2011.

[4] X. Zhou, J. Menche, A. L. Barabási, and A. Sharma, "Human symptoms-disease network," *Nat. Commun.*, vol. 5, no. May, 2014.

[5] T. Arodz, D. Bonchev, and R. F. Diegelmann, "A Network Approach to Wound Healing," *Adv. Wound Care*, vol. 2, no. 9, pp. 499–509, 2013.

[6] A.-L. Barabási, "Network Medicine — From Obesity to the ' Diseasome ,'" *N. Engl. J. Med.*, pp. 404–407, 2007.

[7] K. Il Goh and I. G. Choi, "Exploring the human diseasome: The human disease network," *Brief. Funct. Genomics*, vol. 11, no. 6, pp. 533–542, 2012.

[8] "¿Qué son las enfermedades raras o poco frecuentes?," *FEDER*, no. [Online] Available at:https://enfermedades-raras.org/index.php/enfermedades-raras.

[9] Y.-Z. Ng *et al.*, "Fibroblast-Derived Dermal Matrix Drives Development of Aggressive Cutaneous Squamous Cell Carcinoma in Patients with Recessive Dystrophic Epidermolysis Bullosa," *Cancer Res.*, vol. 72, no. 14, pp. 3522–3534, 2012.

[10] "Vesiculobullous Diseases Part 3," *what-when-how*, no. [Online] Available at: http://what-when-com/acp-medicine/vesiculobullous-diseases-part-3/.

[11] "Understanding EB, EB in depth, About EB," *Debra Am.*, no. [Online] Available at: http://www.debra.org/abouteb.

[12] S. Shinkuma, "Dystrophic epidermolysis bullosa : a review," *Dove Press J. Clin. Cosmet. Investig. Dermatology*, pp. 275–284, 2015.

[13] "DOSSIER DE PRENSA DE LA ASOCIACIÓN DEBRA-PIEL DE MARIPOSA ASOCIACIÓN DEBRA-PIEL DE MARIPOSA," no. [Online] Available at: https://www.pieldemariposa.es/fotos/file/Dossier%20Prensa%20actualizado%202018/Dossier%20prensa%20DEBRA%20Piel%20de%20Mariposa%202018.pdf.

[14] M. J. Escámez Toledano, "Reunión de la Red Española de Registros de Enfermedades Raras para la Investigación (SpainRDR) - Registro Nacional de Epidermolisis Bullosa en 2014 : primeros pacientes registrados," no. [Online] Available at: https://spainrdr.isciii.es/es/Documents/Dic14SpainRDR/SESION_5/Registro_Nacional_EB2014.pdf.

[15] Á. Villarino, "Functional Analysis of altered expression profiles in patients with Epidermolysis Bullosa." Final degree project, October 2015.

[16] "Epidermólisis bullosa. Tipologías," *Debra España*, no. [Online] Availabe at: https://www.pieldemariposa.es/epidermolisis-bullosa/tipologias/.

[17] "Kindler Syndrome," *Genet. Home Ref. –NIH*, no. [Online] Available at: https://ghr.nlm.nih.gov/condition/kindler-syndrome#resources.

[18] L. Mendes, "Kindler syndrome: report of two cases," *An Bras Dermatol*, vol. 87, no. 5, pp. 779–81, 2012.

[19] S. Ghosh, S.K., Bandyopadhyay, D., Das, J., Chatterjee, G., Sarkar, "Kindler's

syndrome: A case series of three Indian children," *Indian J Dermatol.*, vol. 55, no. 4, pp. 393–396, 2010.
[20] A. Das Navya Handa, Dilip Kachhawa, Vinod Kumar Jain, Pankaj Rao, "Kindler's Syndrome: A Tale of Two Siblings," *Indian J Dermatol.*, vol. 61, no. 4, p. 468.
[21] E. Zapatero-solana *et al.*, "Oxidative stress and mitochondrial dysfunction in Kindler syndrome," *Orphanet J. Rare Dis.*, pp. 1–10, 2014.
[22] A. R. Lehmann, D. McGibbon, and M. Stefanini, "Xeroderma pigmentosum," *Orphanet J. Rare Dis.*, vol. 6, no. 1, p. 70, 2011.
[23] N. Orens Viedma, "Xeroderma Pigmentoso, una enfermedad rara," no. [Online] Available at: https://integrasaludtalavera.com/xeroderma-pigmentoso-una-enfermedad-rara/.
[24] J. DiGiovanna and K. Kraemer, "Shining a light on xeroderma pigmentosum," *J. Invest. Dermatol.*, vol. 132, no. 3, pp. 785–796, 2012.
[25] "Detalles de la patología: Xeroderma Pigmentoso," *FEDER*, no. [Online] Available at:https://www.enfermedades-raras.org/index.php/component/content/article?id=3100&idpat=411.
[26] K. Appasani, "Abstract," *MicroRNAs From Basic Sci. to Dis. Biol.*, p. i.
[27] F. Wahid, A. Shehzad, T. Khan, and Y. Y. Kim, "MicroRNAs: Synthesis, mechanism, function, and recent clinical trials," *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1803, no. 11, pp. 1231–1243, 2010.
[28] S.-L. Ying, S-Y., Miller, J.D., Lin, "Non-coding RNAs – development of man-made vector-based intronic microRNAs (miRNAs)," *MicroRNAs From Basic Sci. to Dis. Biol.*, pp. 22–41.
[29] V. R. Ambros, "Foreword," *MicroRNAs From Basic Sci. to Dis. Biol.*, pp. xxvii–xxviii.
[30] "Central dogma," *BioNinja*, no. [Online] Available at: http://ib.bioninja.com.au/standard-level/topic-2-molecular-biology/27-dna-replication-transcri/central-dogma.html.
[31] C. Catalanotto, C. Cogoni, and G. Zardo, "MicroRNA in control of gene expression: An overview of nuclear functions," *Int. J. Mol. Sci.*, vol. 17, no. 10, 2016.
[32] Y. Huang, X. J. Shen, Q. Zou, S. P. Wang, S. M. Tang, and G. Z. Zhang, "Biological functions of microRNAs: A review," *J. Physiol. Biochem.*, vol. 67, no. 1, pp. 129–139, 2011.
[33] J. T. Kathryn A. O'Donnell, K.A., and Mendell, "Dysregulation of microRNAs in human malignancy," *MicroRNAs From Basic Sci. to Dis. Biol.*, pp. 295–308.
[34] E. Chacón-solano *et al.*, "Fibroblasts activation and abnormal extracellular matrix remodeling as common hallmarks in three cancer-prone genodermatoses," no. being submitted.
[35] G. N. Corney, D.C., Basturea, "RNA-Seq using Next Generation Sequencing," *Mater Methods*, vol. 3 (203), 2013.
[36] A. Kozomara and S. Griffiths-Jones, "MiRBase: Annotating high confidence microRNAs using deep sequencing data," *Nucleic Acids Res.*, vol. 42, no. D1, pp. 68–73, 2014.
[37] J. C. Oliveros, "Venny. An interactive tool for comparing lists with Venn's diagrams.," no. Available at: http://bioinfogp.cnb.csic.es/tools/venny/index.html.
[38] Y. Fan, K. Siklenka, S. K. Arora, P. Ribeiro, S. Kimmins, and J. Xia, "miRNet - dissecting miRNA-target interactions and functional associations through

network-based visual analysis," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W135–W141, 2016.
[39] The Gene Ontology Consortium, "Expansion of the Gene Ontology knowledgebase and," vol. 45, no. November 2016, pp. 331–338, 2017.
[40] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG : new perspectives on genomes , pathways , diseases and drugs," vol. 45, no. November 2016, pp. 353–361, 2017.
[41] A. Fabregat *et al.*, "The Reactome Pathway Knowledgebase," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D649–D655, 2018.
[42] Š. T. Supek F, Bošnjak M, Škunca N, "REVIGO summarizes and visualizes long lists of Gene Ontology terms," *PLoS One*, 2011.
[43] S. Zafari, C. Backes, P. Leidinger, E. Meese, and A. Keller, "Regulatory MicroRNA Networks: Complex Patterns of Target Pathways for Disease-related and Housekeeping MicroRNAs," *Genomics, Proteomics Bioinforma.*, vol. 13, no. 3, pp. 159–168, 2015.
[44] M. Fabbri, R. Garzon, A. Cimmino, G. A. Calin, and C. M. Croce, "High throughput microRNAs profiling in cancers," *MicroRNAs From Basic Sci. to Dis. Biol.*, pp. 309–321.
[45] D. Tehler, N. M. Høyland-Kroghsbo, and A. H. Lund, "The miR-10 microRNA precursor family," *RNA Biol.*, vol. 8, no. 5, pp. 728–734, 2011.
[46] A. Bayat, "Clinical review - Science, medicine and the future: Bioinformatics," *BMJ*, vol. 324, no. July, pp. 1018–33, 2002.
[47] M. Li *et al.*, "Down-Regulation of miR-129-5p Inhibits Growth and Induces Apoptosis in Laryngeal Squamous Cell Carcinoma by Targeting APC," *PLoS One*, vol. 8, no. 10, p. e77829, 2013.
[48] N. Zhu *et al.*, "MiR-195-5p regulates hair follicle inductivity of dermal papilla cells by suppressing Wnt/ β-catenin activation," *Biomed Res. Int.*, vol. 2018, no. 1, pp. 1–13, 2018.
[49] L. Y. Geer *et al.*, "The NCBI BioSystems database," *Nucleic Acids Res.*, vol. 38, no. SUPPL.1, pp. 492–496, 2009.
[50] S. Breitenbach *et al.*, "Transcriptome and ultrastructural changes in dystrophic Epidermolysis bullosa resemble skin aging," vol. 7, no. 6, pp. 389–410, 2015.
[51] R. Naylor, D. Baker, J. van Deursen, A. Freund, C. K. Patil, and J. Campisi, "Senescent Cells: A Novel Therapeutic Target for Aging and Age- Related Diseases," *EMBO J.*, vol. 30, no. 8, pp. 105–116, 2011.
[52] M. Toutfaire, E. Bauwens, and F. Debacq-Chainiaux, "The impact of cellular senescence in skin ageing: A notion of mosaic and therapeutic strategies," *Biochem. Pharmacol.*, vol. 142, pp. 1–12, 2017.
[53] M. Schosserer, J. Grillari, and M. Breitenbach, "The Dual Role of Cellular Senescence in Developing Tumors and Their Response to Cancer Therapy," *Front. Oncol.*, vol. 7, no. November, 2017.
[54] M. Hayasi, "Roles of Oxidative Stress in Xeroderma Pigmentosum," *Mol. Mech. Xeroderma Pigment.*, no. Landes Bioscience and Springer Science+Business Media, 2008., pp. 120–127, 2008.
[55] B. Poljšak, R. G. Dahmane, and A. Godić, "Intrinsic skin aging: The role of oxidative stress," *Acta Dermatovenerologica Alpina, Pannonica Adriat.*, vol. 21, no. 2, pp. 33–36, 2012.
[56] M. C. Velarde, J. M. Flynn, N. U. Day, S. Melov, and J. Campisi, "Mitochondrial oxidative stress caused by Sod2 deficiency promotes cellular senescence and

aging phenotypes in the skin," *Aging (Albany. NY).*, vol. 4, no. 1, pp. 3–12, 2012.
[57] D. Jin and H. Lee, "Prioritizing cancer-related microRNAs by integrating microRNA and mRNA datasets," *Sci. Rep.*, vol. 6, no. July, pp. 1–14, 2016.
[58] Y. Xiao *et al.*, "A novel significance score for gene selection and ranking," *Bioinformatics*, vol. 30, no. 6, pp. 801–807, 2014.
[59] S. E. Lazic, "Ranking, selecting, and prioritising genes with desirability functions," *PeerJ*, vol. 3, p. e1444, 2015.
[60] S. M. Corsello *et al.*, "The Drug Repurposing Hub: a next-generation drug library and information resource," *Nat Med*, vol. 23, no. 4, pp. 405–408, 2017.
[61] "miRNA AAV and Adenovirus," *Vector Biolabs*, no. [Online], Available at: https://www.vectorbiolabs.com/mirna-aav-adenovirus/.

ANNEX

Table with the 28 significant miRNAS in EDBR vs Healthy

Gene_id	logFC	logCPM	PValue	FDR
hsa-miR-10a-5p	8.824	15.506	5.040e-13	6.310e-10
hsa-miR-10a-3p	8.590	6.421	1.107e-12	6.931e-10
hsa-miR-556-5p	-4.536	-1.136	1.383e-11	5.772e-09
hsa-miR-6507-5p	-4.779	-2.072	5.103e-08	1.597e-05
hsa-miR-6842-3p	-2.425	1.260	7.527e-08	1.885e-05
hsa-miR-195-3p	2.902	4.292	3.217e-07	6.712e-05
hsa-miR-556-3p	-3.859	-2.270	7.267e-07	1.300e-04
hsa-miR-129-5p	3.604	4.105	1.863e-06	2.915e-04
hsa-miR-129-2-3p	2.975	2.725	4.151e-06	5.774e-04
hsa-miR-29b-2-5p	-2.150	0.624	7.285e-06	8.440e-04
hsa-miR-10b-5p	4.379	15.553	7.415e-06	8.440e-04
hsa-miR-29c-5p	-1.901	2.887	1.737e-05	1.813e-03
hsa-miR-29c-3p	-2.172	7.219	2.798e-05	2.694e-03
hsa-miR-10b-3p	4.185	4.142	3.675e-05	3.286e-03
hsa-miR-1295a	-3.015	-0.935	6.489e-05	5.416e-03
hsa-miR-4488	-2.192	-1.061	7.257e-05	5.679e-03
hsa-miR-615-3p	3.925	9.194	8.503e-05	6.262e-03
hsa-miR-148a-5p	2.112	6.922	1.078e-04	7.499e-03
hsa-miR-146b-5p	-1.976	9.142	1.146e-04	7.552e-03
hsa-miR-615-5p	4.028	4.577	2.479e-04	1.531e-02
hsa-miR-874-5p	-1.612	1.284	2.686e-04	1.531e-02
hsa-miR-497-5p	2.303	7.525	2.690e-04	1.531e-02
hsa-miR-195-5p	2.555	7.207	5.443e-04	2.931e-02
hsa-miR-935	-2.624	-1.623	5.618e-04	2.931e-02
hsa-miR-629-3p	1.973	1.196	6.219e-04	3.011e-02
hsa-miR-148a-3p	2.069	12.783	6.253e-04	3.011e-02
hsa-miR-1468-5p	-2.008	4.891	6.669e-04	3.092e-02

Table with the 99 significant miRNAs in KS vs Healthy

Gene_id	logFC	logCPM	PValue	FDR
hsa-miR-1291	5.774	2.764	1.049e-17	1.011e-14
hsa-miR-10a-5p	10.496	16.335	2.217e-17	1.069e-14
hsa-miR-1260a	5.360	1.685	1.503e-15	4.830e-13
hsa-miR-10a-3p	9.535	6.274	3.761e-13	9.064e-11
hsa-miR-483-3p	-6.817	2.232	6.627e-13	1.278e-10
hsa-miR-483-5p	-7.343	2.243	1.383e-12	2.221e-10
hsa-miR-556-5p	-6.282	-0.557	4.896e-12	6.742e-10
hsa-miR-3607-5p	4.149	0.746	8.124e-11	9.789e-09
hsa-miR-1295a	-4.676	-0.373	2.420e-09	2.593e-07
hsa-miR-1468-5p	-3.589	5.281	3.161e-08	3.047e-06
hsa-miR-129-5p	3.582	3.331	4.865e-08	4.263e-06
hsa-miR-129-2-3p	3.237	2.552	1.004e-07	7.597e-06
hsa-miR-1247-5p	-5.513	0.273	1.025e-07	7.597e-06
hsa-miR-3605-3p	-2.427	3.054	1.662e-07	1.144e-05
hsa-miR-7974	2.918	4.238	3.096e-07	1.989e-05
hsa-miR-6842-3p	-2.770	1.902	6.803e-07	3.773e-05
hsa-miR-10b-5p	5.137	15.669	7.019e-07	3.773e-05
hsa-miR-29c-3p	-2.119	7.451	7.044e-07	3.773e-05
hsa-miR-148a-3p	2.276	12.498	1.562e-06	7.927e-05
hsa-miR-3615	-2.124	6.401	4.902e-06	2.363e-04
hsa-miR-137	2.503	5.253	5.442e-06	2.498e-04
hsa-miR-195-5p	2.936	6.994	6.427e-06	2.816e-04
hsa-miR-6507-5p	-5.372	-1.443	7.726e-06	3.108e-04
hsa-miR-3184-3p	-1.852	9.164	7.739e-06	3.108e-04
hsa-miR-31-3p	1.948	7.390	1.090e-05	4.202e-04
hsa-miR-561-5p	-1.929	5.927	1.152e-05	4.271e-04
hsa-miR-3529-3p	2.033	3.652	1.671e-05	5.965e-04
hsa-miR-4767	2.362	-0.0198	1.929e-05	6.641e-04
hsa-miR-10b-3p	4.932	4.236	2.146e-05	7.088e-04
hsa-miR-4517	2.388	-0.114	2.206e-05	7.088e-04

hsa-miR-4653-5p	3.374	-1.156	2.941e-05	9.145e-04
hsa-miR-221-5p	1.722	10.263	4.374e-05	1.318e-03
hsa-miR-29b-2-5p	-2.095	1.059	5.826e-05	1.702e-03
hsa-miR-877-5p	-1.821	5.758	6.856e-05	1.944e-03
hsa-miR-142-5p	3.386	2.908	8.156e-05	2.162e-03
hsa-miR-149-5p	-1.864	8.024	8.172e-05	2.162e-03
hsa-miR-5701	-6.849	-1.448	8.297e-05	2.162e-03
hsa-miR-155-3p	3.482	-1.475	1.150e-04	2.918e-03
hsa-miR-142-3p	3.617	1.494	1.246e-04	3.081e-03
hsa-miR-181a-5p	-1.970	13.824	1.466e-04	3.533e-03
hsa-let-7i-3p	1.954	6.066	1.542e-04	3.625e-03
hsa-miR-548ab	-1.939	1.601	2.066e-04	4.735e-03
hsa-miR-133a-3p	-3.795	1.366	2.112e-04	4.735e-03
hsa-miR-328-3p	-1.558	5.970	2.410e-04	5.281e-03
hsa-miR-128-3p	-1.500	8.466	2.565e-04	5.495e-03
hsa-miR-4521	2.209	4.993	2.626e-04	5.504e-03
hsa-miR-624-5p	2.134	-0.376	2.851e-04	5.847e-03
hsa-miR-1305	2.312	2.268	3.011e-04	6.047e-03
hsa-miR-100-3p	1.602	7.980	3.220e-04	6.336e-03
hsa-miR-629-3p	1.893	0.756	3.352e-04	6.463e-03
hsa-miR-98-3p	-1.439	4.571	3.613e-04	6.830e-03
hsa-miR-6784-3p	-2.060	-0.052	4.059e-04	7.525e-03
hsa-miR-1247-3p	-6.422	-1.805	4.445e-04	8.085e-03
hsa-miR-148a-5p	2.012	6.647	4.925e-04	8.792e-03
hsa-miR-887-3p	-1.571	7.305	5.145e-04	9.019e-03
hsa-miR-454-5p	-1.491	3.088	5.389e-04	9.276e-03
hsa-miR-326	-1.541	3.472	5.721e-04	9.564e-03
hsa-miR-556-3p	-3.428	-1.918	5.754e-04	9.564e-03
hsa-miR-7-5p	2.166	0.287	6.960e-04	1.137e-02
hsa-miR-145-5p	1.770	10.846	7.213e-04	1.159e-02
hsa-miR-6868-3p	-2.724	-0.975	7.424e-04	1.173 e-02
hsa-miR-6737-3p	-1.711	0.259	7.964e-04	1.238 e-02
hsa-miR-6769b-3p	-1.581	0.986	8.849e-04	1.354 e-02

hsa-miR-19b-1-5p	1.875	-0.054	9.072e-04	1.367 e-02
hsa-miR-496	1.472	3.573	9.418e-04	1.397 e-02
hsa-miR-195-3p	2.011	3.061	1.012e-03	1.478 e-02
hsa-miR-1972	1.740	0.585	1.148e-03	1.652 e-02
hsa-miR-135b-5p	2.387	0.980	1.179e-03	1.671 e-02
hsa-miR-1268a	1.630	2.690	1.202e-03	1.679 e-02
hsa-miR-6716-3p	-1.554	1.501	1.219e-03	1.679 e-02
hsa-miR-144-5p	-2.106	-0.822	1.414e-03	1.910 e-02
hsa-miR-3677-3p	1.400	2.773	1.426e-03	1.910 e-02
hsa-miR-615-3p	3.246	7.932	1.502e-03	1.983 e-02
hsa-miR-423-3p	-1.340	11.372	1.523e-03	1.984 e-02
hsa-miR-1910-5p	-1.637	0.666	1.549e-03	1.991 e-02
hsa-miR-5683	3.181	-0.348	1.585e-03	2.011 e-02
hsa-miR-605-5p	1.655	0.935	1.665e-03	2.085 e-02
hsa-miR-3152-5p	2.936	3.544	1.835e-03	2.242 e-02
hsa-miR-196b-5p	-1.531	9.238	1.842e-03	2.242 e-02
hsa-miR-573	2.556	-1.061	1.861e-03	2.242 e-02
hsa-miR-130b-5p	-1.310	5.899	1.200e-03	2.380 e-02
hsa-miR-4803	2.149	-0.995	2.079e-03	2.444 e-02
hsa-miR-2116-3p	-1.470	1.469	2.124e-03	2.467 e-02
hsa-miR-33b-5p	-1.445	3.862	2.184e-03	2.506 e-02
hsa-miR-668-3p	-1.428	2.642	2.447e-03	2.775 e-02
hsa-miR-6750-3p	-2.200	-1.509	2.513e-03	2.817 e-02
hsa-miR-3152-3p	2.117	-1.325	2.664e-03	2.952 e-02
hsa-miR-4731-3p	2.260	-1.141	2.792e-03	3.059 e-02
hsa-miR-423-5p	-1.194	9.065	3.026e-03	3.278 e-02
hsa-miR-106b-3p	-1.241	6.007	3.327e-03	3.531 e-02
hsa-miR-153-3p	-2.482	2.443	3.350e-03	3.531 e-02
hsa-miR-34c-3p	1.376	3.152	3.370e-03	3.531 e-02
hsa-miR-365a-3p	1.176	7.691	3.514e-03	3.642 e-02
hsa-miR-6824-3p	-1.627	-0.293	3.600e-03	3.691 e-02
hsa-miR-29c-5p	-1.312	3.440	3.797e-03	3.853 e-02
hsa-miR-3157-5p	2.022	-0.851	3.860e-03	3.876 e-02

hsa-miR-656-5p	1.817	-0.913	4.191e-03	4.165 e-02
hsa-miR-3613-3p	1.315	2.188	4.278e-03	4.208 e-02
hsa-miR-1254	-1.345	1.726	4.758e-03	4.633 e-02

Table with the 148 significant miRNAs in XPC vs Healthy

Gene_id	logFC	logCPM	PValue	FDR
hsa-miR-10a-5p	10.442	16.236	4.086e-74	3.641e-71
hsa-miR-10a-3p	9.483	6.258	1.155e-44	5.144e-42
hsa-miR-1260a	7.239	3.462	1.617e-23	4.803e-21
hsa-miR-6087	6.211	2.099	2.032e-14	4.526e-12
hsa-miR-129-5p	3.999	3.675	2.297e-13	4.093e-11
hsa-miR-129-2-3p	3.639	2.870	4.422e-13	6.567e-11
hsa-miR-195-5p	3.145	7.139	2.842e-10	3.618e-08
hsa-miR-1295a	-5.668	-0.282	2.501e-09	2.785e-07
hsa-miR-1291	5.148	2.109	1.005e-08	9.951e-07
hsa-miR-6842-3p	-3.317	1.830	3.823e-08	3.186e-06
hsa-miR-10b-5p	5.152	15.644	3.934e-08	3.186e-06
hsa-miR-155-3p	4.731	-0.526	5.253e-08	3.900e-06
hsa-miR-1468-5p	-3.251	5.261	1.197e-07	8.207e-06
hsa-miR-137	2.870	5.524	2.004e-07	1.266e-05
hsa-let-7i-3p	2.504	6.472	2.300e-07	1.266e-05
hsa-miR-5683	5.262	1.388	2.316e-07	1.266e-05
hsa-miR-556-5p	-4.168	-0.377	2.493e-07	1.266e-05
hsa-miR-146a-5p	3.763	7.167	2.557e-07	1.266e-05
hsa-miR-3607-5p	3.859	0.432	6.783e-07	3.181e-05
hsa-miR-148a-3p	2.188	12.377	8.616e-07	3.838e-05
hsa-miR-147b	3.088	2.324	1.132e-06	4.802e-05
hsa-miR-548ab	-2.636	1.483	1.697e-06	6.873e-05
hsa-miR-877-5p	-2.057	5.661	3.409e-06	1.321e-04
hsa-miR-4454	2.278	3.991	4.725e-06	1.754e-04
hsa-miR-10b-3p	4.989	4.256	4.996e-06	1.781e-04
hsa-miR-100-3p	1.974	8.221	5.690e-06	1.950e-04

hsa-miR-195-3p	2.365	3.304	6.012e-06	1.984e-04
hsa-miR-3605-3p	-1.940	3.106	1.082e-05	3.442e-04
hsa-miR-3117-3p	2.637	4.832	1.236e-05	3.798e-04
hsa-miR-19b-1-5p	2.456	0.334	1.423e-05	4.221e-04
hsa-miR-142-5p	2.802	2.358	1.468e-05	4.221e-04
hsa-miR-483-5p	-4.169	2.298	2.110e-05	5.876e-04
hsa-miR-1247-5p	-4.264	0.381	2.401e-05	6.481e-04
hsa-miR-7974	2.170	3.551	2.798e-05	7.331e-04
hsa-miR-146a-3p	6.520	0.908	3.530e-05	8.987e-04
hsa-miR-15a-3p	2.503	-0.058	3.795e-05	9.392e-04
hsa-miR-3615	-1.724	6.436	3.980e-05	9.585e-04
hsa-miR-4524a-5p	2.469	0.042	4.376e-05	1.026e-03
hsa-miR-3184-3p	-1.577	9.179	4.560e-05	1.042e-03
hsa-miR-129-1-3p	5.018	-0.896	4.733e-05	1.054e-03
hsa-miR-3614-5p	4.423	1.783	5.032e-05	1.078e-03
hsa-miR-155-5p	1.753	9.458	5.080e-05	1.078e-03
hsa-miR-4488	-3.355	-0.471	5.820e-05	1.206e-03
hsa-miR-4473	-2.102	0.712	6.801e-05	1.377e-03
hsa-miR-891a-5p	7.110	-0.711	7.145e-05	1.415e-03
hsa-miR-183-5p	3.802	3.166	7.650e-05	1.482e-03
hsa-miR-6507-5p	-6.470	-1.239	8.237e-05	1.561e-03
hsa-miR-624-5p	2.423	-0.209	9.000e-05	1.671e-03
hsa-miR-561-5p	-1.840	5.899	9.521e-05	1.731e-03
hsa-miR-4517	2.201	-0.295	1.018e-04	1.814e-03
hsa-miR-196b-5p	-1.569	9.180	1.115e-04	1.948e-03
hsa-miR-142-3p	2.619	0.620	1.261e-04	2.160e-03
hsa-miR-454-5p	-1.637	3.014	1.286e-04	2.162e-03
hsa-miR-1908-3p	-2.026	0.945	1.415e-04	2.334e-03
hsa-miR-668-3p	-1.885	2.501	1.965e-04	3.183e-03
hsa-miR-615-3p	3.474	8.101	2.180e-04	3.468e-03
hsa-miR-3529-3p	1.677	3.325	2.483e-04	3.882e-03
hsa-miR-615-5p	3.663	3.620	2.771e-04	4.257e-03
hsa-miR-21-3p	1.419	12.740	3.466e-04	5.234e-03

hsa-miR-4773	1.970	-0.0757	3.713e-04	5.513e-03
hsa-miR-6737-3p	-2.056	0.219	4.175e-04	6.098e-03
hsa-miR-98-3p	-1.477	4.517	4.295e-04	6.173e-03
hsa-miR-146b-3p	-2.345	5.170	4.685e-04	6.625e-03
hsa-miR-874-5p	-1.858	1.408	4.786e-04	6.654e-03
hsa-miR-326	-1.874	3.356	4.891e-04	6.654e-03
hsa-miR-31-3p	1.421	6.942	4.929e-04	6.654e-03
hsa-miR-34c-3p	1.744	3.375	5.020e-04	6.676e-03
hsa-miR-20a-5p	1.370	8.011	5.965e-04	7.816e-03
hsa-miR-487b-3p	-1.321	7.471	6.156e-04	7.933e-03
hsa-miR-486-3p	-1.808	3.338	6.257e-04	7.933e-03
hsa-miR-887-3p	-1.518	7.268	6.321e-04	7.933e-03
hsa-miR-3613-3p	1.528	2.291	7.211e-04	8.924e-03
hsa-miR-1268a	1.681	2.677	7.725e-04	9.428e-03
hsa-miR-149-5p	-1.478	8.066	8.263e-04	9.949e-03
hsa-miR-19a-5p	2.202	-0.506	9.320e-04	1.107e-02
hsa-miR-549a	-1.617	6.501	1.017e-03	1.192e-02
hsa-miR-431-3p	-1.369	3.480	1.090e-03	1.261e-02
hsa-miR-4767	2.370	-0.080	1.157e-03	1.322e-02
hsa-miR-6868-3p	-2.539	-0.829	1.183e-03	1.334e-02
hsa-miR-29b-2-5p	-1.709	1.138	1.226e-03	1.358e-02
hsa-miR-3613-5p	1.427	4.489	1.235e-03	1.358e-02
hsa-miR-1299	2.329	1.775	1.286e-03	1.398e-02
hsa-miR-6747-3p	-2.775	-1.139	1.305e-03	1.401e-02
hsa-miR-328-3p	-1.237	6.010	1.607e-03	1.704e-02
hsa-miR-16-1-3p	1.630	0.855	1.719e-03	1.802e-02
hsa-miR-424-5p	1.460	10.492	1.743e-03	1.806e-02
hsa-miR-1233-3p	-3.758	-1.677	1.870e-03	1.915e-02
hsa-miR-2116-3p	-1.627	1.414	1.915e-03	1.939e-02
hsa-miR-128-3p	-1.219	8.496	1.999e-03	2.001e-02
hsa-miR-585-5p	1.921	2.799	2.038e-03	2.018e-02
hsa-miR-944	6.216	-1.562	2.108e-03	2.062e-02
hsa-miR-483-3p	-2.997	2.372	2.129e-03	2.062e-02

hsa-miR-5701	-4.472	-1.201	2.164e-03	2.074e-02
hsa-miR-548h-3p	1.768	-0.383	2.236e-03	2.120e-02
hsa-miR-3194-5p	3.070	-1.733	2.352e-03	2.195e-02
hsa-miR-503-3p	1.662	2.792	2.365e-03	2.195e-02
hsa-miR-584-5p	-1.178	5.978	2.488e-03	2.263e-02
hsa-miR-1270	1.620	0.654	2.498e-03	2.263e-02
hsa-miR-4524a-3p	2.686	-1.151	2.515e-03	2.263e-02
hsa-miR-6787-3p	-2.832	-1.038	2.566e-03	2.287e-02
hsa-miR-548t-3p	1.652	0.229	2.616e-03	2.308e-02
hsa-miR-181a-5p	-1.479	13.887	2.711e-03	2.368e-02
hsa-miR-486-5p	-1.578	9.533	2.776e-03	2.340e-02
hsa-miR-1296-3p	1.947	-0.787	2.800e-03	2.340e-02
hsa-miR-744-5p	-1.171	7.878	2.866e-03	2.420 e-02
hsa-miR-431-5p	-1.139	7.452	2.920e-03	2.420e-02
hsa-miR-133a-3p	-2.232	1.544	2.928e-03	2.420e-02
hsa-miR-423-3p	-1.118	11.388	2.933e-03	2.420e-02
hsa-miR-548z	1.907	0.105	3.125e-03	2.554e-02
hsa-miR-450a-1-3p	1.921	-0.382	3.414e-03	2.765e-02
hsa-miR-629-3p	1.511	0.438	3.446e-03	2.766e-02
hsa-miR-485-5p	-1.145	5.607	3.573e-03	2.842e-02
hsa-miR-1273g-3p	1.661	0.763	3.631e-03	2.863e-02
hsa-miR-598-3p	-1.566	3.942	3.663e-03	2.863e-02
hsa-miR-7641	1.958	1.882	3.854e-03	2.986e-02
hsa-miR-543	-1.128	5.442	4.052e-03	3.112e-02
hsa-miR-181b-5p	-1.385	10.516	4.124e-03	3.140e-02
hsa-miR-6720-3p	-1.790	0.367	4.215e-03	3.183e-02
hsa-miR-7706	-1.129	5.186	4.336e-03	3.237e-02
hsa-miR-1252-5p	-3.811	-1.081	4.379e-03	3.237e-02
hsa-miR-140-5p	1.180	6.202	4.396e-03	3.237e-02
hsa-miR-221-5p	1.091	9.758	4.487e-03	3.277e-02
hsa-miR-548am-5p	1.678	0.846	4.563e-03	3.306e-02
hsa-miR-34b-5p	1.315	5.898	4.635e-03	3.331 e-02
hsa-miR-148a-5p	1.530	6.224	5.180e-03	3.692 e-02

hsa-miR-27a-5p	1.102	6.044	5.315e-03	3.758 e-02
hsa-miR-708-3p	1.873	5.737	5.375e-03	3.771 e-02
hsa-miR-3607-3p	-1.284	4.756	5.485e-03	3.818 e-02
hsa-miR-374a-5p	1.135	6.558	5.543e-03	3.829 e-02
hsa-let-7d-5p	-1.154	9.847	5.708e-03	3.891 e-02
hsa-miR-4707-3p	-2.406	-0.957	5.743e-03	3.891 e-02
hsa-miR-190a-5p	1.155	7.091	5.764e-03	3.891 e-02
hsa-miR-1254	-1.364	1.700	6.095e-03	4.060 e-02
hsa-miR-1972	1.512	0.377	6.106e-03	4.060 e-02
hsa-miR-4521	-1.795	2.843	6.193e-03	4.087 e-02
hsa-miR-432-5p	-1.086	6.448	6.298e-03	4.126 e-02
hsa-miR-18a-5p	1.135	4.900	6.548e-03	4.259 e-02
hsa-miR-30c-2-3p	-1.086	4.818	7.089e-03	4.554 e-02
hsa-miR-106a-5p	1.507	-0.307	7.105e-03	4.554 e-02
hsa-miR-655-5p	-1.560	-0.367	7.289e-03	4.639 e-02
hsa-miR-145-5p	1.205	10.385	7.524e-03	4.752 e-02
hsa-miR-1305	1.537	1.634	7.611e-03	4.752 e-02
hsa-miR-874-3p	-1.231	4.357	7.627e-03	4.752 e-02
hsa-miR-374a-3p	1.046	8.234	7.849e-03	4.799 e-02
hsa-miR-3120-3p	1.233	3.528	7.853e-03	4.799 e-02
hsa-miR-3157-5p	1.941	-0.923	7.863e-03	4.799 e-02
hsa-miR-675-5p	4.852	-1.038	8.154e-03	4.942 e-02
hsa-miR-1180-3p	-1.086	5.860	8.253e-03	4.968e-02