Cell Signalling and MicroRNAs: Regulation and Evolution

Mohab Helmy Elbishbishy

A thesis submitted for the degree of Doctor of Philosophy in Biological Sciences

School of Biological Sciences

University of Essex

October 2018

Summary

Cell signalling is among the most studied topics in modern biology, it allows cells to communicate with each other and their environment and thus, orchestrates the entire functioning of organisms. Despite being studied for more than a century, the field of cell signalling has recently evolved as cell signalling started to be looked at in the context of sophisticated networks that incorporates several loops and regulatory mechanisms. One of the major regulators of cell signalling are microRNAs. MicroRNAs are small non-coding sequences that regulate gene expression post-transcriptionally. As all other non-coding sequences, the significance of microRNAs has only been established recently. Today, microRNAs are known as one of the major regulators of gene expression that are able to target more than 60% of all human protein-coding genes as well as being involved in many diseases. While the role of microRNAs in regulating several components of signalling networks is known, our current knowledge lacks a systematic overview of the patterns of microRNA-mediated regulation of signalling networks. In this work, I provide a comprehensive analysis of the evolution of microRNA-mediated regulation through the incorporation of several bioinformatic tools. The results of this work show that microRNA-mediated regulation in signalling networks is particularly important on receptors. In addition, the evolutionary analysis shows that rodents and humans microRNA-mediated regulation of receptors have diverged significantly, limiting the validity of these animals models to study human disease related to cell signalling. Finally, the analysis of the precision of microRNA target prediction shows that multiple target sites close to each other significantly increase the chances of microRNA regulation. In summary, the main addition to knowledge provided by this work is a novel representation of a comprehensive evolutionary overview of microRNA regulation among different cell signalling networks in addition to tackling some of the issues currently present in microRNA target prediction.

Acknowledgments

First and foremost, I would like to express my appreciation and gratitude for my supervisor, Dr. Antonio Marco. Toni, your constant advice, support, and encouragement have meant the world to me these last few years. Thank you for inspiring me and thank you for allowing me to develop my own ideas. Thanks for always being there - be it for an informal chat or for a long scientific discussion with my endless questions- and thank you for not just being there, but for always going that one step beyond. I could not imagined having a better supervisor.

I am extremely thankful to my board member Professor Glyn Stanway for providing his insightful feedback throughout the length of my PhD and I would like to thank the School of Biological Sciences, University of Essex for the funding of my studentship.

Patrick, Tyler, Stuart, and Louis I can not thank you enough for your friendship and for providing me with the best work environment anyone can dream of. Beside the friendship and the environment, I am extremely grateful to Tyler and Patrick for their endless R help, to Stuart for once bringing my computer back from the dead, and to Louis for nothing.

I am very grateful to Dr. Andrea Hatlen and Dr. David Clark for their valuable help, feedback, and advice throughout the length of my PhD. I am also very thankful to my dear colleagues and friends Fionnuala, Myles, and Mila for their constant support, encouragement, and for believing in me.

Finally, I would like to thank my parents, the sole reason I made it this far.

List of abbreviations

3' UTR Three prime untranslated region

5' UTR Five prime untranslated region

CDS Coding sequences

CLIP Crosslinking and Immunoprecipitation

DNA Deoxyribonucleic acid

GTP Guanosine-5'-triphosphate

HITS-CLIP High Throughput Sequencing Crosslinking and Immunoprecipitation

ID-miRs immediate-down regulated

INDELs Insertions and deletions

IRES Internal ribosome entry site

mRNA messenger RNA

NICD Notch intracellular domain

NMD Nonsense mediated decay

PAR-CLIP Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation

piRNAs Piwi-interacting RNAs

pre-miRNA precursor microRNAs

pri-miRNAs primary-microRNAs

RISC RNA-induced silencing complex

RNA Ribonucleic acid

siRNAs Small interference RNAs

SNP Single Nucleotide Polymorphisms

TRBP TAR RNA binding protein

Table of contents

1. Introduction

At the dawn of the new millennium, the first draft of the human genome project was published and advances in genomic technologies made it possible to effectively sequence whole genomes of many species. The sequencing of the genomes of different eukaryotes has revealed two main surprises: a huge impact of transposable elements and the relatively-low number of protein coding sequences when compared to those that does not code proteins. For instance, 44% of human genome derives from transposable elements and 98.5% of the genome does not encode proteins (Mills et al., 2007; Beniaminov et al., 2008). Surprisingly, the number of protein coding genes is quietly steady among humans, *Caenorhabditis elegans* and sea sponges (Lindsay and Griffiths-Jones 2013). Later on, evolutionary studies have shown that even this 1.5% of coding genes is well conserved between species that are evolutionarily distant (Meader et al., 2010). As a result, it has became clear that the more rapidly evolving non-coding genome makes most of the difference. Only then, scientific interest has been attracted to the once-ignored non-coding genome resulting in a huge number of studies that led to our current vast knowledge.

1.1. The non-coding genome; not junk anymore

The major role played by non-coding sequences is regulating many processes by binding to protein, RNA or DNA through structure or sequence compatibility. Accordingly, it has become increasingly apparent that the non-protein-coding portion of the genome is of crucial functional importance for normal development and physiology and it is involved in many diseases (Esteller 2011). Majorly, non-coding RNAs are classified into two classes according to their sizes: (i) long

non-coding RNAs, with lengths more than 200 nucleotides, there are not big differences in size between long non-coding RNAs and coding mRNAs, (ii) small non-coding RNAs: these are non-coding sequences that are less than 200 nucleotides in length and usually processed from longer transcripts by RNA cleavage enzymes (Mattick and Makunin 2006). By definition, long non-coding RNAs do not have functional open reading frames unlike mRNAs. In addition, long non-coding RNAs are not as highly conserved and expressed as mRNAs. On the functional side, long non-coding RNAs, as other non-coding RNAs, mainly regulate gene expression. However, unlike other non-coding RNAs, long non-coding RNAs regulate gene expression at both the pre and post-transcriptional levels. On the other hand, based on their functions and sizes, small non-coding RNAs are further sub-classified into different classes. Small interference RNAs (siRNAs), these small RNAs range in size between 20 and 25 nucleotides and interfere with the expression of genes through RNA interference pathway that usually results in mRNAs being cleaved at the post-transcriptional level (Agrawal et al., 2003). Interestingly, siRNAs can be synthetically produced and introduced into cells to inhibit the growth and metastasis of oncogenes which open the door for further studies on using them therapeutically (Li et al., 2008). Piwi-interacting RNAs (piRNAs) form another class of small non coding RNAs that are able to silence gene expression at the post-transcriptional level through interacting with piwi proteins and forming RNAs-Proteins complexes (Thomson and Lin, 2009). However, piRNAs are only found in animals and are expressed from larger single-stranded precursors. PiRNAs roles are significant in specific processes such as DNA methylation, gametogenesis and neuroplasticity (Aravin et al., 2008). However, the best known class of small non-coding RNAs are microRNAs.

1.2. MicroRNAs

MicroRNAs are a class of small (~22 nt), single stranded, evolutionary conserved, non-protein-coding RNAs that are able to regulate gene expression post-transcriptionally. After binding to their target mRNA, microRNAs either degrade or suppress the translation of the targeted mRNA to proteins and thus, regulate gene expression (Kim and Nam 2006). The microRNAs story started in the year 1993 when two groups (Lee et al., 1993; Wightman et al., 1993) reported that a gene called *lin-14*, which regulates the development of the roundworm *C. elegans*, is targeted through complementarity by a small single-strand molecule that is encoded by the *lin-4* gene. Thus, they both reported that *lin-14* was regulated by a small sequence of RNA. However, the role of microRNAs has been under-appreciated for nearly a decade as they were thought to be limited to *C. elegans*. However, this all changed when the microRNA let-7 was discovered by Reinhart et al., (2000), again in *C. elegans*. Strikingly, this particular microRNA was found in multiple species, including humans (Pasquinelli et al., 2000). Since then, microRNAs started to capture the attention of biological community which resulted in a fair re-estimation of their roles in gene regulation. Such a time gap may have been just due to the fact that microRNAs were not known to be expressed by any other species but *C. elegans* in addition to the inefficiency of genomic and sequencing technologies at the time (Li and Kowdley 2012). After all, microRNAs have come a long way from being neglected to currently being recognised as one of the major regulators of gene expression among eukaryotes.

1.2.1. Evolutionary conservation

Evolutionary conservation was a milestone in microRNAs studies. It was their conservation among species that initially attracted the attention to them (Berezikov 2011). However, evolutionary conservation of microRNAs between different species appears to be much more complicated than it was initially thought. Interestingly, while some microRNAs are conserved among all species of the animal kingdom, each species has its own microRNAs that are exclusively expressed in it (Campo-Paysaa 2011). For instance, some microRNAs are conserved among species that are as far as humans and starlet sea anemone (Grimson et al., 2008) while others have been lost between primates and humans (Berezikov et al., 2006) (Figure 1.1).

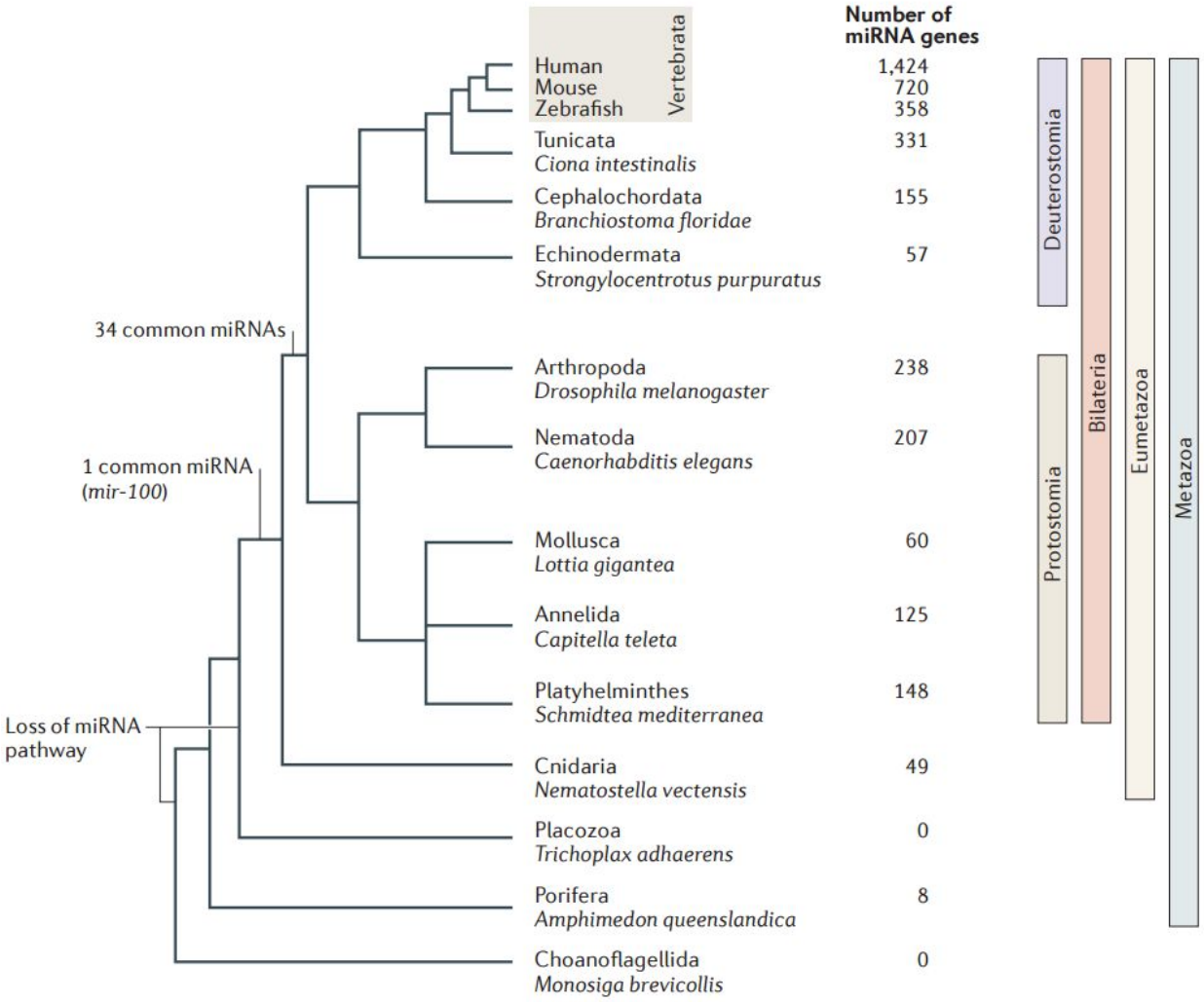

Figure 1.1 | MicroRNAs conservation across species: MicroRNAs seems to be heavily conserved across the whole animal kingdom with one microRNA (mir-100) present across the whole Bilateria and Cnidaria (adapted with permission from Berezikov 2011).

1.2.2. MicroRNA biogenesis

On its way to maturity, a microRNA molecule passes through a sophisticated multi-step biogenesis pathway before being able to target gene transcripts and consequently regulate gene expression (Figure 1.2). First, in the cell nucleus and by the aid of the enzymes RNA polymerase II or III, microRNAs are transcribed from the genome as a long transcript known as primary-microRNA (pri-miRNAs). Pri-miRNAs are then cleaved by a complex of the enzymes DGCR8 and Drosha to produce a shorter (~70bp) stem-loop structure known as precursor microRNAs (pre-miRNA) which consists of a terminal loop and two arms (5p and 3p). Pre-miRNAs are then exported to the cytoplasm by exportin-5, GTP, and Ran complex in a process known as nuclear exportation (Murchison and Hannon 2004).

Once in the cytoplasm, pre-miRNAs are further cleaved by the Dicer, TRBP and AGO enzymes complex that trims pre-miRNAs terminal loop producing only a ~22 nt duplex of two arms known as the miRNA/miRNA* duplex. Usually, only one of the two arms (dominant arm) will then be associated with a RNA-induced silencing complex (RISC) that result in a mature microRNA while the other arm is considered a waste-product and is usually degraded in a process known as arm-sorting. However, recent studies have proved that mature products from both arms of a precursor can be functional, resulting in two microRNAs with different functions (Marco et al., 2012). In some cases, microRNAs can be transcribed from genes that are physically adjacent hence they are collectively referred to as “microRNA clusters” that are known for having a synergetic effect in gene regulation (Yu et al., 2006) although this view has been recently challenged (Marco 2018).

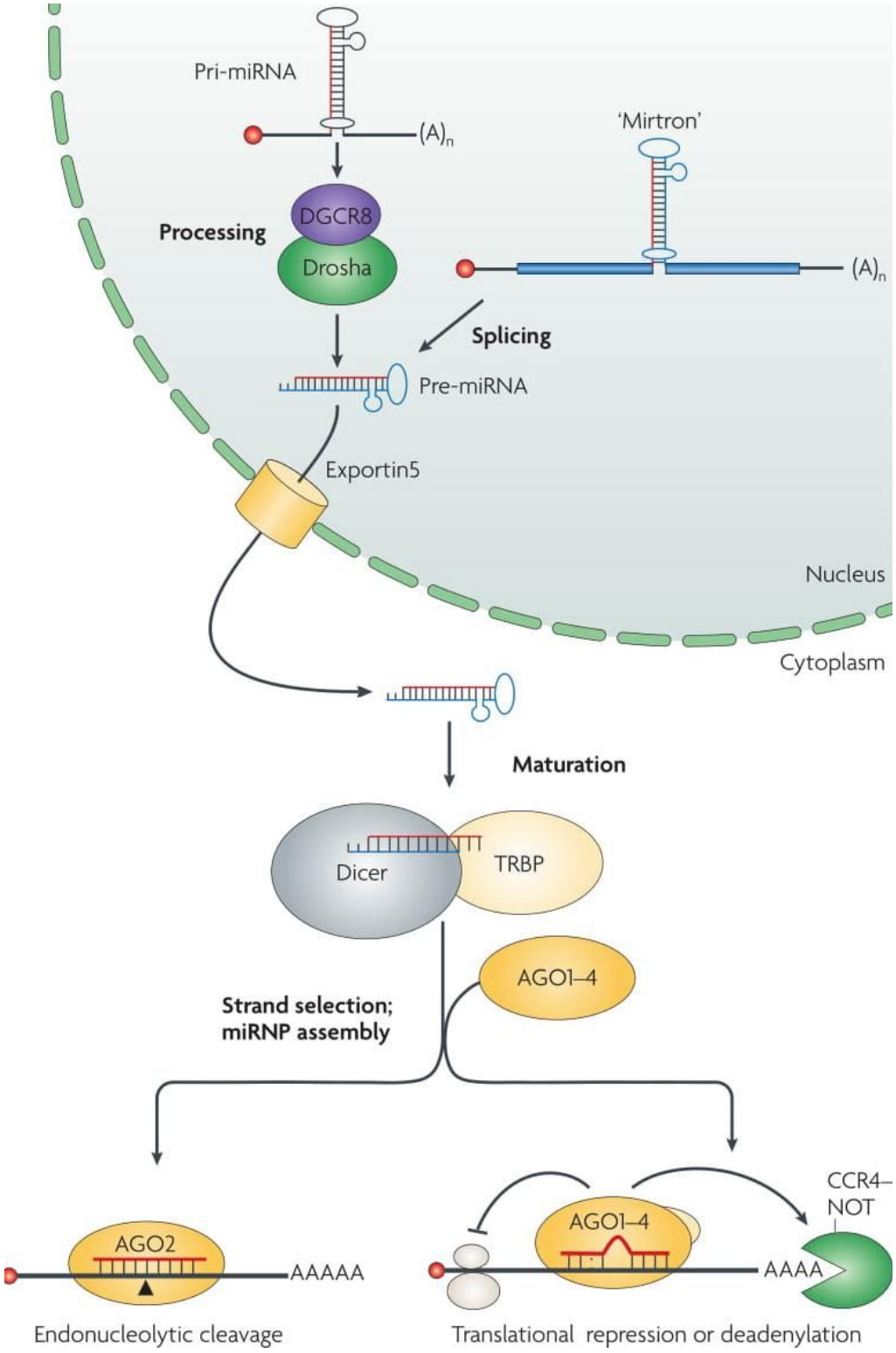

Figure 1.2 | microRNAs biogenesis: A complex multistep process that takes place across cell's nucleus and cytoplasm is mediated by many enzymes that convert the pri-miRNA to a mature microRNA (adapted with permission from Filipowicz et al., 2008).

1.2.3. MicroRNA-target interactions

Mature microRNAs, depending on sequence complementarity, mainly bind to the 3'UTR of a target mRNA. However, binding to other regions like 5'UTRs, coding DNA sequences, and promoters is also possible (Xu et al., 2014). MicroRNAs pairing with their targets relies on a certain region of the microRNA that is known as "seed region". Seed regions are nucleotides 2-7 of the microRNAs that are believed to be the most conserved regions of the microRNAs of all metazoan species. Since the average length of 3'UTR in human is nearly 1000 nucleotides (Sood et al., 2006), and given the relatively short length of the seed regions (only 6 nucleotides), it is not surprising that each human microRNA has the potential of having multiple target sites at most of human genes. However, the efficiency of the pairing between microRNAs and their targets is not always guaranteed. In fact, pairing between seed regions of microRNAs and their targets might not, on its own, be sufficient for influencing the targets expression levels. Hence, other sequences (beyond the seed) must be taken into consideration in predicting microRNAs target sites. Accordingly, seed regions are further classified as canonical (7mer and 8mer), marginal (6mer) or atypical (supplementary and compensatory) (Lim et al., 2003; Bartel 2009) (Figure 1.3).

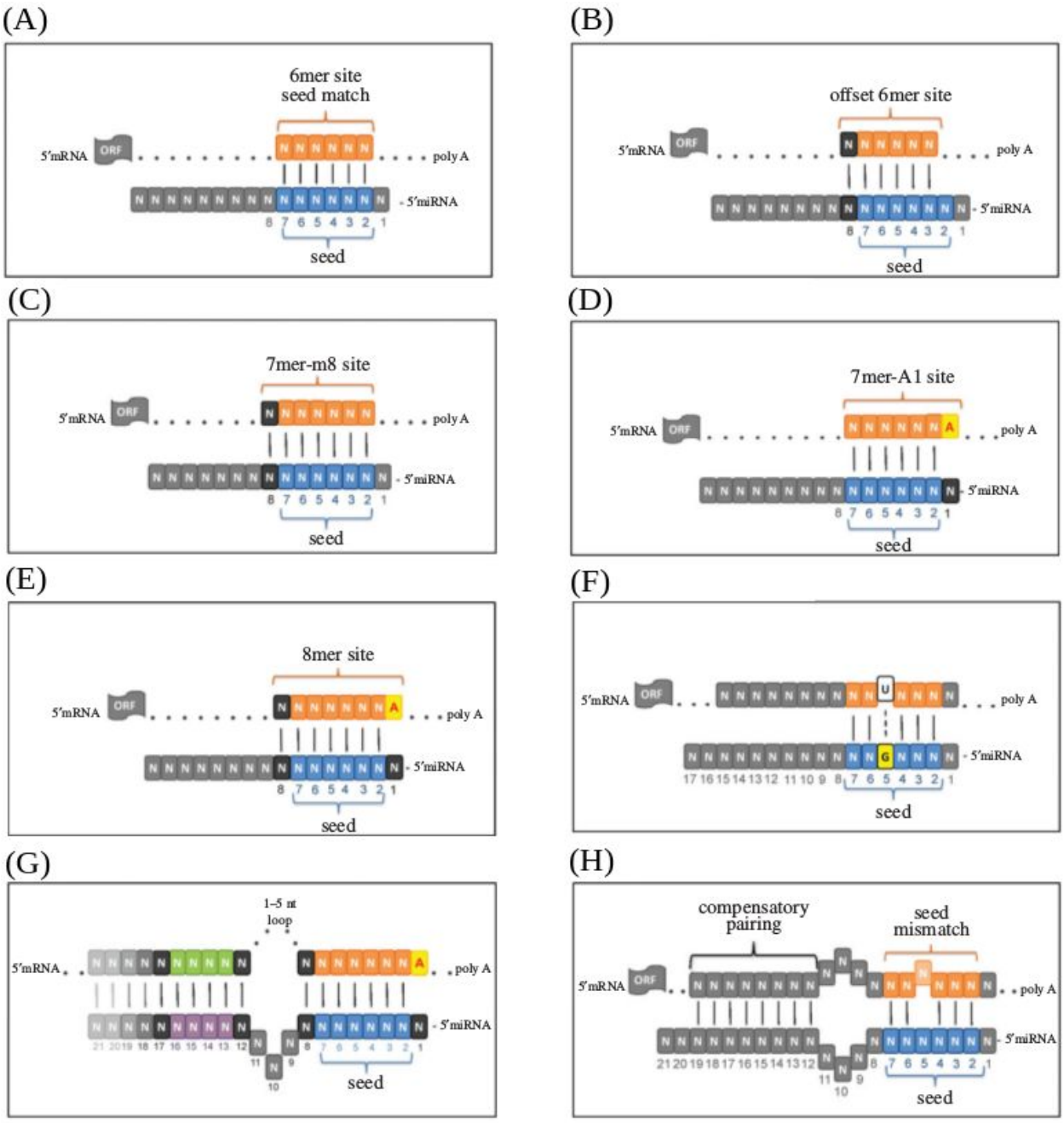

Figure 1.3 | Types of microRNA-target interactions: (A), 6mer: perfect complementarity between the seed region and the target 3'UTR. (B), 6mer-offset: complementarity between nucleotides 3-8 of the microRNA (instead of 2-7). (C), 7mer-m8: the 6mer pairing is complemented by a complementarity at the 8th position of the microRNA. (D), 7mer-A1: the 6mer pairing is complemented by an adenine adjacent to 1st nucleotide of the microRNA. (E), 8mer: the 6mer pairing is complemented by a complementarity at the 8th position and an adenine adjacent to 1st nucleotide of the microRNA. (F), Wobble pairing: A G:U base pairing between the seed region and the target. (G), Supplementary pairing: the 6mer, 7mer or 8mer pairing is complemented by pairing between positions 12 and (18-21) of the microRNA. (H), Compensatory pairing: the mismatch in the seed region is compensated by the pairing between 12 and (18-21) of the microRNA (adapted with permission from Moszyńska et al., 2017).

1.2.4. MicroRNA annotation

Due to size, structure and function similarities, distinguishing microRNAs from other small non-coding sequences is a challenging process (Ambros et al., 2003). The criteria for detecting a microRNA usually starts by detecting an 18-24 nucleotide sequence through sequencing, cloning or Northern blot, and c-hybridization of a size fractionated RNA sequence. This must be coupled with the presence of a 60-80 nucleotide fold back precursor that have such sequence as one arm of its hairpin. In addition, due to the fundamental role played by Dicer in microRNAs maturation, the high number of microRNAs precursor in organisms where Dicer functions are inhibited is a sign of a possible microRNA sequence. Finally, the phylogenetic conservation of the potential microRNA sequence is studied. As a result, a sequence should be an 18-24 nucleotide long, processed by Dicer and, from a phylogenetically conserved stem-loop structure to be called a microRNA (Kozomara and Griffiths-Jones 2013).

1.2.5. MicroRNAs nomenclature and online database

After the ever-growing interest in microRNAs, the necessity of creating a comprehensible database was proponed. That lead to the development of miRBase (formerly the microRNA registry). miRBase is the online repository of published data about microRNA sequences and their annotations. In addition, miRBase is responsible for giving unique names to all discovered microRNAs. Also, miRBase provides its data in forms that are easily readable and analysable for both humans and computers (Kozomara & Griffiths-Jones, 2013).

In addition, it was important to agree on a uniform naming scheme for all microRNAs, the agreed-upon naming system starts with a three letter genus-species code (e.g. hsa for *Homo sapiens*) followed by "mir" (or "miR" if referring to mature sequence). The previous suffix is followed by a unique number. This number can either be sequential (i.e. miR-7 can be the first microRNA discovered after miR-6) or in reference to a homology between the newly discovered microRNA and its human family member. In cases of paralogues, this number is followed by a dash and another number (e.g. hsa-mir-7-2) while closely related sequences are distinguished by the addition of a letter (hsa-miR-125a). Finally, if the same hairpin structure of the microRNA gene gives rise to two different mature sequences from its opposite arms (3' and 5' arms), a further "-3p" or "-5p" is added to the name referring to the 3' arm or the 5' arm origin of the mature sequence respectively (e.g.hsa-miR-7-5p). If only one arm gives rise to a mature sequence, the other subordinate sequence is given an asterisk (e.g. hsa-miR-7-1*) (Ambros et al., 2003). However, this nomenclature has been replaced by "-3p" and "-5p" throughput in the latest version of miRBase.

1.2.6. Mechanisms of regulation by microRNAs

MicroRNAs exert their regulatory roles on gene expression at the translational (post-transcriptional) phase where they either destabilise or repress the translation of mRNAs. In contrast to the very fast widening of understanding of microRNAs importance and conservation, the understanding of the exact molecular mechanisms by which they function remained unclear for years. The agreement on a precise set of mechanisms of microRNA-mediated translation inhibition was always controversial due to the unanticipated diversity of such mechanisms (Wu

and Belasco 2008). The first instance of microRNAs ability to inhibit translation was reported in the year 1993, Wightman et al., highlighted the imbalance between the concentration of mRNAs and levels of the proteins they are translated to which led to the discovery of the ability of *lin-14* (the only first known microRNA) to represses the translation of mRNAs to proteins.

In eukaryotes, the process of RNA translation to proteins consists of three stages: initiation, elongation and termination. MicroRNAs usually bind to mRNAs at the initiation step of translation. The initiation step starts upon the recognition and binding of the eIF4F subunit of the translation initiation factor to the cap structure at the 5' terminal of the mRNA. Alternatively, initiation can happen as a result of the interaction between the RISC complex and the internal ribosome entry site (IRES). The initiation step is followed by elongation where several amino acids are transported by tRNA and finally, a protein is terminated. It is believed that microRNAs inhibition of translation can take place at the post-initiation levels as well. However, the exact mechanisms by which post-initiation inhibition takes place remains unclear (Filipowicz et al., 2008) (figure 1.4 (A-C)).

Based on sequence complementarity, degradation of target mRNAs by microRNAs is divided into two categories: (i) Site-specific cleavage (also known as RNA interference) which is restricted to the relatively rare cases where the complementarity is perfect and in animals is carried out only by Argonaute 2 (AGO2). This process ends up with two unprotected RNA fragments that are susceptible to endonuclease attack and thus mRNA decay (figure 1.4 (D)) (ii) Non-cleavage repression which usually results from the more-common imperfect

complementarity and is carried out by any of the four Argonautes. In this case, by the aid of the Ccr4/Not complex, the poly-A tail of the 3' end is removed and the mRNA is exposed to exonucleolytic digestion (Gu and Kay, 2010) (figure 1.4 (E)).

Depending on the strength of binding, expression levels of both microRNAs and their target mRNAs, microRNAs effects can vary from just fine tuning the expression level of their targets to almost complete elimination of the expression (Bartel and Chen 2004). Until very recently, it was believed that translation inhibition is the major role that microRNAs plays in animals gene regulation and that degradation is more of a plant-related process (Millar & Waterhouse 2005). However, recent studies shows that translation inhibition posses rather a minor effect in gene regulation as it is responsible for ~25% of successful microRNA-induced regulation of gene expression. In contrast, mRNA destabilization is the dominant way of microRNA-induced gene regulation (Eichhorn et al., 2014). In addition, the translational repression by microRNAs can in some cases be reversible (Bhattacharyya et al., 2006). Interestingly, in some cases microRNAs can in fact upregulate the expression of their target mRNAs. However this remains rather an exception to the downregulation rule (reviewed in Vasudevan 2012).

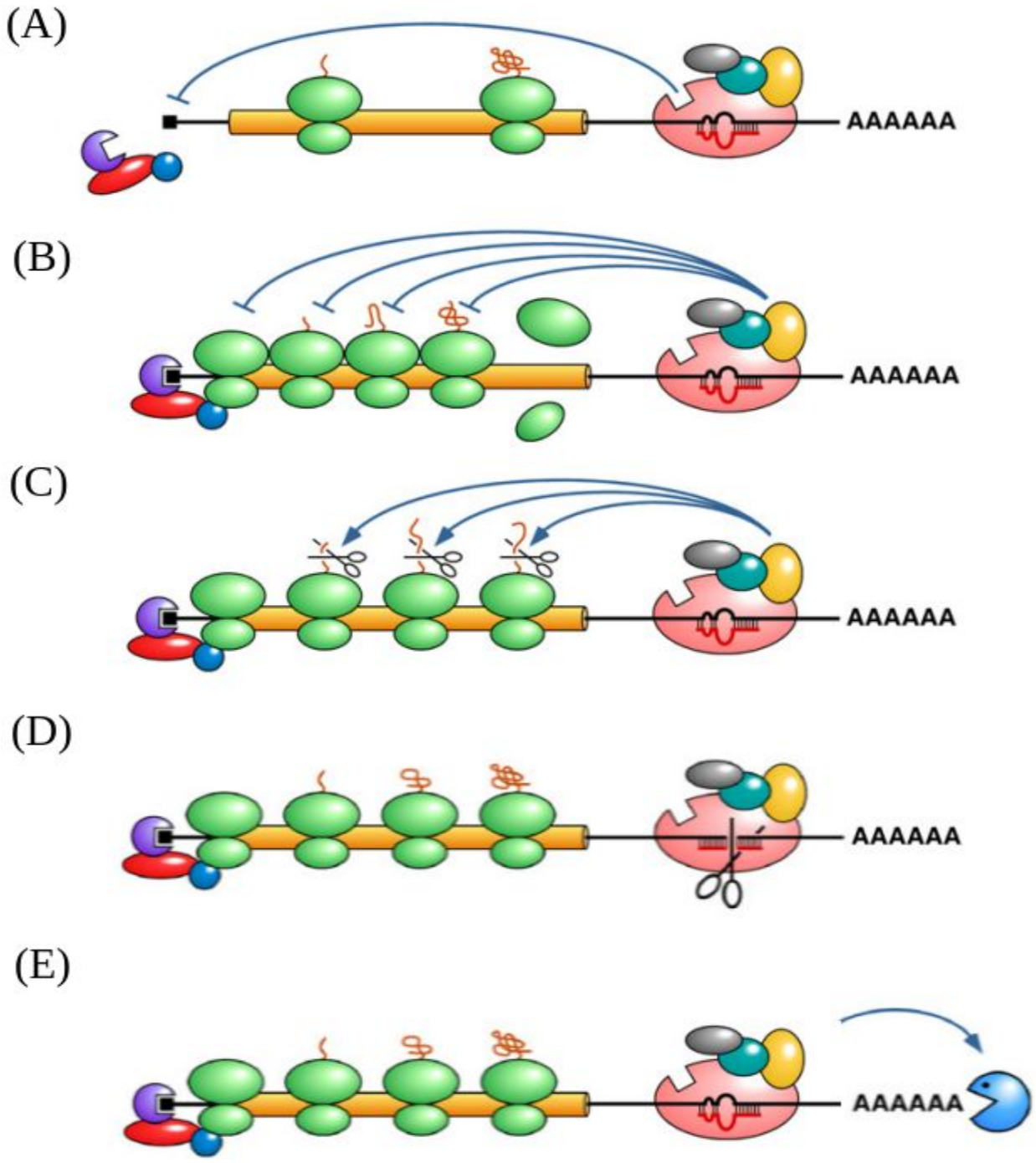

Figure 1.4 | Mechanisms of microRNAs regulation: (A) The RISC complex competes with eIF4E for the binding of the m7G cap resulting in translational repression at the initiation step. (B) Inhibition of elongation by inducing premature termination. (C) Inhibition of elongation by degrading polypeptides . (D) In case of full complementarity between the microRNA and the target, endonucleolytic cleavage of the mRNA takes place by the aid of Ago2 resulting in mRNA decay. (E) Poly-A tail of the mRNA being removed by the Ccr4/Not complex resulting in mRNA decay (adapted with permission from Wu and Belasco 2008).

1.2.7. MicroRNA functions and target prediction

Determining a microRNA function is in fact finding its possible targets and the role of those targets. MicroRNAs and their target mRNAs are engaged in a many to many relationship in which each microRNA can target several gene transcripts and each gene transcript is usually targeted by several microRNAs. For this instance, much research has been done in order to predict different targets to microRNAs which resulted in many software that use different algorithms in order to predict possible targets of microRNAs bioinformatically. On the other hand, many experimental approaches are used to confirm the results of bioinformatic predictions (Ameres and Zamore 2013).

Target prediction algorithms detect complementarity between the 3' UTR (sometimes, coding regions too) of a mRNA and the seed region of a microRNA (Zorc et al., 2012). However, predicting microRNAs targets in animals is usually more challenging than in plants, this is due to the fact that in plants a perfect complementarity is usually required between the whole microRNA sequence and their targets which is not the case in animals where complementarity is usually just limited to the seed region of the microRNA (Bentwich 2005). The pipeline used by most algorithms starts with identifying possible binding sites with accordance to base pairing pattern. This is usually followed by assessing the thermodynamic properties of the binding between the microRNA and its target (Maziere and Enright 2007). However, many interactions will be predicted with high probability of false positives. To eliminate this, comparative sequence analysis is usually done to assess the conservation of such interactions between species

across evolution (Maziere and Enright 2007). Finally, the possibility of having multiple target sites in the mRNA is taken in consideration as it boosts the chance of translation suppression as it allow coordination of multiple microRNAs in targeting a single mRNA (Stark et al., 2003). The previous pipeline is the basic methodology that is used in bioinformatic predictions of microRNA targets. However, each algorithm differs from the others by having extra features that allow more precise prediction of microRNA-mRNA interactions and thus different programs will predict different target sets. Therefore, a single approach to predict microRNA-targets interaction may not always be sufficient. In most cases a combination of different software can be used to predict possible targets then one of the experimental procedures are usually done to confirm such prediction (Kuhn et. al., 2008). However, the combination of several methods produces a high level of false negatives (Alexiou et al., 2009).

Any of the above algorithms will typically result in a large list of possible interactions between microRNAs and their target gene transcripts. However, this will always stay only an assumption as not all predictions are correct and not all predicted targets are functional. Thus, a way to reduce the incidence of false positives is to use the fact that closely related species will conserve functional microRNA target sites. Evolutionary conservation is a robust strategy that is used to reduce the number of false positives (Maragkakis et al., 2009). Specific target sites can be validated with specific assays (Kuhn et al., 2009). For this instant, different laboratory approaches like northern blot, miRNA microarray, transgenics and real-time PCR can be used to assess and validate the prediction as well as to determine the biological significance of the targeting (Watanabe et al., 2007).

Recently, a new class of software has been developed in order to enhance the reliability of prediction by adding extra features. Basically, most of these programs uses one or many of the already existing databases as their starting point. For instance, TarBase7.0 (Vlachos et al., 2015) is a database that allows a user to view only experimentally validated predictions by linking each predicted interaction to scientific studies documenting its validation. Likewise, miRecords (Xiao et al., 2009) uses the results of 11 different target prediction databases and links them to published experimental validations. MiRGator (Nam et al., 2009) provides more information about target predictions by linking gene expression, functional analysis and genome annotations to each prediction which allow more detailed assessing to each prediction. Finally, miRDB is the only interactive database for miRNA targets prediction; it allows its users to edit target prediction results according to their findings (Wong et al., 2014).

1.2.8. MicroRNA and diseases

The link between microRNAs and human disease was always obvious. Firstly, microRNAs are usually expressed in a tissue-specific pattern (Lagos-Quintana et al., 2002) Also, it is believed that defects in microRNA expression are often a hallmark of human diseases (Esteller 2011). The first evidence of the role of microRNAs in human disease was supplied by the result of Calin et al., (2002) when they detected events of down regulation or loss of two microRNAs (mir-15 and mir-16) in 68% of chronic lymphocytic leukaemia patients. In addition, many developmental process are believed to be regulated by microRNAs as stem cell division (Hatfield et al., 2005), cell metabolism, proliferation and apoptosis (Croce and Calin 2005), neural cell fate (Smirnova

et al., 2005), stress resistance (Ambros 2003), muscle differentiation (Chen et al., 2006), and brain morphogenesis (Giraldez et al., 2005). As a result, many pathological states such as Alzheimer (Wang et al., 2008), Parkinson (Nelson et al., 2008), arrhythmogenesis (Yang et al., 2007), myopathy (Eisenberg et al., 2007), psoriasis (Bostjancic and Glavac 2008), rheumatoid arthritis (Nakasa et al., 2008), schizophrenia (Beveridge et al., 2010), Downs Syndrome (Kuhn et al., 2010), and cardiovascular diseases (Fichtlscherer et al., 2011) are among the disease that microRNAs play roles in. Furthermore, microRNAs are particularly significant in cancer. "A rule rather than an exception" was how Jansson and Lund (2012) described aberrant expression of microRNAs in cancer cells. In fact, since their discovery, there was always a link between microRNAs and cancer as the first discovered microRNA have showed a regulatory effect on the division of cells in *C. elegans* which pointed to a presumed role in cancer (Lu et al., 2005). To help in identifying the role of microRNAs in different diseases, Li et al., (2013) manually collected and analysed literature data that describe the association between microRNAs and human disease and used them to build the human microRNA disease database (HMDD) which provides a web interface for researchers to browse, download and/or submit disease-related microRNA data.

1.3. Cell Signalling

The formation of complex cellular communication mechanisms was an inevitable step towards the evolution of multicellular organisms from their unicellular ancestors (Weinberg 2014). In fact, the proper functioning of organisms depends on the ability of their cells to develop and maintain proper means of communication between each other as well as their environments. Cells communicate by signals that are sent to and received by other cells, the environment or even the same cell (Alberts et al., 2002). However, different cells may react to different signals in different ways. Understanding the mechanisms by which cells communicate is essential in our understanding of how different organisms function.

The process of cell signalling starts by the synthesis of a signalling molecule (also known as ligand). In some cases, cell signalling may also be initiated by physical signals from the environment such as light or heat (Powar, 2010). In multicellular organisms, proteins, neurotransmitters, peptides, hormones, and steroids are among the various types of signalling molecules that can initiate cell signalling. A signalling cell usually secrets its signalling molecules by a process known as exocytosis into the extracellular space. In fact, signalling molecules can also be diffused through a cell's plasma membrane or interact with other members of the extracellular space while still bound to the surface of the signalling cell in a process known as juxtacrine signalling. However, this remains a contact dependant process that occurs only between neighbouring cells (Anklesaria et al., 1990). Depending on the distance between the cell that synthesize the signalling molecule and the cell that receives it, cell signalling can be

classified into two main types: (i) Endocrine signalling: the most common type of signalling in animals, were the signalling cell and the target cell are distant, the signaling molecule (usually a hormone) may have to travel through the circulatory system for relatively long distance before it reaches its distant targets. (ii) Paracrine signalling: where the signalling molecules act on nearby targets, the secreted molecule acts locally on targets. In addition, a cell can send signals to its own self. In some cases, a cell secrets ligands that can be recognized by and bound to its own receptors. Such process, known as autocrine signalling, usually happens in cells during development (Sporm and Roberts 1985). Despite the means of signal production and transmission, cells in multicellular organisms maintain a range of cell surface receptors that are responsive to extracellular stimuli. Such receptors, usually transmembrane ones, need to recognize and bind to the signalling molecule. However, in some rare cases, signalling molecules may need to make their ways through cell membranes to find an intracellular receptor. In addition, a signalling molecule can be transported by cellular extensions known as cytonemes (Ramírez-Weber & Kornberg 1999). Based on the type of signal and receptor (and the way they interact) their interaction usually activates a series of cascades that are known as signaling pathways that in turn activate (or inhibit the activation) of specific proteins which then decide cells actions or behaviour (Alberts et al., 2002).

1.3.1. Cell signalling and disease

While cell signalling is vital to maintain the normal physiological functions of an organism, it does not always happen in a perfect way. In fact, transduction of signals through the pathways occasionally fails resulting in various types of diseases. Depending on the type of error of

signalling, different types of diseases may occur. For instance, when insulin receptors fails to bind to insulin this usually results in type II diabetes (Pessin 2000). Also, excessive production of the signalling molecule glutamate after strokes may lead to a complete brain damage (Rothman 1986). Finally, cancer is significantly related to cell signalling malfunction. Cancer mainly results from the ability of cells to grow and divide independent on signals. Cancer cells will also need to be able to ignore the apoptotic (suicidal) signals that are produced when cells starts to grow uncontrollably (Yamazaki et al., 2005). Also, it is very common that cancer cells secrete growth factor signals to themselves and to other cells to grow more rapidly (Powar 2010). Unsurprisingly, to prevent diseases, cell signalling is tightly regulated by a wide range of regulators and mong such regulators are microRNAs (Chen 2015).

1.3.2. The evolution of the study of cell signalling

The importance of cell signalling was acknowledged since the beginning of the 19th century (Freeman & Gurdon 2002). However, for long time, the study of cell signalling relied on the identification of simple, straight and hierarchically-organised downstream cascades of signalling from receptors to enzymes and transcription factors. Fortunately, the substantial progress of biochemistry and the availability of high throughput technologies has led to the identification of an increasingly large number of cell signalling components as well as the diverse mechanisms made it clear that signalling pathways do not necessarily fall into such a linear fashion. Instead, cell signalling falls into pathways of a highly complex nature that is usually regulated by the presence of many forward and backward loops. Furthermore, in most cases, such pathways together form larger networks of signalling and not as previously believed are stand-alone

entities (Eungdamrong & Iyengar 2004). Cell signalling networks emerges when different pathways interact in a way that makes a single signalling component able to receive, process and pass-on signals in different ways. Such signalling molecules, known as interconnections, can act either as nodes or junctions where junctions act as signalling integrators and nodes act as signalling splitters (Jordan et al., 2000). In some cases, the very same signalling molecule can act as a node and a junction (Schmelzle & Hall 2000) (Figure 1.5). Given this complex nature, the use of mathematical and computational modelling was always necessary to explore the diverse mechanisms that underlie cell signalling. Thanks to mathematical and computational modeling, we are now able to model and simulate signalling networks with tens of components and regulatory loops that are impossible to predict and validate using the current experimental technologies (Eungdamrong & Iyengar 2004).

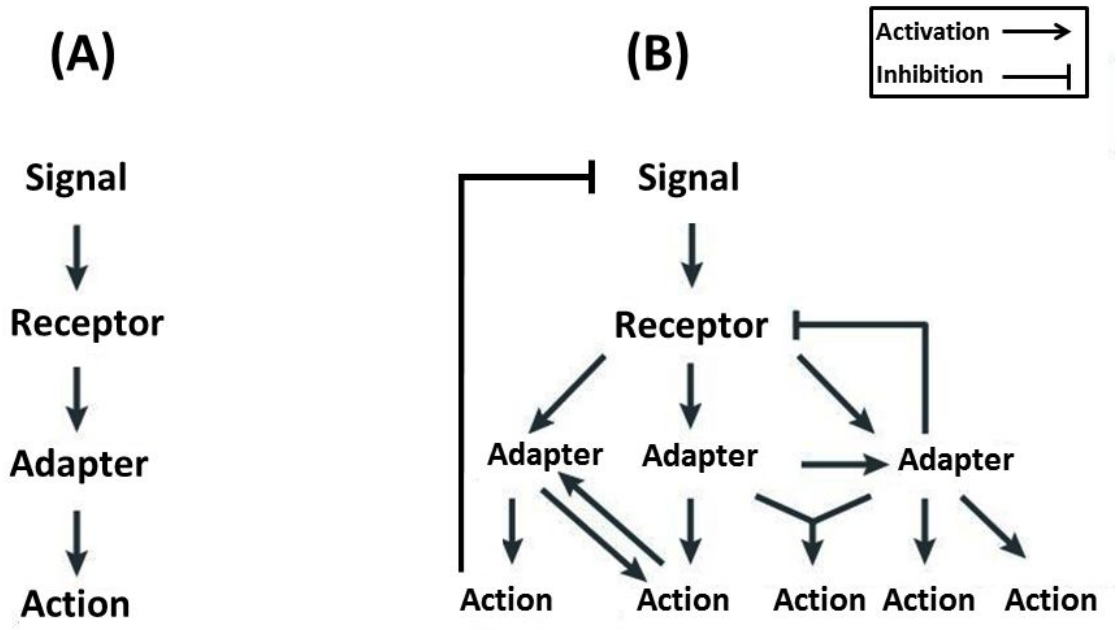

Figure 1.5 | The evolution of cell signalling understanding: (A) The initial thought of cell signalling as a linear process where the binding of a ligand to a receptor activates a cellular adaptor that results in a cellular action. (B) The complex, multi-step process of cell signalling where the activation of signaling is regulated by many forward and backward loops as well as cross talking that together, activate or inhibit various steps of the signalling process before a cellular action takes place (adapted with permission from Taniguchi et al., 2006).

One good representation of how our understanding of the field of cell signalling has rapidly evolved is the textbook "Molecular Biology of the Cell". In its first edition more than 30 years ago, the famous book only mentioned cell signalling briefly within a chapter named "Chemical signalling between cells" . Throughout the years, the growth of our knowledge of cell signalling has been reflected in the book through its different editions. Since its 2nd edition in 1989, Alberts et al., has devoted a complete chapter to cell signalling that has evolved over the years

being named “Cell signalling” in the second and third editions, “Cell communications” in the fourth edition, “Mechanisms of cell communications” in the fifth edition and finally back to “Cell signalling” in the current sixth edition of the book. However, since its establishment in the second edition, the cell signalling chapter almost doubled in size which reflects the importance as well as the complex nature of cell signalling (Alberts et al., 1983, 1989, 1994, 2002, 2008, 2014).

1.4. ErbB signalling network

ErbB stands out as the most studied receptor system in biomedical research (Samaga et al., 2009). Many factors contributes to ErbBs fame, first is it's very complex nature as the binding of an ErbB ligand to an ErbB receptor can activate different signalling pathways each containing several regulatory loops that can in turn activate or deactivate different cellular actions. However, beside its complex nature, ErbB is able to regulate several cellular process including cells differentiation, migration, survival/apoptosis and proliferation in almost all multicellular organisms and thus, ErbBs plays a profound role in regulating many diseases including cancer. Unsurprisingly, these regulatory roles made the ErbB network a prime target to many of the recently developed drugs (Yarden and Sliwkowski 2001).

The ErbB discovery dates back to the early 1950s when the epidermal growth factor (EGF) was firstly isolated in mouse by Stanley Cohen and his colleagues. While studying another growth factor, the nerve growth factor (NGF), Cohen treated a group of mice with salivary gland extract, these mice showed a significant acceleration in growth and development compared to their

controls. Only then, Cohen suggested that one substance is responsible for that, EGF (Cohen 1962). Later on, this discovery won Cohen (joined by his colleague Levi-Montalcini) a Nobel Prize for their joint discoveries of the epidermal and nerve growth factors, respectively. Cohen's discovery sparked the scientific interest to discover the biological and physiological roles that EGF plays. As a result, its was found that EGF is expressed in other human tissues and body fluids. In addition, the role of EGF in regulating many cellular process was established. All these led to the discovery of first EGF receptor (EGFR) which was firstly described as a transmembrane glycoprotein with a tyrosine kinase activity (Cohen 1982). The key components of the ErbB signalling network are the four tyrosine kinase receptors and the eleven epidermal growth factors ligands (Barberan et al., 2016). The receptors family consists of four structurally related members: *ErbB1* (*EGFR*/*HER1*), *ErbB2* (*HER2*/*neu*), *ErbB3* (*HER3*) and *ErbB4* (*HER4*) that are able to engage in homodimers or heterodimers. However, only ErbB1 and ErbB4 can be considered fully functioning ErbB receptors (i.e. they are both able to bind to ligands and initiate an intrinsic tyrosine-kinase activity). On one hand, *ErbB2* is known as the orphan receptor as its lacks the ability to bind to any ligand however it remains one of the favourite dimerization partners of its ErbB siblings. *ErbB3*, on the other hand, is able to bind ligands but it lacks the ability of initiating a tyrosine-kinase activity. However, it also can form dimers with any of the other ErbB receptors (Yarden and Sliwkowski 2001).

1.4.1. The conservation of ErbBs

A single ErbB receptor is expressed by roundworms and flies. In more complex organisms, mammals for instance, the number of expressed genes increases to four for the aim of

modularity, redundancy and therefore, robustness (Citri and Yarden 2006). In roundworms, the ErbB-like gene (*LET-23*) is responsible for regulating ectoderm development, development of male tail and development of vulvae (Moghal and Sternberg 2003). Likewise, the ErbB-like gene (*DER*) is expressed in the fruit fly *Drosophila melanogaster* and is responsible for regulating many processes mainly during embryogenesis (Shilo 2003). After comparing different amino acid sequences of ErbBs between vertebrates and invertebrates, Stein and Staros (2000) suggested that a process of gene duplication may have resulted in two different ancestral genes; *ErbB1*/*ErbB2* precursor and *ErbB3*/*ErbB4* precursor in vertebrates from the single ErbB gene of invertebrates. Subsequently, each of these precursors furtherly duplicated generating the four ErbBs that are expressed in vertebrates. While evolving, ErbBs have witnessed two main accumulated mutation events. One that prevents the *ErbB2* gene from binding to ligands (Burgess et al., 2003), the other inhibits the activation of the kinase domain of *ErbB3* after it binds to ligands (Guy et al., 1994). Those mutations in fact resulted in a complex signalling network and stopped the near-perfect linear configuration of the receptor-ligand pathways (Citri and Yarden 2006). Such a complex signalling network is believed to be responsible for the higher robustness of the system in vertebrates (Bray and Lay 1994). The characterization of an ErbB tyrosine kinase receptor relies on the identification of three components: two cysteine-rich clusters at a ligand-binding extracellular domain, a transmembrane domain, and cytoplasmic domain that exerts a tyrosine kinase activity. Likewise, all ErbB ligands contain an ErbB-like domain in their extracellular domain which is characterized by the presence of 6 cysteines motifs that enhance the binding to the receptors (Barberán et al., 2016).

1.4.2. The complexity of ErbB signalling

One of the major reasons behind the complexity of the ErbB signalling network is the ability of some G-protein-coupled-receptors (GPCRs) to activate some of the signalling cascades that are usually activated by ErbBs (Thomas et al., 2006). As a result, an ever wider range of signalling pathways can be activated by these events of cross talking which add more diversity to the possible cellular actions. For instance, ErbBs activation activates the MAPK signalling pathway which in turn initiates the transcription of ErbBs ligand which upon binding to receptors can activate MAPK signalling in a positive feedback loop (Yarden and Sliwkowski 2001). Normally, ErbB receptors localises and functions at the cell membrane. However, in some cases, ErbBs can translocate to the nucleus and exerts different functions from there. For instance, it has been proposed that nucleic *ErbB1* can act as a transcription factor (Wang and Hung 2009). In addition, ErbBs nuclear translocation has been linked with poor prognosis of human cancers (Wieduwilt and Moasser 2011). The exact mechanisms by which such relatively large proteins free themselves from the cell membrane endosomes and travel to the cell nucleus remain unclear. However, different mechanisms including endoplasmic-reticulum-mediated translocation have been proposed (Hu et al., 2006).

Another interesting feature of the ErbB network is its bow-tie like structure (Citri and Yarden 2006). A bow-tie is a common structure of biological systems that enhances their robustness. Simply, a bow-tie structure consists of two wide ends and a narrow (conserved) core. In signalling pathways, bow-tie structures are common. Usually, cells receptors are exposed to a high number of signals. However, between the binding to extracellular signals and the expression

of reactions (proteins) lies a “knot”. In which, a number of biochemical interactions form a core process that decides the fate of the received stimuli (Kitano 2004). Adding more to ErbBs complexity, the activation of the ErbB network through *ErbB1* and *ErbB2* in fact activates a subset machinery of interactions that decrease the abundance of a set of tumour suppressor microRNAs. By this means, oncogenic ErbBs protect themselves from being targeted by tumor suppressing microRNAs (Avraham et al., 2010).

1.4.3. ErbBs in human diseases

A wide range of human diseases stem from aberrant ErbB signaling. For instance, drugs inhibiting *ErbB2* activity have been proven to be cardiotoxic (Feldman et al., 2000). Also, the ability of *Mycobacterium leprae* to bind to and activate ErbB receptors causes leprosy (Tapinos et al., 2006) and continuous activation of *ErbB1* is a known cause of psoriasis (King et al., 1990). Finally, the dysregulation of the ErbB ligand *NRG1* is one characteristic of several disorders including Alzheimer's disease (Chaudhury et al., 2003), schizophrenia (Hahn et al., 2006) and multiple sclerosis (Brinkmann et al., 2008). Among all different diseases, cancer is one of the most studied diseases that are regulated by ErbBs. The first link between ErbBs and cancer dates back to the 1980s when the over-expression of *ErbB2* was reported in neoplastic rat cells (Schechter et al., 1984). Since then numerous studies have been carried out to uncover the roles played by ErbB signalling in cancer as it was always clear that a network that regulate cell growth, apoptosis and proliferation is indeed a network that can be linked to cancer. The established knowledge of the profound impact of ErbBs on human disease made them targets of many recently-developed drugs (Weiduwilt and Moasser 2011). Unsurprisingly, most of these

are anticancer ones. The first clinically used ErbB-targeting drug was trastuzumab. As a monoclonal antibody, trastuzumab interferes with the extracellular domain of ErbBs disrupting their ligand binding. However, the non-neglectable side effects of trastuzumab coupled with advances molecular biology techniques that allowed the electron-microscopic visualization of ErbBs dimers has led to the development of more advanced monoclonal antibodies such as pertuzumab, cetuximab and panitumumab. In addition to monoclonal antibodies, tyrosine kinase inhibitors (TKIs) are another class of ErbB targeting cancer therapies. Through binding to the ATP kinase domain, TKIs like erlotinib, gefitinib and lapatinib inhibit the enzymatic functions of ErbBs activation and thus block the initiation of downstream ErbB signaling. In addition, very recently, another TKI (neratinib) has been introduced with its promising ability to co-regulate different ErbBs at once. Unfortunately, despite the promising initial response rates resistance to ErbB-directed therapies is the rule rather than the exception (Wang and Greene 2008). Among the possible reasons behind such resistance is the ability of signalling pathways to maintain ErbB independent activations through cross talking with other signalling networks. Also, mutations at the TKIs binding sites of ErbBs may inhibit the sufficient binding of TKIs. For instance, Sergina et al (2007) suggests that a negative regulatory loop in the ErbB network enables ErbBs to escape TKIs targeting. Despite the initial promising response, resistance has resulted in an re-estimation of anti-ErbBs powers. Today, the therapeutic use of anti-ErbB agents is limited to combining them with chemotherapy.

1.5. Toll-like signalling network

The Toll-like signalling network is involved in one of the most critical process in all living organisms which is immunity. Throughout their lives, all living organisms will be exposed to a variety of pathogens within their environments that they will have to protect themselves from (Akira and Takeda 2004). The family of Toll-like receptors (10 receptors in humans) is able to recognize and bind to different pathogens-related molecules that then activates a cascade of intracellular interactions which in turn result in an immune response. In particular, Toll-like signaling is associated with one of the two classes of immune response that is innate immunity (the other being acquired immunity). As its name suggest, innate immunity is the first immune response taking place once a pathogen is detected (Akira 2006). Thus, Toll-like signalling acts as the front line defense mechanism of a host to a given pathogen (Medzhitov 2001).

The name "Toll-like" stems from the referral to the previously discovered *Toll* gene in *Drosophila* which is involved in embryogenesis (Hashimoto et al., 1988). Later on, it was shown that *Toll* plays an important role in initiating an immune response in *Drosophila* against fungal infections (Lemaitre et al., 1996). Such observation have sparked an interest in exploring the gene and subsequently, similar homologous sequences of the *Drosophila Toll* gene have been found in several species including humans and where then given the name Toll-like (Akira and Takedo 2004).

Toll-like receptors mainly respond to one of two signals, either a pathogen-related compound or to a molecule produced endogenously by the host as response to its infection by a pathogen. Pathogen-related compounds that are usually referred to as PAMPs (pathogen associated

molecular patterns) can be proteins, microbial lipids, carbohydrates or even, a nucleic acids of the pathogen. Out of the 12 Toll-like receptors that have been so far identified in mammals, some of them (*TLR1*, *TLR2*, *TLR4*, *TLR5*, *TLR6*, and *TLR11*) are known to localize to the cell membrane while others (*TLR3*, *TLR7*, *TLR8*, and *TLR9*) act intracellularly by localizing into the endosomes and thus they are known as Toll-like endosomal receptors (Lin et al., 2012). However, in human, only *TLR1* to *TLR10* are expressed while two extra Toll-like receptors are expressed in mouse (*TLR11* and *TLR13*). While the existence of *TLR10* is recognized, it has for long been known as the only Toll-like receptor with no biological function. However, recently, a study has suggest an inhibitory role of *TLR10* in immune response (Oosting et al., 2014). Beyond the recognition of (and binding to) a ligand by a toll-like receptor, it is up to one of two main adaptors to integrate the activation of the Toll-like receptors. The first and most dominant route is through the myeloid differentiation factor (*MyD88*) pathway where several intracellular pathways are activated. The other less-dominant way of signal transduction is known as MyD88-independent pathway in which, other adaptors like *TRAM*, *TRIF* or *TIRAP* and activate different pathways. In fact, only activations of Toll-like signalling through *TLR3* and *TLR4* are associated with the MyD88-independent pathway (Lin et al., 2012). Finally, the activation of Toll-like receptors and the initiation of signalling cascades by the adaptors usually ends in an increase of the expression of several defense molecules such as proinflammatory cytokines (like, tumor necrosis factor and interleukins) and costimulatory molecules (like interferon) depending on the pathways activated by the adapters (MyD88-dependant pathway or MyD88-independent one) (Borad et al., 2007). Finally, the expression of such molecules result in the activation of

macrophages and neutrophils in the aim of killing the pathogens detected (Takeda and Akira 2005) (Figure 1.6).

1.5.1. Conservation of the Toll-like network

The bioinformatic studies done on Toll-like receptors shows no presence of Toll-like orthologues in neither prokaryotes nor fungi. In plants, however, few receptors have been related to the animals Toll-like ones based on the fact that they serve similar functions and their sequences share some similarities . However, given the low level of sequences similarities and the fact that these receptors activate totally different signalling pathways when compared to those activated by animal Toll-like receptors indicates that the sequences detected in plants are not orthologues of the animals Toll-like receptors but are rather plants-specific receptors that preserved functions similar to Toll-like receptors through convergent evolution (Ausubel 2005). As a result, the origin of Toll-like receptors appears to belong to the animal kingdom. Within the animal kingdom, the conservation of Toll-like receptors varies greatly between several lineages. First, the lack of any Toll-like receptors in the genome of a sponges species (*Amphimedon queenslandica*) coupled with the presence of typical Toll-like receptors in the cnidarian species *Nematostella vectensis* and several bilateria species, suggests that the precise origin of Toll-like receptors goes back to the last common ancestor of all eumetazoans (all animals except sponges and Placozoa) (Voogdt and van Putten 2016). While Toll-like receptors are conserved among almost all animals, the number Toll-like genes varies largely between species. For instance, while humans conserve 10 different Toll-Like receptors, the number is expanded to 13 in rodents, 21 in most amphibians, and up to 222 in the purple sea urchin (Huang 2008). Such a

massive expansion in the number of Toll-Like genes might stem from the lack of an adaptive immune system in some species so the reliance is massively on innate immunity which is majorly activated by Toll-like receptors (Ward and Rosenthal 2014).

1.5.2. Toll-like signalling in human disease

Based on their fundamental role in initiating immunity, mutations (or alterations) among Toll-like genes have been heavily linked to a number of infectious diseases. For instance, mutations among Toll-like receptors are associated with higher susceptibility of leprosy (Kang et al., 2002), asthma (Qian et al., 2010), malaria (Mockenhaupt et al., 2006), and tuberculosis (Ben-Ali et al., 2004). However, the pleiotropic nature of Toll-like signalling makes them involved in a wide range of diseases. For instance, recent studies have shown an important role of *TLR3* in the regulation of several liver diseases including hepatitis (Yin and Gao 2010) and of *TLR9* in renal disease (Bossaller et al., 2016). In addition, a common polymorphism in *TLR4* (D299G) is related to a wide range of disorders including diabetes (Weyrich et al., 2010), asthma and atopy (Raby et al., 2002), and ulcerative colitis (Torok et al., 2004). However, the same polymorphism is linked with a reduced risk of atherosclerosis (Kiechl et al., 2002) and rheumatoid arthritis (Radstake et al., 2004).

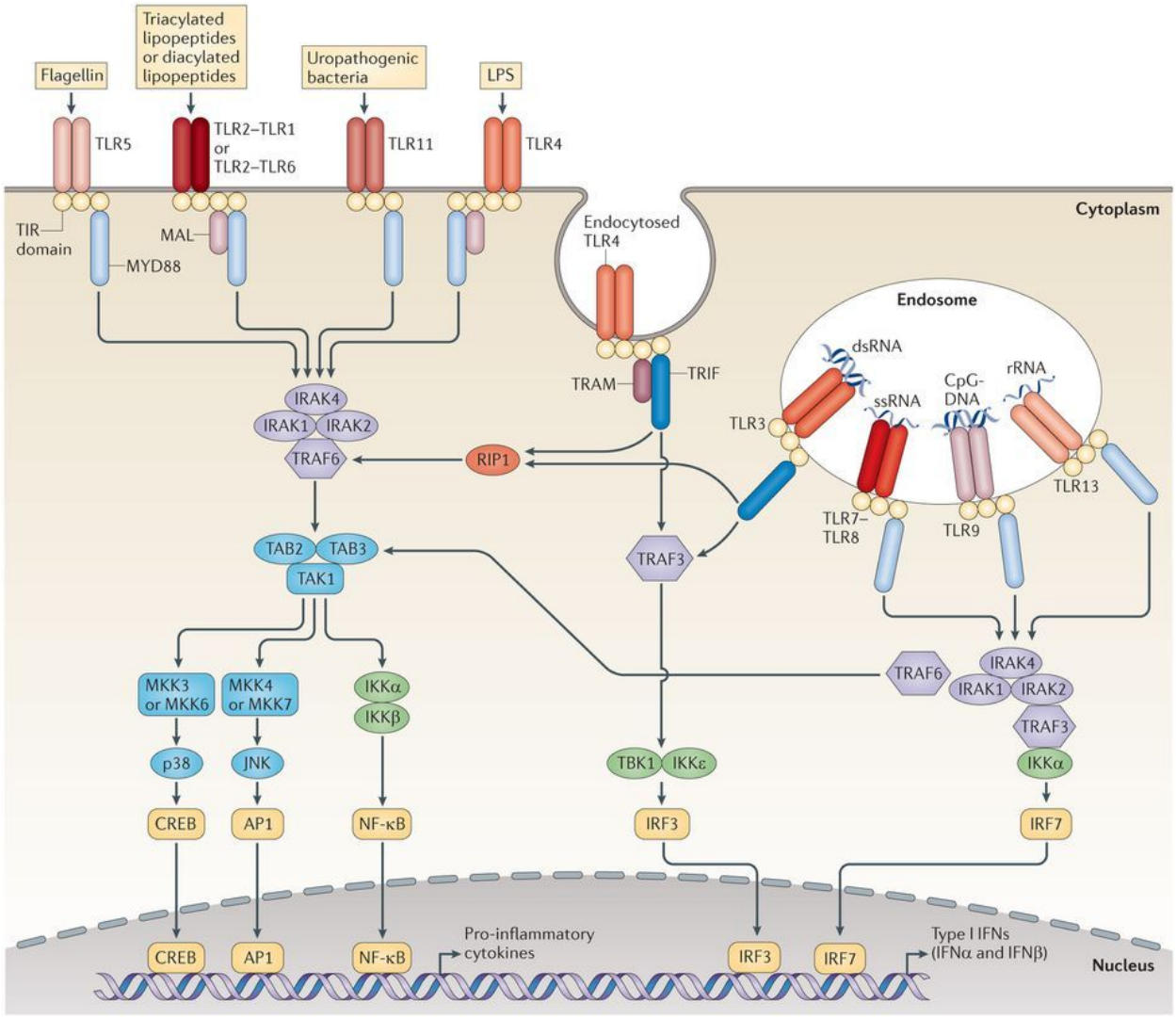

Figure 1.6 | The Toll-like signalling network: The signaling is initiated upon the detection of pathogens (or pathogen related molecules) by one of the 11 Toll-like receptors either on the cell surface (*TLR1, TLR2, TLR4, TLR5, TLR6, and TLR11*) or in the endosome (*TLR3, TLR7, TLR8, TLR9, and TLR13*). Consequently, the signal is transmitted via one of the two adapting pathways (MyD88 dependant and MyD88 independent). Finally, the signal reached the cell nucleus where it interacts with transcription factors that usually results in cellular actions. (adapted with permission from O'Neill et al., 2013).

1.6. Notch signalling network

At the dawn of genetics, in the year 1919 the observation of a mild wing deformation in *Drosophila* has opened the door to a long history of research that have finally led to our current understanding of one of the most widely studied signalling networks, the Notch signalling network (Mirandola et al., 2009). Notch signalling is one of the networks that relies on juxtacrine mode of signalling where the signal-sending and signal-receiving cells are in direct contact with one another. The signal sent is one of the two families of ligands in mammals, Jagged (*JAG1* and *JAG2*) and Delta-Like (*DLL1*, *DLL3*, and *DLL4*). While attached to the signal-sending cell, the ligand bind to its receptor on the membrane of the signal receiving one. Upon the binding of a ligand to one of the four members of the family of mammalian Notch receptors (*NOTCH1, NOTCH2, NOTCH3,* and *NOTCH4*), a series of proteolytic reactions takes place cleaving the receptors and releasing the Notch intracellular domain. Once released, the Notch intracellular domain translocates and enters the cell nucleus where it acts as a transcriptional coactivator (Bray 2016). Interestingly, the rather simple design of the Notch signalling network is not reflected in the diversity of its outcomes as Notch signalling functions uniquely among a wide range of cell types and during different developmental and physiological stages. However, such diversity of outcomes might also be due to the cross-talking events between some components of the Notch network with other signalling networks including TGF-beta and Wnt signalling pathways (Blokzijl et al., 2003; Collu et al., 2014).

1.6.1. Conservation of Notch

A recent comparative-genomics study of the Notch signalling network have revealed that some of the components of the Notch signalling network is conserved among all animals and some of the genes are even shared with placozoa and sponges which suggest that Notch signalling was present in the last common ancestor of all metazoans and predates bilateral symmetry, and was then strongly conserved during the course of animal evolution (Gazave et al., 2009). Furthermore, a few components of the Notch signalling network are also conserved among other non-metazoan eukaryotes which might even suggest an older origin of Notch signaling. However, the majority of the core signalling components of the Notch network appear to be an oddity of animals as no homologues of Notch receptors and their Delta-like ligands were found to be conserved beyond metazoans (Gazave et al., 2009).

1.6.2. Notch functions in vertebrates

Despite being related to *Drosophila* for years, the relatively recent discovery of the conservation of several homologues of Notch signalling among most animals suggested it might have roles in other species. Consequently, a substantial amount of research has been done to explore the roles played by Notch signalling in vertebrates development. As a result, it has been shown that Notch plays an important role in neurogenesis (Chitnis et al., 1995), lymphoid development (Robey et al., 1996), somitogenesis (De Angelis et al., 1997), and vasculogenesis (Roca and Adams 2007) during development as well as in the preservation of tissue homeostasis in adult life (Bigas and Espinosa 2012). Thus, it has become apparent that Notch signalling functions in several stages of life is a common feature in most animals and not restricted to *Drosophila* as it was once thought.

1.6.3. Notch in human disease

Given the major roles played by Notch signalling in the regulation of neural development, it comes unsurprisingly that mutations in Notch are related to several neurologic disorders. For instance, mutations in *NOTCH1* has been deemed responsible for the pathogenesis of human T-cell Acute Lymphoblastic Leukemia (Weng et al., 2004), and mutations in *NOTCH3* has been linked with the hereditary stroke disorder CADASIL (Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy) (Joutel et al., 1996). Furthermore, Alagille syndrome is usually caused due to mutations in both *NOTCH2* and *JAG1* (Li et al., 1997; McDaniell et al., 2006). Finally, mutations in *NOTCH4* have been shown to increase the susceptibility of developing schizophrenia among some individuals (Wei and Hemmings 2000). However, despite the wide range of neurological disorders that are caused by Notch signalling mutations, Notch mutations has also been described in other non-neurological disorders. For instance, mutations in the *NOTCH2* are heavily related to the severe bone loss disorder Hajdu-Cheney syndrome (Simpson et al., 2011). In addition, due to the roles played by Notch signalling in somitogenesis, mutations among several genes in the Notch signalling network have been linked with several abnormal vertebral segmentation disorders like spondylothoracic dysostosis (STD) and spondylocostal dysostosis (Dunwoodie 2009). Finally, mutations in *JAG1* and *NOTCH1* has been related to range of cardiovascular diseases (De la Pompa and Epstein 2012).

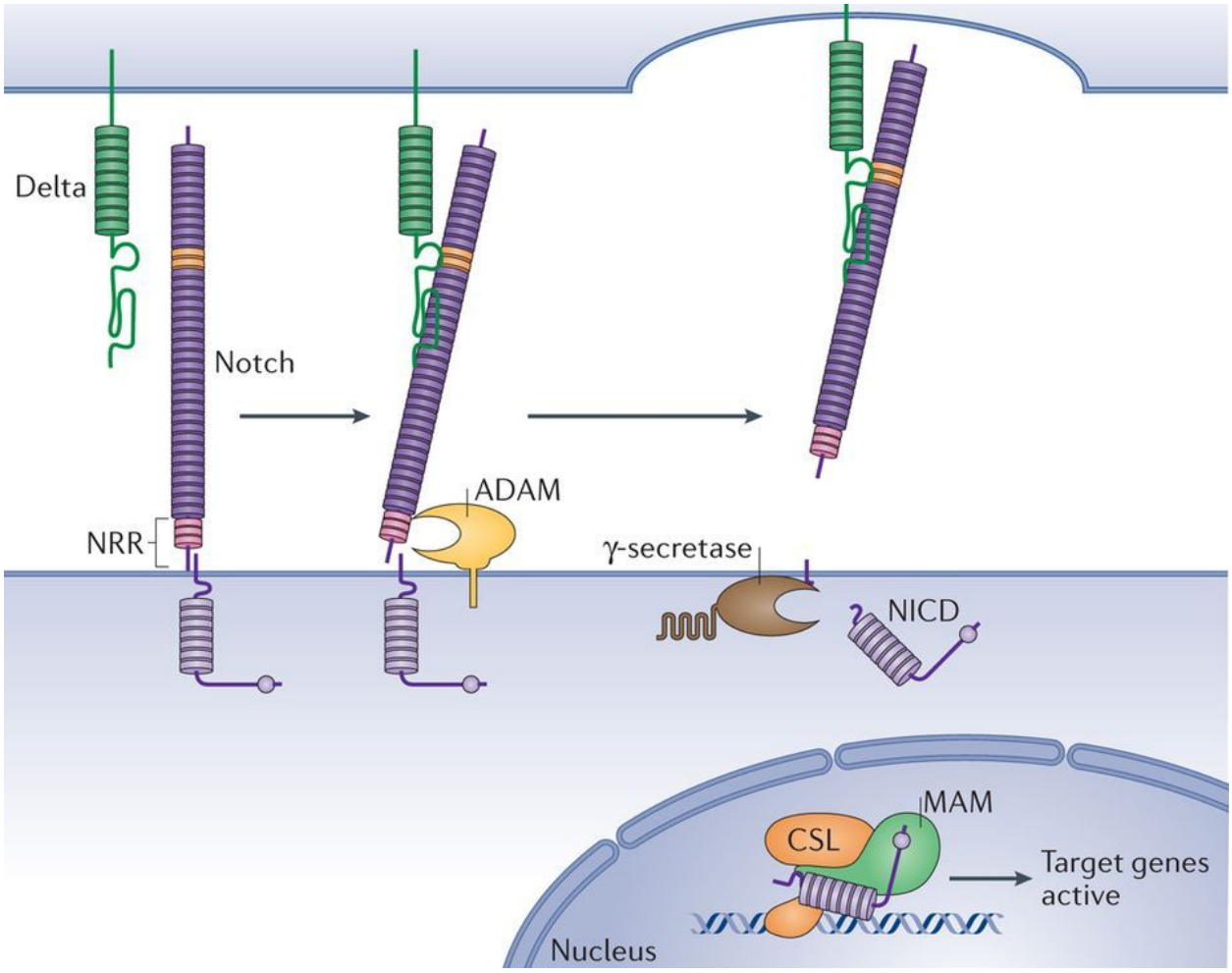

Figure 1.7 | The Notch signalling network: The core Notch signalling map where the signaling is initiated by the binding of one of the ligands (while still bound to the surface of the signal-sending cell) to one of the receptors on the surface of the signal receiving cell. Consequently, by the aid of ADAM, the intracellular domain of the activated Notch receptor is cleaved and translocates to the nucleus where it interact with the transcription factors (adapted with permission from Bray 2006).

1.7. Aims

For more than a century, cell signalling has been recognised as a crucial process for the proper functioning of all living organisms. Moreover, the diversity and complexity of cell signalling has been uncovered as well as its huge impact on human disease. On the other hand, relatively recently, microRNAs became well recognised as one of the major regulators of gene expression in mammals. All this led to a great interest in both topics. However, a comprehensive understanding of the regulatory roles played by microRNAs in different cell signalling networks is missing. In this work, I provide a broad overview of the regulatory relation between two of the most exciting topics in today's biology, microRNAs and cell signalling. In addition, this work aims to provide an understanding of the evolutionary bases of these regulatory interactions. To achieve this, I highlight the general aspects of microRNAs regulation among some of the most widely studied signalling networks using different sets of microRNAs. In addition, I provide answers to some of the known issues in microRNA research as well as highlighting other potential issues that need to be considered.

2. Methods

2.1. Reconstruction of signalling networks

The complex nature of the ErbB signalling network made the process of reconstructing its signalling map challenging. Many genes were reported to be involved in ErbB signalling in some publications but were not mentioned in others, Therefore, the current map (chapter 3) was assembled after compiling data from several publications. Based on the findings of Citri and Yarden (2006), the genes that were reported to be involved in ErbB signalling were then directed into one of four signalling levels based on the roles they play in the signalling network. The four levels were as follow: level zero of the network were the ligands; level one (the first level in the signal-receiving cell) were the Erbb receptors; level two, the signalling pathways that are activated by the binding of the ligands and receptors; and finally level three (inside the cell nucleus) which enclosed the transcription factors. In contrast with ErbB, the assembly of the Notch, Toll-like and TRAIL networks was straight forward as most of the published papers present matching networks of signalling. However, I mainly used the reviews of Bray (2016) and O'neill et al., (2013) to construct the maps of Notch and Toll-like networks respectively. For all four networks, the Genecards database (Stelzer 2016) was used to check synonymous names of the genes reported in literature to match them to the gene names provided by Ensembl (For a full list of the genes used in the analysis of this networks and their corresponding levels, see Appendix 1).

2.2. Genomic data

2.2.1. 3'UTRs

As microRNA regulation relies mainly on complementarity between the microRNA seed regions and the 3'UTRs of their target mRNAs, microRNAs target prediction in this work was conducted by finding such complementarity. For this, the BioMart webtool from the Ensembl genome browser (release 82) (Cunningham et al. 2015) was used to retrieve 3'UTRs sequences of the studied genes. For genes where more than one transcript were available the transcripts that have the longest 3'UTRs were chosen to avoid possible loss of microRNA target sites in transcripts with shorter 3'UTRs. However, in some cases, the longest 3'UTRs belonged to non-functional protein coding transcripts known as nonsense mediated decay (see Chapter 5). In such cases, all nonsense mediated decay transcripts were excluded and the longest protein-coding transcript was used instead.

All human cell receptors were identified using the PANTHER classification system (version 13.1) (Mi et al., 2013). The gene names reported by PANTHER were then used as a filter in Ensembl BioMart query and the 3'UTRs start and end were selected as attributes. As a result, the length of the 3'UTR of each of the receptors were calculated. However, unavailable or incomplete 3'UTRs were discarded.

2.2.2. MicroRNAs

Mature microRNAs sequences for human and mouse were obtained from miRBase (release 21) (Kozomara and Griffiths-Jones 2013). MicroRNA sequences were then clustered using an in-house PERL script into families according to their extended seed regions (nucleotides 2-7) based on the definitions from Bartel (2009). As such, different microRNAs that share the same seed region, now belong to the same family since they will have the same canonical targets.

2.3. MicroRNA expression and conservation levels

In order to identify microRNAs that are highly expressed, I used the RNA sequencing data of Meunier et al (2013) collected from five different human organs and available on NCBI gene expression omnibus (https://www.ncbi.nlm.nih.gov/geo/) with the following accession numbers: frontal cortex (GSM995300), cerebellum (GSM995301), heart (GSM995302), kidney (GSM995303), and testis (GSM995304). Furthermore, to count the expression levels for each microRNA, miRcounts software (Hatlen 2016) was used to map these sequencing reads to mature human microRNA sequences that were obtained from miRBase. By default, miRcounts tolerates zero mismatches between the reads and the reference query. Based on miRcounts results, microRNAs that had an expression level of (>1/1000) in at least one of the five tissues were identified as highly expressed ones and a list of these 474 microRNAs belonging to 380 seed region families were used in the analysis. Furthermore, to restrict the analysis to only microRNAs that are broadly conserved between species, the list of broadly conserved microRNAs of Friedman et al., (2009) was used and microRNAs were identified and were then used in the analysis. Finally, the lists of highly-expressed and highly-conserved microRNAs

were joined and 158 microRNAs belonging to 86 families were identified as both highly expressed and highly conserved.

2.4. MicroRNA target prediction

All microRNA target predictions were done using seedVicious (v1.1) (Marco 2018) with default parameters unless specified. SeedVicious requires an input of a transcript file (3'UTRs sequences retrieved from BioMart) and a microRNA file (microRNA sequences retrieved from miRBase). In cases where more than one transcript existed for the same gene, the longest transcript function of seedVicious (-l) was used. In addition, to avoid the length bias of the 3'UTRs, the numbers of nucleotide per each 3'UTR were kept and microRNA interactions were represented as a number of interaction per 100 nucleotides of 3'UTR. The prediction of microRNA targets that have more than one target site was done using the pairs function of seedVicious (-x 4) with all other parameters as default. Also, a BASH script was written to identify only microRNAs with multiple target sites where a minimum distance of between targets is set. Finally, the energy function of seedVicious (-e) was used to calculate the thermodynamic stability of the microRNA-target duplexes.

2.5. Evolutionary analysis of microRNAs regulation

I used seedVicious' ancestral state function (-x 3) which implements a maximum parsimony model to predict events of gains and/or losses of microRNA target sites between different species based on the given tree. Seven species were selected for such comparison; human, chimpanzee, rhesus macaque, mouse, rat, rabbit, and chicken. The choice of these species were determined by

few factors. First, the availability of their genomic data as for some other species there were not enough data (especially microRNA data). In addition, the rodents group is one of the most widely used group of animal models.

To run seedVicious' ancestral state function, an alignment of the 3'UTR should be provided in addition to the microRNAs and a newick tree file. The exact genomic locations of each 3'UTR was parsed from Ensembls BioMart as a chromosome number, 3'UTR start, and 3'UTR end positions. Based on this information, the pre-aligned sequences of the 3'UTRs where obtained from the table browser feature of the UCSC genome browser were the group was set to "comparative genomics" and the table was set to "Multiz align". However, as UCSC genome browser outputs the alignments in a multiple alignment format (MAF), the MAF to FASTA webtool (Blankenberg et al., 2011) available on the public galaxy server (https://usegalaxy.org/) was used to obtain the alignment in fasta format so that it can be used as seedVicious input. The MAF to FASTA tool was provided by the maf outputs and the type of output was set to "one sequence per species" and the studied species were selected from the table pre-detected by the maf to fasta tool. However, for some genes, not all species were available in the alignments which has been noted and the Newick trees were adjusted accordingly. In addition, Newick trees files were manually constructed based on the species present in the alignments. Finally, ClustalX (version 2.0) (Larkin et al., 2007) was used to eliminate gap-only columns from the alignments and to visualize the alignments. Finally, the TimeTree database (http://www.timetree.org/) (Kumar et al., 2017) was used to calculate the divergence times between species in order to calculate evolutionary turn over of gains and losses.

2.5. Evolutionary analysis of ErbB ligands

The dN/dS ratio was used to study the rates of evolution of ligands proteins, amino acid sequences of the conserved protein domains of ErbB ligands were obtained from Stein and Staros (2006) and their coding sequences (CDS) were retrieved from Ensembl genome browser (release 82) (Cunningham et al. 2015). Blast2 (Tatusova and Madden 1999) was used to retrieve the exact region of the CDS that codes for the conserved amino acid sequence. PAL2NAL (version 14) (Suyama et al., 2006) was used to calculate the dN/dS ratios between the amino acid and the coding sequences between two species at a time per each ligand with the "remove gaps in frame stop codons" parameter set to "yes" and all other default parameters.

2.6. SNPs analysis of ErbB microRNA targets

PolymiRTS database (version 3.0) (Bhattacharya et al., 2013) was used to parse information about the reported SNPs and INDELs in either the target sites or the seed regions of microRNAs targeting ErbBs. Using the search tool of the database, I downloaded all microRNAs that are involved in regulating ErbBs (while having SNPs and/or INDELs) and filtered out for only those that were highly expressed as identified earlier.

2.7. Nonsense mediated decay transcripts

In order to count for the numbers of nonsense mediated decay transcripts used in pre-computed microRNA target prediction software. All interactions data were downloaded from the following software: TargetScan (7.2) (Agarwal et al., 2015), DIANA-microT-CDS (v5) (Paraskevopoulou

et al., 2013), RNA22 (v.2) (Miranda et al., 2006), miRanda (Enright et al., 2003), miRDB (v5.0) (Wong and Wang 2014). Depending on the format of the interactions data given by each software, the Ensembls transcripts IDs were parsed. Finally, the transcript IDs were matched to a list of transcript ID and transcript type fetched from Ensembls BioMart.

2.8. Precision of microRNA target predictions

Lists of experimentally validated microRNA interactions were downloaded from miRTarBase (release 7.0) (Chou 2017) for both human and mouse. Such lists were used as the gold-standard lists of interactions that seedVicious predictions were then compared to. As miRTarbase presents its data in a tabular form where a single microRNA (i.e. not microRNA families) is reported to target a gene name, seedVicious was run again using single mature microRNAs and 3'UTRs were obtained again from Ensembls BioMart but with gene names instead of gene IDs. In order to calculate precision of microRNAs target interactions, a bash script was written to first identify (and keep counts of) interactions that either appears in the seedVicious list but not the miRTarBase one (false positives), in the miRTarBase but not seedVicious one (false negatives), or in both lists (true positives). Finally precision was calculated using the following formula

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

2.9. High complex microRNA identification

In order to identify high complex microRNAs, a PERL script was used to exclude microRNAs whose seed regions consists entirely of mono-, di- or tri-nucleotide repeats. This has resulted in

the elimination of 29 human microRNAs and 15 mouse ones. Once these microRNAs were identified, the analysis was carried out again using a new list of microRNAs which includes all microRNAs except the ones identified as low-complex. Finally, in order to test the effect of the eliminated microRNAs, regular expression bash scripting was used to eliminate miRTArBase interactions that were only reported by high-throughput experiments.

2.10. Data visualization

All the plots in this work were done using R (version 3.3.1) (R core team, 2014). The p-values on the box plots are results of Wilcoxon non-parametric test that was used to test the significance of the difference between each box (representing each level of signalling) and the adjacent one. While the non-parametric Kolmogorov–Smirnov test was used in other plots as specified in text.

2.11. Scripts

All the scripts used in this work can be accessed online (see Appendix 2)

3. The evolution of microRNA-mediated regulation of the ErbB signalling network

---

Aim: On their own, microRNAs and ErbBs are very well studied. Also, the targeting of some ErbBs by single microRNAs and the effect of this on the regulation of different diseases is well established. However, the global regulatory role microRNAs play on the ErbB network is yet to be fully understood as the current knowledge misses a comprehensive view of microRNA-mediated regulation of the whole ErbB signalling network. In this chapter, I provide an extensive, systematic, and global view of the regulatory relationships between microRNAs and ErbBs. In addition, I highlight the evolutionary events that, I believe, should be taken into consideration when studying microRNA-mediated regulation of the ErbB network.

MicroRNAs, on one hand, are among the major regulators of gene expression in eukaryotes, that are able to regulate around 30% of all genes (Stark et al., 2005). The ErbB network, on the other hand, is one of the most studied signalling networks due to its important roles in normal physiology and disease (Samaga et al., 2009). Unsurprisingly, microRNAs regulate ErbBs signalling through targeting several genes among multiple levels of the signalling network (Barker et al., 2010). Starting from the top of the network, microRNAs have an established role in targeting and regulating several ligands that are essential for the initiation of signalling

through binding to one of the four tyrosine-kinase receptors (*ErbB1*, *ErbB2*, *ErbB3*, and *ErbB4*) (Kedmi et al., 2015). Also, the regulatory interactions between microRNAs and ErbBs receptors is even more perceived due to their clinical and therapeutic importance as they are heavily linked to different diseases (Wieduwilt and Moasser 2017). Each of the four receptors are targets of many microRNAs and a few of this regulatory interactions have been validated experimentally and are thought as important factors in regulating regular functioning of the network as well as its role in diseases including cancer (Giles et al., 2011) (Table 1). Besides ligands and receptors, the regulation of different genes among several levels down the network by microRNAs is also known, for instance, MAPK (Chakraborty et al., 2016), PI3K/AKT (Josse et al., 2014), RAS (Johnson et al., 2005), and STAT (Kohanbash and Okada 2012) are all regulated by different microRNAs.

Table 1 | Experimetally-validated interactions between microRNAs and ErbBs

Gene	Experimentally-validated microRNA Interactions	Reference
EGF	miR-223-3p	Fabris et al., 2016
HBEGF	miR-132-3p	Molnar et al., 2013
TGFA	miR-374a	Wu et al., 2013
AREG	miR-34c-5p	Tung et al. 2017
NRG1	miR-125a-3p	Yin et al., 2015
EREG	miR-192-5p	Morimoto et al., 2017
ErbB1	miR128b; miR-7	Chan et al., 2012; Webster et al., 2009
ErbB2	miR-331-3p; miR-1296-5p	Epis et al., 2009; Shan et al., 2017
ErbB3	miR-22	Ling et al., 2012
ErbB4	miR-146b	Zhu et al., 2016

The binding of an ErbB ligand to one of its receptors initiates a simultaneous series of signalling cascades that are collectively known as the ErbB signalling network. The ErbB network is majorly divided into three layers of signalling: (i) the input layer, comprising the eleven ligands and the four tyrosine kinase receptors, (ii) the signal processing layer, consisting of the signalling pathways and their nuclear transcription factors, (iii) the output layer, cellular actions (Yarden and Sliwkowski 2001). In fact, ErbB ligands and receptors are considered the two key levels of the ErbB network (Barberan et al., 2016). However, beyond the ligands and receptors, there exist two more levels of intracellular signalling that culminates in one of the cellular actions. The activation of the ErbB receptors by the ligands is followed by an activation of one of different intracellular signalling pathways including RAS/MAPK and PI3K/AKT. Consequently, the activation of such pathways is translated in the nucleus to transcription factors that finally decide the fate of the signalling by inducing a cellular action (Yarden and Sliwkowski 2001). As a result, the ErbB signalling network can be classified into four vertical levels of signalling: ligands, receptors, adapters (signalling pathways), and transcription factors. However, the exact extent of which genes take place in such cascades remains controversial as despite the agreement on the four receptors and eleven ligands, the fact that some signalling cascades can be activated via cross-talking and some are only activated in diseases makes it challenging to determine which genes are exactly involved ErbB signalling processing (Yarden and Sliwkowski 2001). I here present the main ErbB signalling map which includes only the genes that have been majorly agreed upon to be taking part in ErbB signalling (figure 3.1) (see Methods chapter).

Interestingly, the regulatory relationship between microRNAs and ErbBs is not restricted to a one-way regulation in which microRNAs target and regulate ErbBs. In fact, the activation of the ErbB signalling through binding of growth factors to *ErbB1* and *ErbB2* in turn activates a subset machinery of cellular interactions that decrease the abundance of a set of tumour suppressor microRNAs that normally regulate the expression of oncogenic ErbB transcription factors like *FOS*. These microRNAs, known as immediate-down regulated (ID-miRs), are usually repressed in brain and breast cancers which implicates their role in regulating tumor progression. By this means, oncogenic ErbBs protect themselves from being targeted by tumor suppressing microRNAs by repressing the expression of such microRNAs which in turn promote the rapid induction of their oncogenic effects (Avraham et al., 2010).

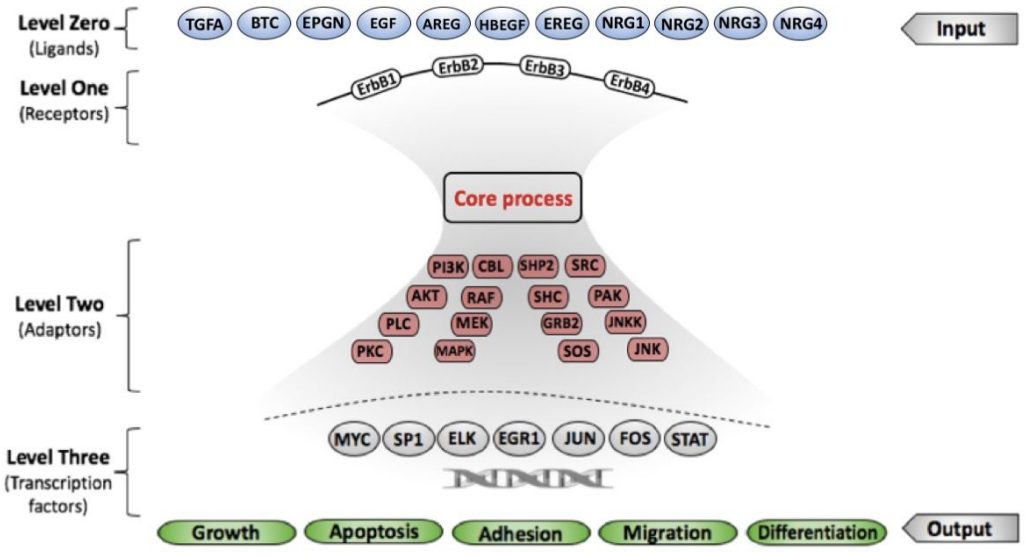

Figure 3.1 | ErbB signalling network: Different levels of ErbB signalling starting with the ligands level (level zero), the receptors level (level one), the adaptors level (level two), and finally the transcription factors level (level three).

Given the clear fundamental role that receptors play in the functioning of the whole ErbB network, I here explore whether this importance is reflected in microRNA regulation of the network. To begin with, I test whether the receptors have longer 3'UTRs (thus more complementarity with microRNAs and more regulatory interactions) to be under a strict regulation by microRNAs. In addition, regardless of the length, I test whether ErbB receptors are favouribally regulated by microRNAs by counting for the presence of target sites per specific number of nucleotides on the 3'UTRs. Also, I test whether they are both longer and are favourably regulated by microRNAs.

3.1. Lengths of the 3'UTRs of ErbBs

Throughout evolution, it is thought that selective pressure has shortened the 3'UTRs of some genes to avoid microRNA regulation (such genes are known as anti-targets and usually have housekeeping functions) (Stark et al., 2005). Oppositely, the 3'UTR of other genes have always been maintained long for the exact opposite reason; to accommodate more microRNA target sites. Interestingly, genes that are involved in signal transduction are specifically known to have relatively longer 3'UTRs, possibly to conserve more microRNAs target sites. Slowly evolving genes, like ErbBs, usually preserve longer 3'UTRs and thus are tightly regulated by microRNAs (Cheng et al., 2009).

Here, I explore trends in the length of the 3'UTRs across the ErbB network in regards to microRNAs and their target sites that might help in figuring out whether ErbBs are intentionally attracting (or avoiding) regulatory microRNA targeting interactions by having longer or shorter

3'UTRs. In figure 3.2, I show the length of ErbB receptors compared to the length of all other genes in the network (figure 3.2 A and 3.2 C), the different levels of signalling (figure 3.2 B), and the length of 3'UTRs between different species (figure 3.2 D). Figure 3.2 C, shows significantly longer 3'UTR for *ErbB1* and *ErbB4* when compared to all other genes of the ErbB network. Interestingly, *ErbB1* and *ErbB4* are the only autonomous receptors that are able to bind to ligands and initiate a tyrosine kinase activity (Yarden and Sliwkowski 2001). In addition, while comparing the lengths of ErbBs 3'UTRs between human and mouse, there is no clear difference in length (figure 3.2 D). In fact, such analysis would have been much more informative if more species were included in the analysis. Unfortunately, the lack of reliable 3'UTR sequences data in other species did not facilitate such analysis. Furthermore, in figure 3.3 B, there is no apparent direction by which different ErbBs have longer or shorter 3'UTRs in regards to the level of signalling. In conclusion, we see that the receptors level and specifically the autonomous receptors (*ErbB1* and *ErbB4*) have longer 3'UTR that potentially host more microRNA target sites, and this pattern is also observed in mouse.

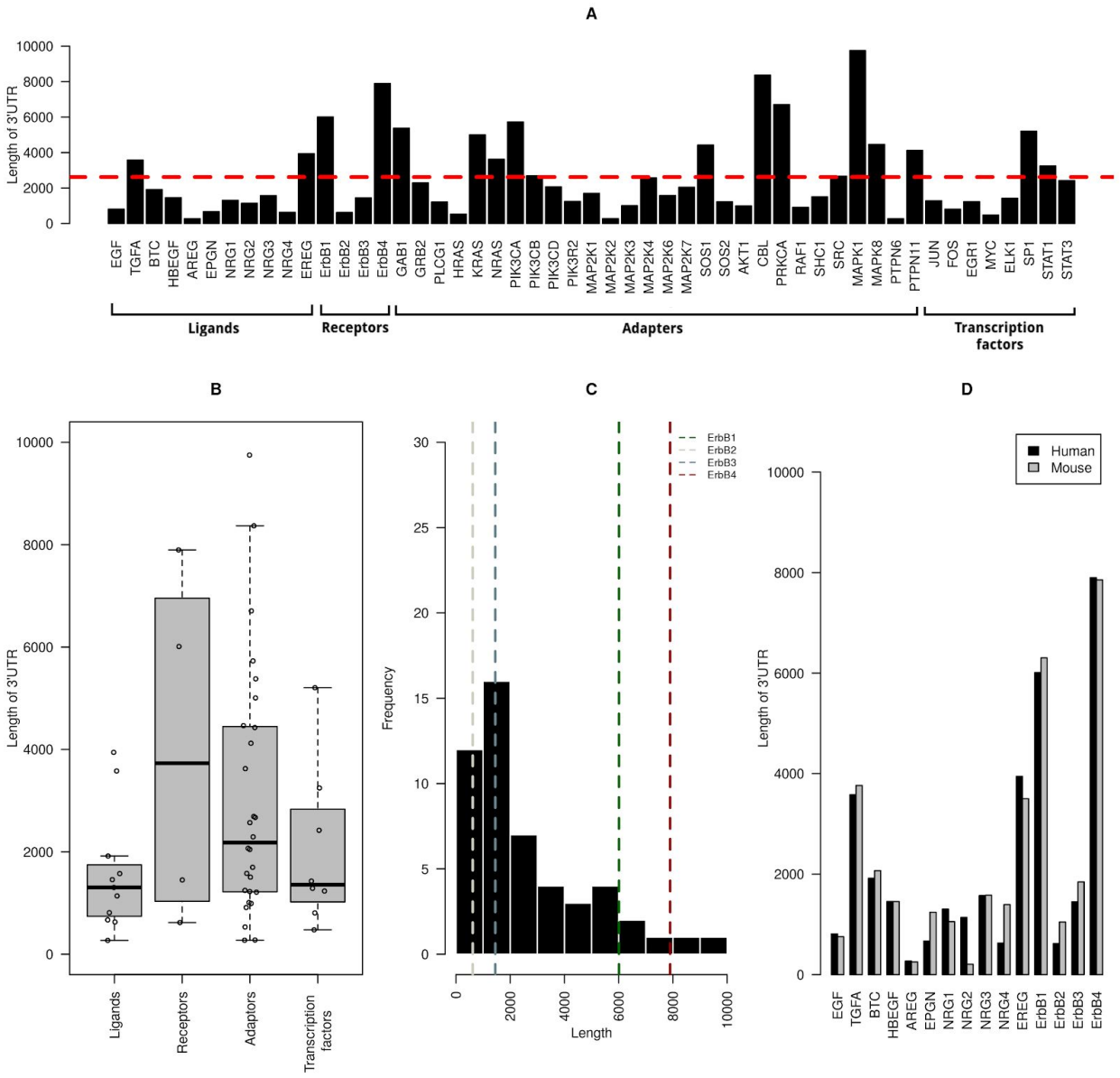

Figure 3.2 | The evolution of the lengths of the 3'UTRs of ErbBs: (A) The lengths of the 3'UTRs of all ErbB genes, red line showing the mean length. (B) Box plots showing the lengths of ErbBs in regards to the levels of signal transduction. (C) The frequency of different lengths of 3'UTRs across the ErbB network with the four receptors highlighted. (D) A comparison of the lengths of the 3'UTRs of ErbB ligands and receptors between human and mouse.

3.2. MicroRNA-mediated regulation of the ErbB network

Since microRNAs bind to the 3'UTRs of their targets through complementarity, the longer the 3'UTR, would logically mean the more the microRNA target sites and thus more regulatory interactions. In fact the length of 3'UTRs might also vary for evolutionary reasons that are not related to microRNAs. However, I here try to avoid this length bias by presenting microRNA-mediated regulation as density rather than absolute numbers (i.e. how many microRNA target sites are there per 100 nucleotides of 3'UTR). In addition, microRNAs with pairs of targets (i.e. a microRNA that have more than one target site on the same 3'UTR) are believed to be more solid predictions as a gene with a single microRNA target site on its 3'UTR is unlikely to be repressed by a microRNA (Stark et al., 2005). For this reason, I present a further layer of analysis that is restricted to those interactions where there are at least two target sites per a microRNA on the 3'UTR. Furthermore, experimental analysis of the regulation efficacy of microRNAs with multiple target sites, shows an optimum efficacy when pairs of target sites are located in close proximity to each other (Sætrom et al., 2007; Marco 2018). So I here present a final, most strict, level of regulation restricted to interactions with pairs of targets that are no more than 50 nucleotides apart. Finally, in addition to complementarity, number of target sites, and distances between them, a microRNA needs to exist in high levels at the same tissue as its target for the inhibitory interaction to take place. For this reason, I use a set of 380 microRNA families that are highly expressed in at least one human tissue in order to have a more clear picture of microRNA-mediated regulation of the ErbB network.

In regards to the four levels of ErbB signalling, ligands and receptors are known to be the key levels of the network. Here, I explore whether there are any patterns by which microRNAs regulate the four layers of the network. Firstly, the distribution of microRNAs target sites among the 3'UTRs among different levels is uniform. For instance, in figure 3.3 A, the median density of the presence of microRNA target sites on the 3'UTR (the second quartile of the boxplot) does not show any difference. Likewise, in figure 3.3 B, the lack of difference is even more clear when restricting the analysis to highly expressed microRNAs only. In figures 3.3 C and 3.3 D, while counting for all microRNAs and highly expressed ones with more than one target site, the receptors level clearly stands out as the most tightly regulated level. However, the p-values (Wilcoxon test) does not indicate that this difference is significant. One explanation for that might be the relatively low number of sample size of the receptor level (only four genes). Finally, restricting the analysis to microRNAs with pairs that are no more than 50 nucleotides apart shows an even distribution of the density of microRNAs targeting among the levels which again might be affected by the small sample size due to the strict filtering (figures 3.3 E and 3.3 F).

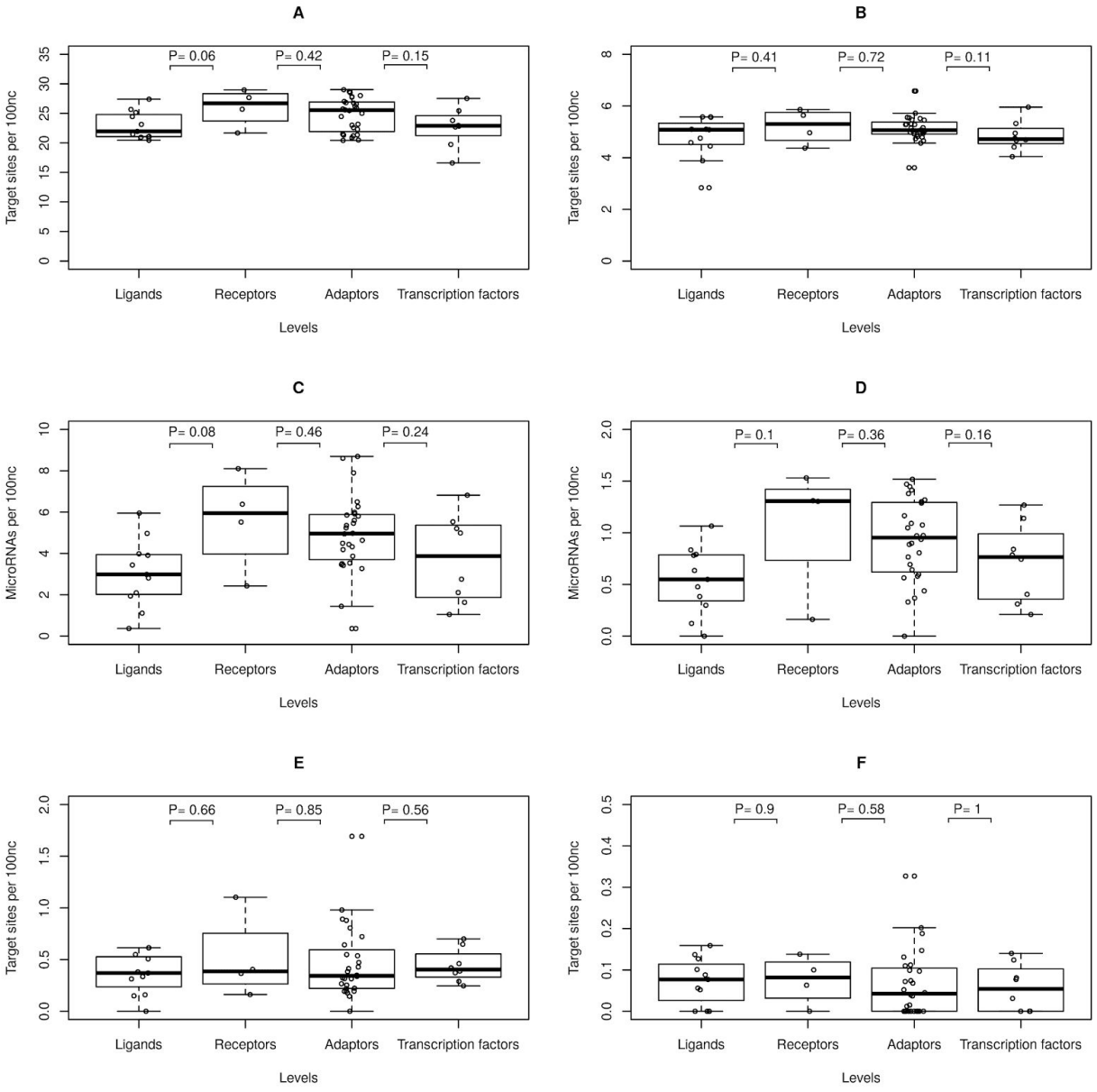

Figure 3.3 | The patterns of microRNA-mediated regulation of ErbB signalling: Boxplots showing the patterns by which microRNAs regulate different levels of ErbB signalling. (A) Target sites for all human mature microRNAs. (B) Target sites for highly expressed microRNAs. (C) All microRNAs with more than one target site. (D) Highly expressed microRNAs with more than one target site. (E) All microRNAs with a maximum distance of 50 nucleotides between pairs of target sites. (F) Highly expressed microRNAs with a maximum distance of 50 nucleotides between pairs of target sites.

Interestingly, according to figures 3.2 B, 3.3 C, and 3.3 D there seems to be a correlation between the lengths of the 3'UTRs and the density of microRNAs with more than one target site. In figure 3.4, I show a clear and significant correlation between the length of the 3'UTRs of all the ErbB genes and the number of microRNAs with multiple target sites per 100 nucleotide of 3'UTR where the longer the 3'UTR the higher the density of microRNAs target sites. According to this result, while restricting to the more fierce regulatory relationships (those having multiple target sites per microRNA) members of the ErbB network with longer 3'UTRs (like receptors) have higher density of microRNA target sites.

In order to explore the densities of microRNA regulation in ErbB receptors, I use the sliding windows smoothing method with a window size of 200 nucleotide and 1 nucleotide offset between windows, I compare the density of the targeting (i.e. how many target sites are there per 200 nucleotides) of highly expressed microRNAs to the observed sequences and a thousand replicates that were generated by shuffling the 3'UTRs while preserving the dinucleotide frequency in each of the 3'UTRs of the ErbBs receptors. In figure 3.5, there seems to be a difference in distribution of densities between the observed and the average of the expected sequences among most receptors. This indicates that the observed sequences of ErbBs 3'UTRs might be under strong selective pressure that, intentionally, preserve their microRNAs target sites not just their lengths.

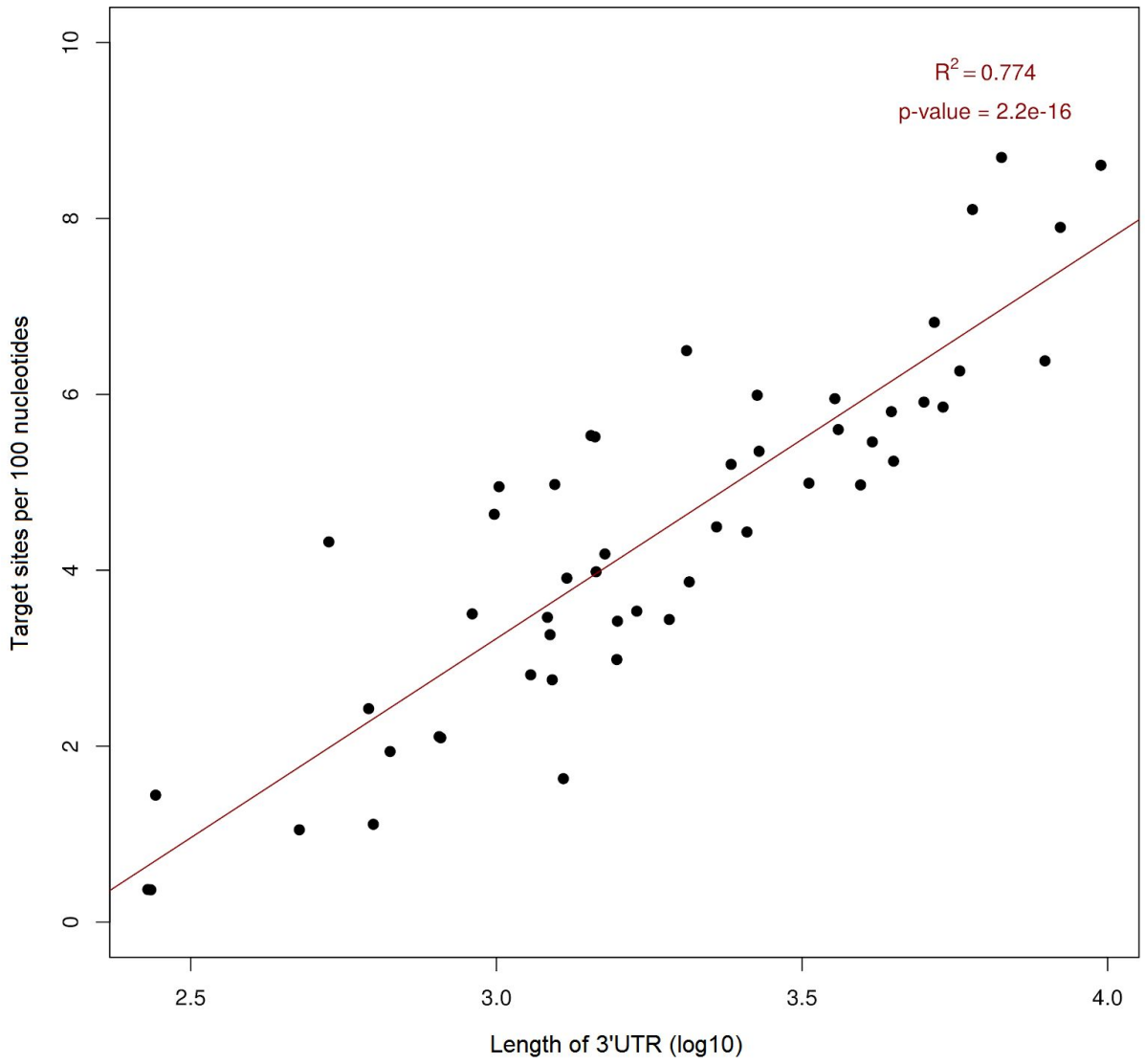

Figure 3.4 | The length of ErbBs 3'UTRs and microRNAs with multiple target sites: The correlation between microRNAs with multiple target sites and the length of their target 3'UTRs among the ErbB signalling network.

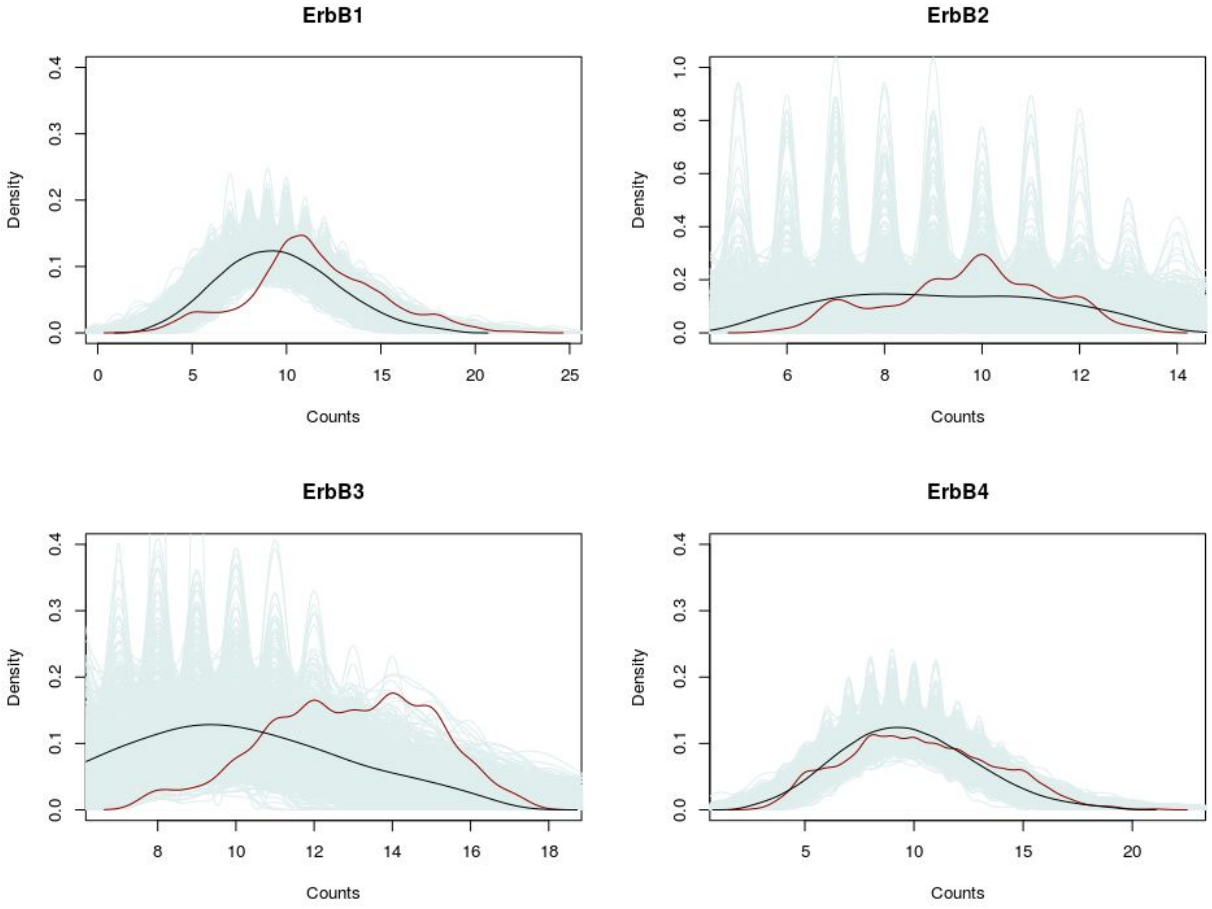

Figure 3.5 | The density of microRNA-mediated regulation of ErbB receptors: The densities of highly-expressed microRNAs target sites among windows (200 nucleotides wide) of 3'UTRs of ErbBs receptors. (Grey lines shows 1000 replicates, black lines shows the average distribution and red lines shows observed sequences).

3.3. The evolutionary conservation of ErbB ligands

Despite the extensive study of ErbBs, their evolutionary conservation, and their cellular functions, it has recently been shown that ErbBs functions are not as conserved among different species (Liu et al., 2013). Ligands, together with receptors, form the key components of the ErbB signalling network (Barberan et al., 2016). While the high conservation of ErbB receptors has long been established, the degree of conservation of ErbB ligands remain unclear (Stein and Staros 2000). Here, I use the protein domain structures of the eleven ErbB ligands to explore their evolutionary conservation as described by Suyama et al., (2006).

One of the most widely used methods of studying proteins is to study the rate by which they evolve. For this, the dN/dS ratio is used to show the rates by which these proteins evolve through showing the rates of nonsynonymous and synonymous mutations where a gene is expected to be neutrally evolving (dN/dS = 1) while a low dN/dS ratio (<1) indicates that the gene is experiencing purifying (negative) selection and a high dN/dS ratio (>1) indicates that the gene is undergoing Darwinian (positive) selection (Kimura 1977). The analysis of the rates by which ErbBs ligands are evolving indicates a low dN/dS ratio between different pairs of species belonging to two groups, primates and rodents. First, the rates between human and chimpanzee within the primates groups were lowest, indicating that they are under strict purifying selection that insures their conservation. Second, the same low rates of evolution is observed between mouse and rat in the rodents group. Finally, the cross-group analysis (human and mouse) also shows relatively low rates of protein evolution (figure 3.6). Such observations, indicate the evolutionary conservation of the ErbBs ligands as key members of the ErbB network. While the

protein domain structures show the conservation of ErbBs ligands, the conservation of the target sites on the 3'UTRs of such ligands is what determines the conservation of microRNA-mediated regulation. As a result, in the next part I explore the conservation of microRNA-mediated regulation among the conserved ErbB ligands and receptors.

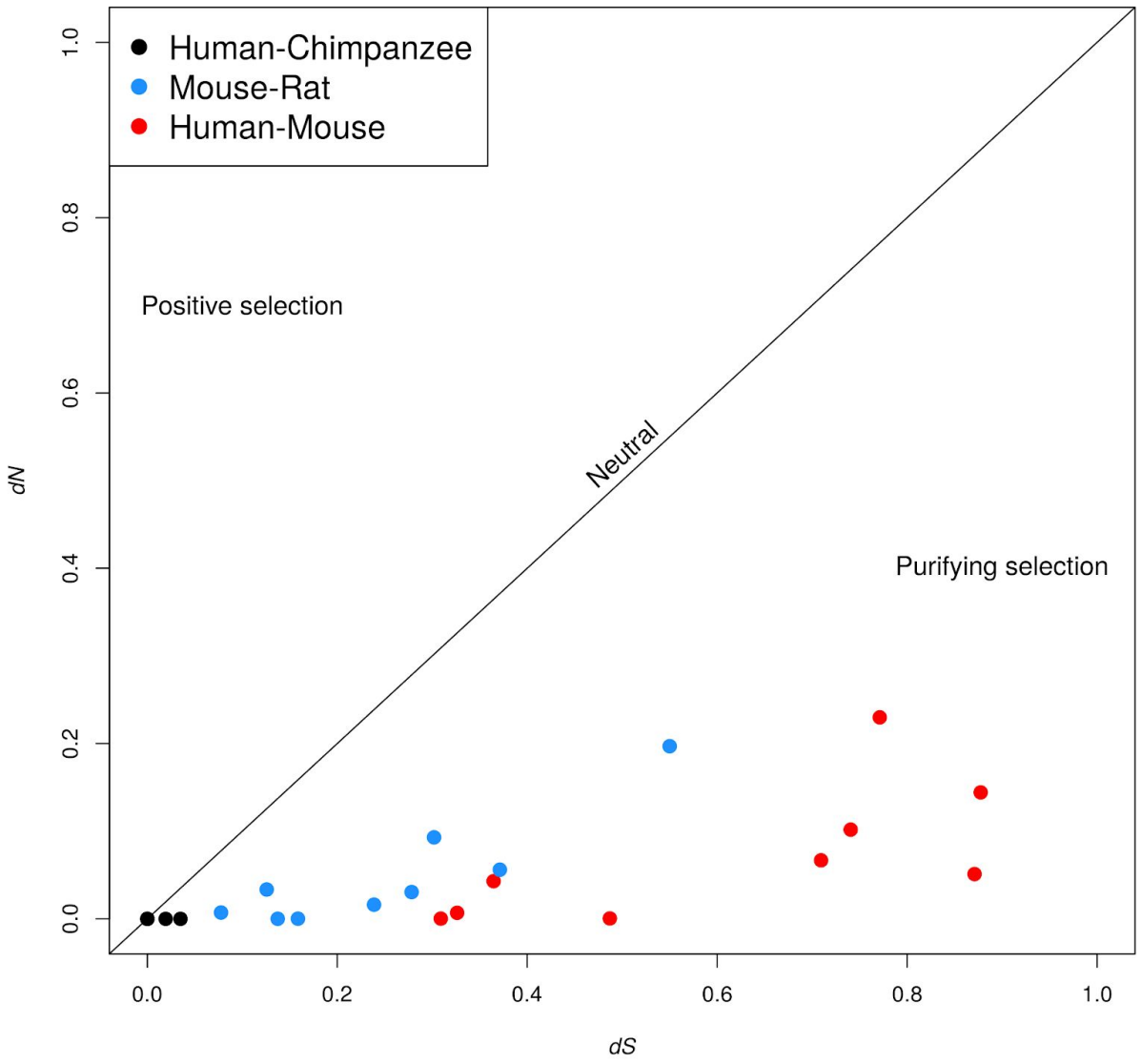

Figure 3.6 | The conservation of ErbB ligands: The rates of protein evolution of the eleven ErbB ligands represented by the dN/dS ratio between primates and rodents (human-mouse) and members of each group (human-chimpanzee and mouse-rat).

3.4. The conservation of microRNA regulation of ErbB signalling

As discussed earlier, both microRNAs and ErbBs are highly conserved across almost all animals. However, this analysis shows that this high conservation is not necessarily the case when it comes to their interactions. In fact, in the majority of microRNAs target prediction software, a big emphasis is given to the conservation of target sites that in many cases, non-conserved target sites are disregarded as false positives (Pinzon et al., 2017). Here, among other conserved interactions, I highlight the existence of many nonconserved ones. I use seedVicious' ancestral state function to predict the conservation of the interactions between microRNAs and ErbBs ligands and receptors by exploring the events of gains and losses of such interactions between seven different species: primates (human, chimpanzee and rhesus macaque), rodents (mouse, rat and rabbit), and a bird (chicken) as an evolutionary outgroup. In figures 3.7 and 3.10 there are relatively high numbers of target sites gains and losses per each node of the phylogenetic tree. However, there seems to be a distinct sum of target sites gains on the primate lineage compared to that of the rodents. In addition, through the course of time, microRNAs have been gaining target sites in ErbB receptors among the tested species which is shown on the timeline of target sites turnover among microRNAs which highlights the necessity of microRNAs interaction in regulating the functioning of the network. Furthermore, I carry out the analysis using a set of highly expressed microRNAs that are conserved among most animals according to the findings of Friedman et al., (2008). In figures 3.8 and 3.11, despite the conservation of microRNAs there still appears to be a high degree of target sites gains and losses between the tested species.

The high numbers of gains and losses of microRNA target sites implied the possibility of target sites emerging (or vanishing) at one stage of mammalian evolution. In addition, the strikingly high number of gains of target-sites on the primate lineage, complemented by the high number of target sites loss on the mouse/rat lineage suggests the possibility of primate-specific interactions between microRNAs and ErbBs. For instance, the highly-expressed microRNA, miR-574-3p is a known tumor-suppressor microRNA through its ability to regulate the expression of *ErbB1* (Chiyomaru et al., 2013). However, this analysis shows that the tumor-suppression powers of miR-574-3p are limited to primates as it does not conserve any target sites in rodents nor chicken. Likewise, the conservation of target sites of both hsa-miR-127-5p and hsa-miR-124-3p in *ErbB1* and *ErbB3,* respectively are limited to the primate lineage where, in the case of the later, a single nucleotide change was the reason behind a target site being conserved only among primates but not any other species (figure 3.9). In fact, these are just few examples of many lineage-specific microRNA regulations among ErbB receptors.

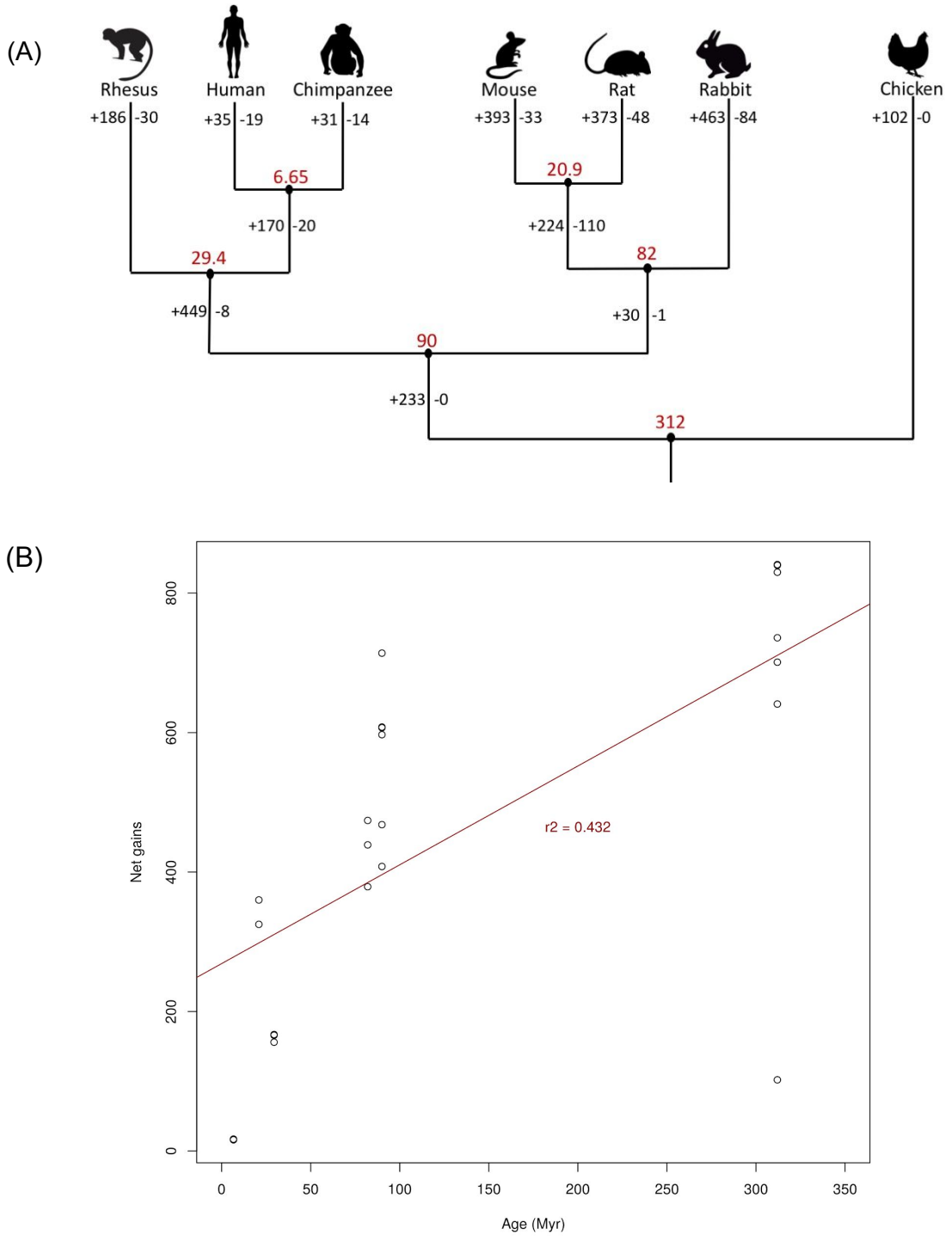

Figure 3.7 | Ancestral state of the regulation of ErbB receptors by highly-expressed microRNAs: (A) seedVicious' predictions of the events of gains (+) and losses (-) of target sites of highly-expressed microRNAs in ErbB receptors. (B) Timeline of turnover by showing the net gains of microRNA target sites in million years (Myr).

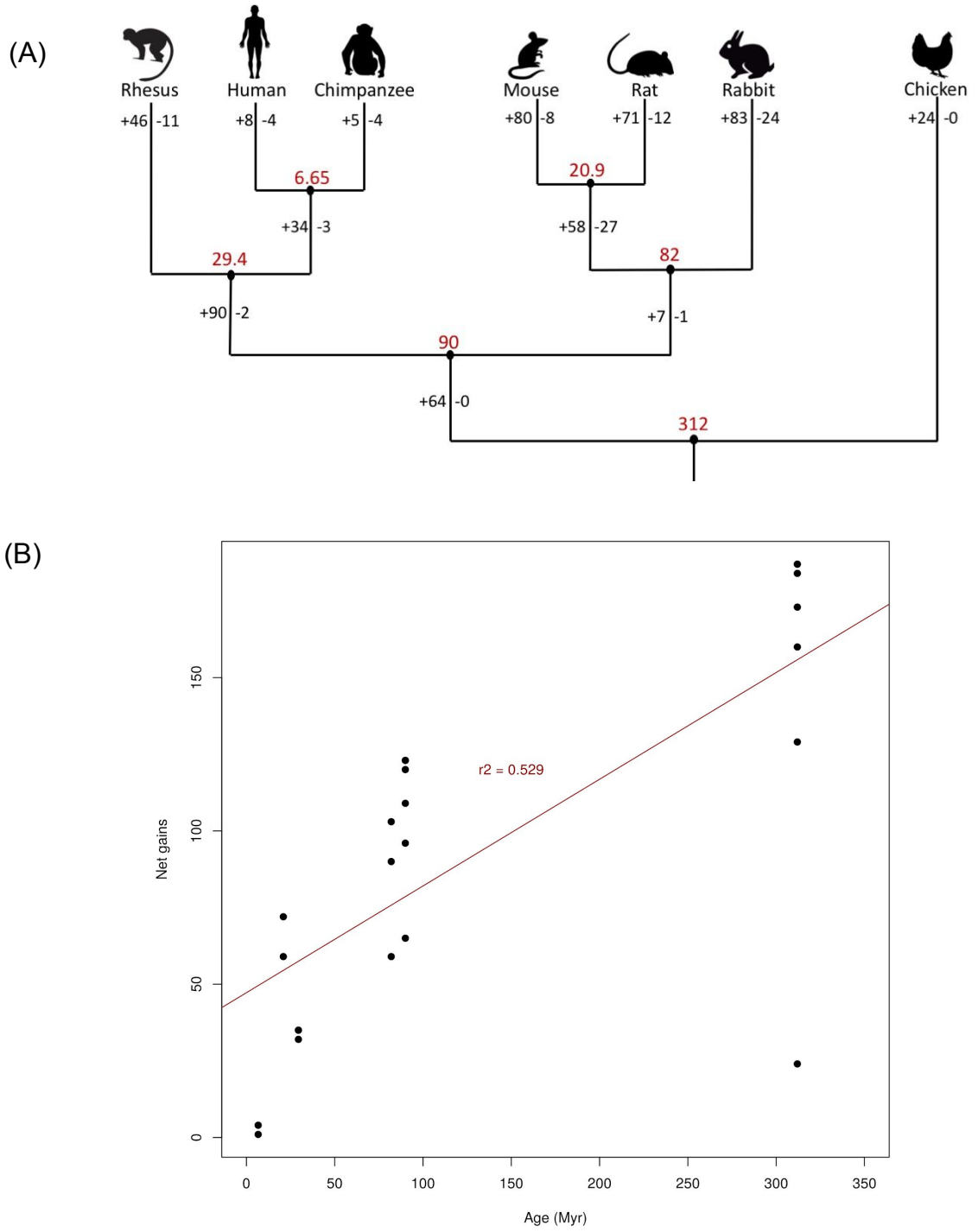

Figure 3.8 | Ancestral state of the regulation of ErbB receptors by conserved and highly-expressed microRNAs: (A) seedVicious' predictions of the events of gains (+) and losses (-) of target sites of highly-expressed microRNAs in ErbB receptors. (B) Timeline of turnover by showing the net gains of microRNA target sites in million years (Myr).

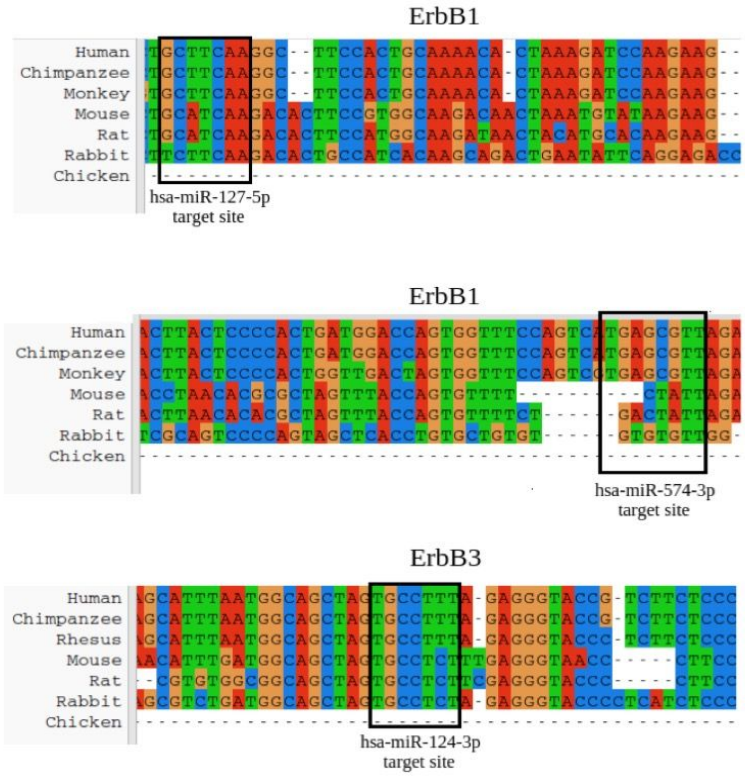

Figure 3.9 | Lineage-specific microRNA regulation in ErbB receptors: Examples of events were target sites of highly expressed microRNA are only conserved among primates in *ErbB1* and *ErbB3*.

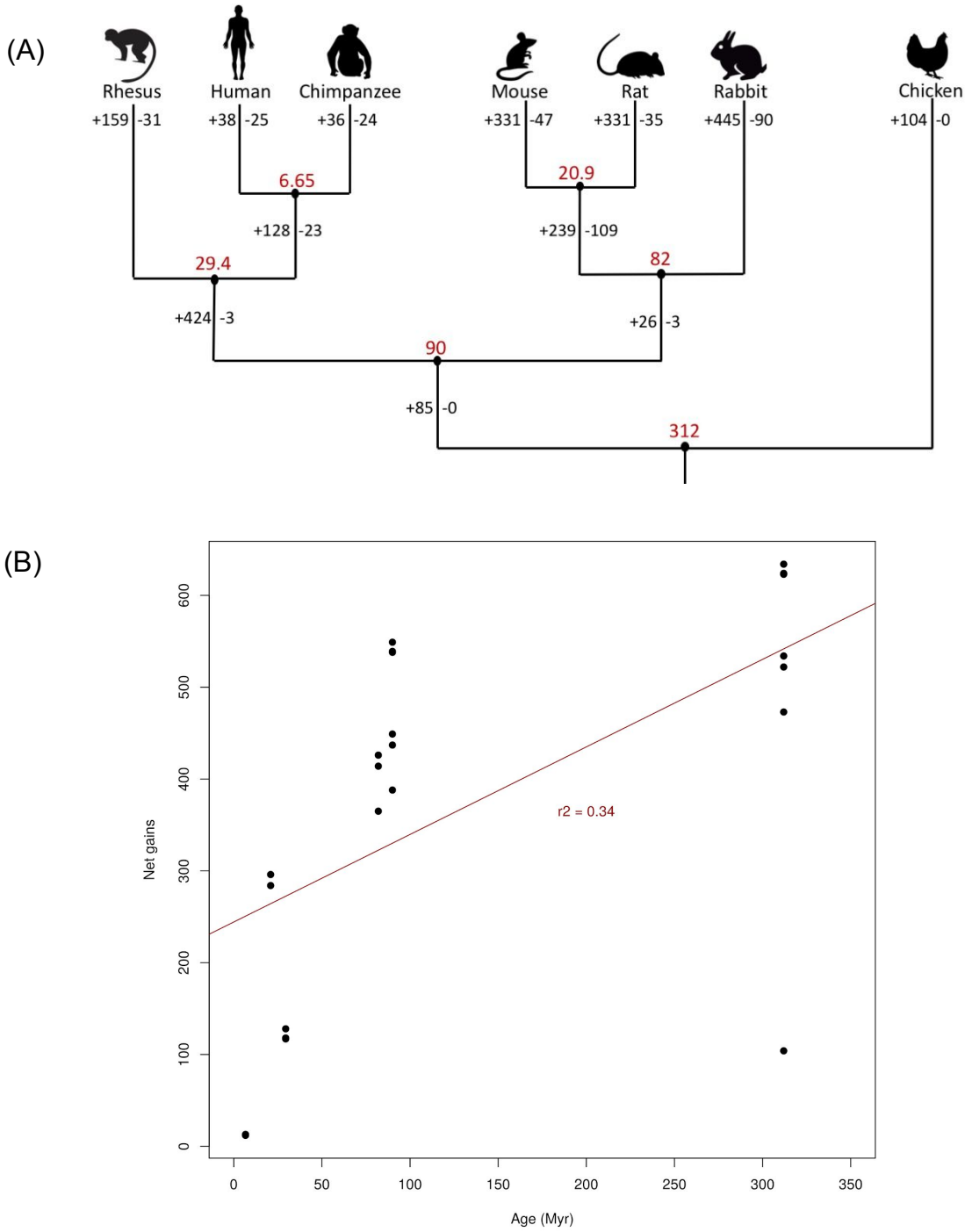

Figure 3.10 | Ancestral state of the regulation of ErbB ligands by highly-expressed microRNAs: (A) seedVicious' predictions of the events of gains (+) and losses (-) of target sites of highly-expressed microRNAs in ErbB receptors. (B) timeline of turnover by showing the net gains of microRNA target sites in million years (Myr).

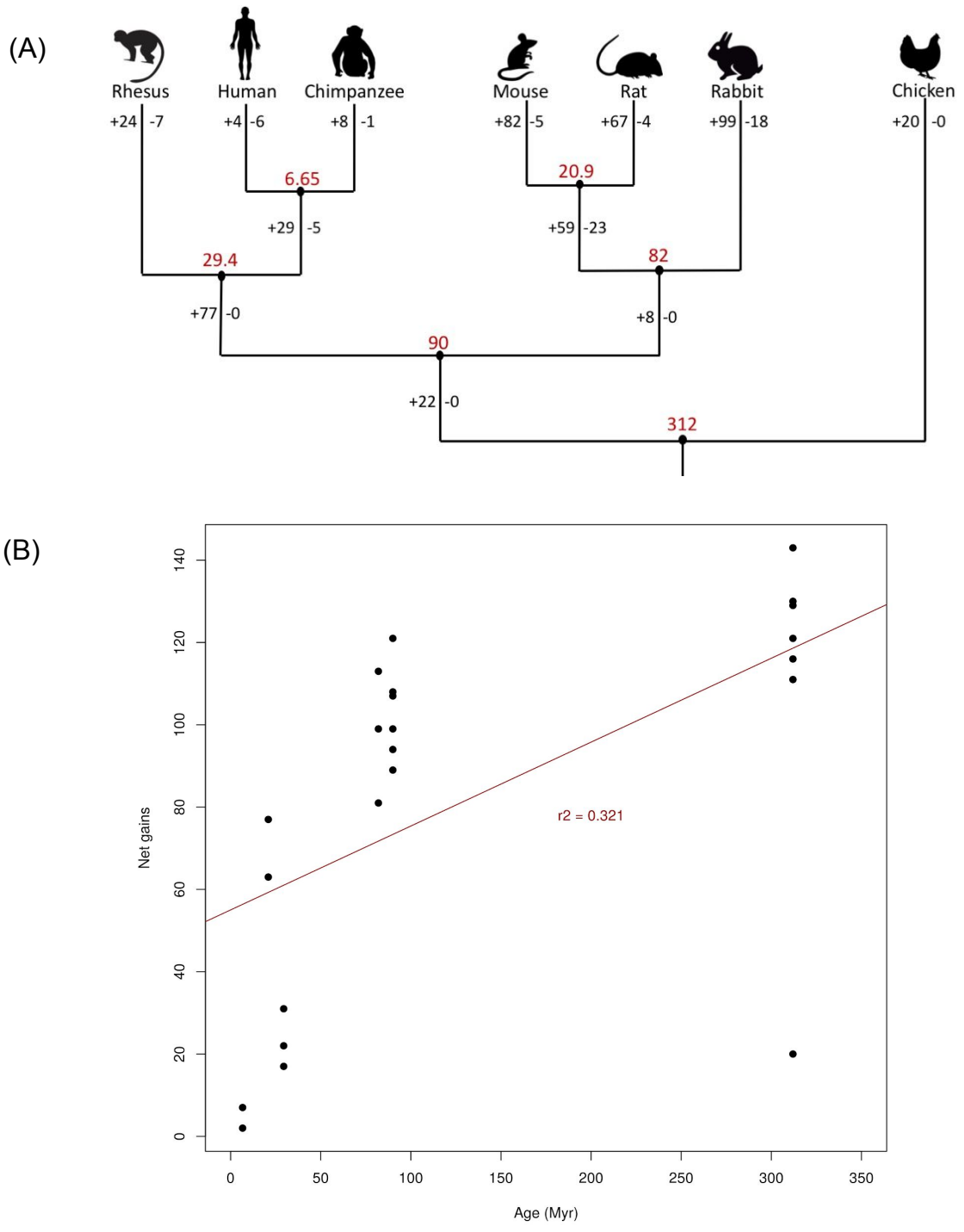

Figure 3.11 | Ancestral state of the regulation of ErbB ligands by conserved and highly-expressed microRNAs: (A) seedVicious' predictions of the events of gains (+) and losses (-) of target sites of highly-expressed microRNAs in ErbB receptors. (B) timeline of turnover by showing the net gains of microRNA target sites in million years (Myr).

3.5. SNPs and microRNA-mediated regulation of ErbBs

Single nucleotide polymorphism (SNP) is the most abundant type of genetic variation in the human genome (Kim and Misra 2007). By definition, a SNP is a single-nucleotide variation that is present in at least 1% of the population (Brookes 1999). The presence and profound impact of SNPs in ErbB genes is well established (Toomey et al., 2016), SNPs in coding regions of ErbBs receptors can alter their structures and/or functions. In addition, such SNPs are strongly linked to ErbBs roles in cancer and their drug resistance (Choura and Rebaï 2009). However, cancer drugs are not the only targeters of ErbBs, microRNAs are there too. In fact, SNPs can possibly alter the whole regulatory functions of microRNAs on ErbBs by either creating a new interaction or destroying an established one. Here, I report the numbers of target sites created or destrdue to SNPs in target sites and/or seed regions of microRNAs as reported by PolymiRTS Database (release 3.0) (Bhattacharya et al., 2013). According to these results, there appears to be a relatively large number of SNPs in both the seed regions of microRNAs targeting ErbBs and the target sites of such microRNAs on the 3'UTRs of ErbBs ligands and receptors. Such SNPs, together with coding regions SNPs, may result in a paradigm shift in our understanding of the whole regulatory role microRNAs play in ErbB signalling (Table 3.2). In addition, the longer the 3'UTRs of the gene the more SNPs there seems to be and the more complementarity with microRNAs having seed regions SNPs. However, despite the relatively-long 3'UTR of *ErbB4*, the abundance of SNPs seems minimal.

Table 3.2 | SNPs in either the target sites or the seed regions of microRNAs targeting ErbBs

	SNPs in microRNAs target sites	Disrupted interactions	Created interactions	Interactions disrupted by SNPs in seed regions	Interactions created by SNPs in seed regions
TGFA	68	149	175	29	36
BTC	4	9	5	11	7
HBEGF	20	56	37	28	27
EREG	46	110	83	28	34
NRG1	5	10	7	-	-
ErbB1	35	81	71	14	10
ErbB2	-	-	-	16	13
ErbB3	13	49	44	26	24
ErbB4	1	1	-	-	-

3.6. Discussion and conclusion

In this chapter, I presented a comprehensive overview of the regulation of the ErbB signalling network by microRNAs. While many other studies have focused on the regulation of a single ErbB gene (or level, like receptors) by a single microRNA in a single species (usually human). I here provide a more diverse and comprehensive view as I believe it is important to look upon the ErbBs signalling network as an intact body where a change (like a regulation by microRNAs) in one level of signalling will often have its influence on the whole functioning of the network.

MicroRNAs play a fundamental role in the regulation of the ErbB signalling. According to the results of this chapter, the autonomous receptors *ErbB1* and *ErbB4* tends to be under a relatively more strict regulation by microRNAs which might reflect the importance of these receptors in initiating the intercellular cascades of ErbB signalling. In addition, most of the therapeutic developments and clinical trials of anti-ErbB agents focuses on *ErbB1* and *ErbB2* as they are the two receptors with the most established roles in cancer so far (Tebbutt et al., 2013). However, several studies have recently shown the important role played by *ErbB4* in various types of cancer (Prickett et al., 2009; Williams et al., 2015; Soung et al., 2006). As I show here, *ErbB4* and ErbB1 are the two most tightly regulated by microRNA in the whole network *ErbB4*, and the two autonomous receptors. As a result, I believe that more focus should be given to *ErbB4* as a potential target in cancer therapy.

Using the recent microRNA target prediction tool, seedVicious (Marco 2018), I was able to explore the evolutionary history of microRNA-mediated regulation of the ErbB network. This

analysis uncovered some of the evolutionary events that resulted in the way we observe microRNA-mediated regulation of the network today. Unfortunately, the lack of reliable genomic data of other species than human and mouse did not allow me to provide a wider view of the evolution of microRNA-mediated regulation of the ErbB network. In addition, I have highlighted the high number of SNPs in microRNA seeds and their complementary target sites among ErbBs. Such SNPs may posses major effects in the regulation of ErbBs by microRNAs in both diseases and normal physiology and should not be ignored.

In fact, the therapeutic potential of ErbBs is too promising to ignore (Arteaga and Engelman 2014). However, their unsuccessful therapeutic use so far is a known issue that have been holding their progress back for long time. While several explanations to such deficiency have been proposed, I believe that, in line of my results, there are more factors that need to be considered. First, while both microRNAs and ErbBs ligands and receptors are heavily conserved my results show that the way microRNAs regulate ErbBs is not as conserved due to the high numbers of gains and losses of microRNAs target sites. In addition, the interspecies variability by the presence of a high number of SNPs affecting microRNA-mediated regulation of the network should be taken into consideration.

4. MicroRNA-mediated regulation of Notch and Toll-Like signalling networks

---

Aim: Further to the previous chapter, I here study the regulation of microRNA to different cell signalling networks namely, Notch, Toll-like, and Trail. ErbB is one of the most complex cell signalling networks. However, this chapter explores the patterns of microRNA-mediated regulation among other networks that also play fundamental roles in many biological process and their mis-regulation have been linked with several diseases.

MicroRNAs target and regulate many cell signalling networks (Ichimura et al., 2011). In the previous chapter, the study of microRNA-mediated regulation to one of the most notable cell signalling networks (ErbB) revealed a few interesting aspects about the regulatory relationship between microRNAs and ErbBs. As a result, it is decisive to explore whether such aspects are common across different cell signalling networks or just specific of ErbB. For instance, while both Notch and Toll-like networks rely on conserved receptors that are important regulators of many processes and are targets to many drugs like ErbBs. Notch, on one hand does not rely on intracellular cascades for the transmission of signals as once a receptor is activated, the intracellular domain of the receptors localizes at the nucleus where it interact with transcription factors and result in cellular actions. Toll-like, on the other hand, is one of the major regulators

of immune responses. Thus, its receptors are not activated by the means of signals sent by other cells but rather they are activated by pathogens. In addition, Notch signalling -unlike ErbB, Toll-like and Trail- relies on juxtacrine signalling in which the cell sending the signal is in direct contact with the cell receiving it and thus, signalling is restricted to neighboring cells. However, each of these networks have their common and in contrast features when compared with ErbB. The three networks studied in this chapter, are less complex in nature as there seems to be a general agreement on which genes take place in the networks.

The Notch signalling network is considered one of the most studied cell signalling networks, as for more than a century, Notch has been providing insights on the development, physiology, and diseases of many species including humans. While it shares some features with ErbB as both networks play important roles in development and are highly related to several diseases including cancer as well as relying on a family of four transmembrane receptors for the initiation of the signalling, Notch is also different in several features. In addition to its juxtacrine signalling nature, the Notch signalling network lacks any intermediate adaptor pathways between the cell membrane (where the signalling starts upon the binding of a Notch receptor to on of its ligands) and the cell nucleus (where transcription takes place) as the intracellular domain of the Notch receptors itself gets translocated to the nucleus and interact with the transcription factors. For such reasons, the Notch signalling network may offer new insights on the structures of cell signalling networks and how it might affect microRNA regulation.

Likewise, the Toll-like signalling network is also considered among the most studied signalling networks due to its paramount role in immune response. Signalling through the Toll-like network is initiated by the binding of a pathogen (or a pathogen related molecule) to one of the nine human toll-like receptors (4 of which are endocytic receptors, and the others are cell surface ones). However, the fact that Toll-like signalling is not initiated by the means of traditional ligands but rather by pathogens makes it a of special importance in exploring the effects of the lack of ligands on the regulation of a signalling network. In addition, despite the obvious direction of regulation between microRNA and the Toll-like signalling network (where microRNAs regulate the expression of genes involved in Toll-like signalling), there seems to be a striking opposite face to this regulation as in fact Toll-like signalling can as well regulate microRNA expression levels (Aalaei-andabili 2013). The changes in microRNA expression levels in response to Toll-Like signaling was first reported by Taganov et al., (2006) when the expression of miR-146a/b was elevated upon the activation of the Toll-like signalling. Shortly after, several studies have reported similar patterns where specific microRNAs are either highly or lowly expressed in response to Toll-Like signalling (Moschos et al., 2007; Bazzoni et al., 2009; O'neill et al., 2011). Furthermore, in a pattern similar to that observed in ErbB, some (early-response) microRNAs have shown elevated expression levels in the tissues upon the stimulation of toll-Like receptors (He et al., 2014), furthermore, another microRNA (miR-21) is expressed at higher level but after longer times post the stimulation of Toll-like receptors thus it is referred to as (late-response) microRNA (Sheedy et al., 2010). As a result, exploring the mechanisms of microRNA regulation of Toll-like signalling might provide new insights into such a relationship.

The results of the ErbB analysis as well as several publications showed a special importance of receptors as being heavily targeted by microRNAs (Adem et al., 2016). Accordingly, it was important to explore whether such pattern applies to other major cell signalling networks or not. The choice of Notch, Toll-like and TRAIL networks stemmed from some similarities they share with ErbB in some cases, and some contrasts they have in others. While integrating the lessons learned from studying ErbB regulation by microRNAs, this chapter will show if the patterns that were observed in ErbB signalling apply to other major cell signalling networks or not. In addition, in order to keep the predictions of microRNA regulation to these networks as precise as possible, I only present seedVicious' predictions of highly expressed microRNAs that have multiple target sites on the 3'UTRs of their targets.

4.1. MicroRNA regulation Toll-like signalling

Unlike ErbB, Notch, and TRAIL, the Toll-like signalling network does not rely on ligands for the activation of the network. In fact, the signalling is initiated through the binding of a pathogen (or a pathogen related molecule) to one of the Toll-like receptors. As a result, the Toll-like signalling network can be classified into three levels: receptors, adaptors (including signalling pathways) and, transcription factors.

Following the analysis of ErbB signalling network, I here explore the patterns of microRNA regulation to the Toll-like signalling network and to explore whether the lack of a ligands level would be reflected in the patterns of microRNA regulation. In figure 4.1, in regards to the above levels, the adaptor level seems to be slightly higher-regulated by highly-expressed microRNAs. However, the differences of the regulation between levels of the Toll-like signalling network are not significant.

In the previous chapter, a strikingly high number of turn over of microRNA target sites among the highly-conserved ErbB receptors level has been reported. In addition, there was a clear bias in the gains of microRNA target sites in the primate lineage. The analysis of the gains and losses of microRNAs target sites on Toll-Like receptors, despite still showing a high number of gains on the primate lineage, does not show a clear direction of gains and losses between primates and mouse/rat as that shown in ErbB. In addition, there appears to be an accumulation of a high number of net gains of target sites at the divergence time between primates and rodents which might stem from the high number of gains among the rodents lineage (figure 4.2).

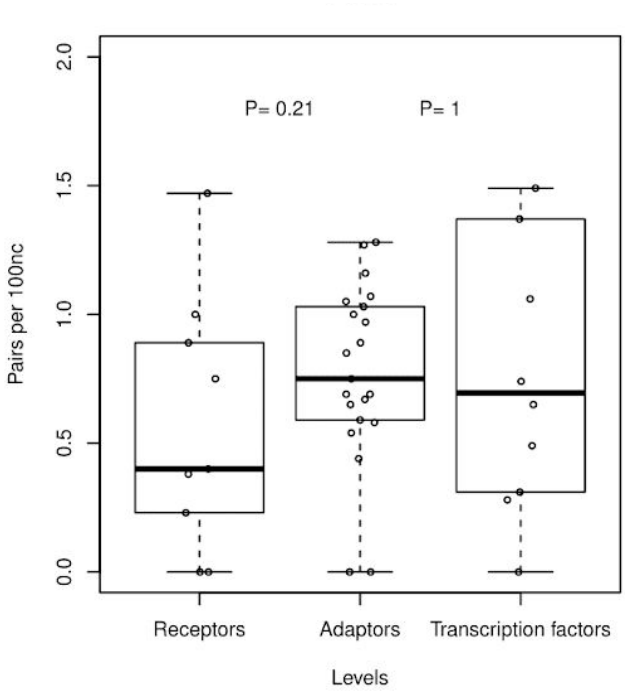

Figure 4.1 | MicroRNA-mediated regulation of Toll-like signalling network: Boxplot showing the regulation of highly-expressed microRNAs that have multiple target sites per each of the different levels of toll-like signalling network.

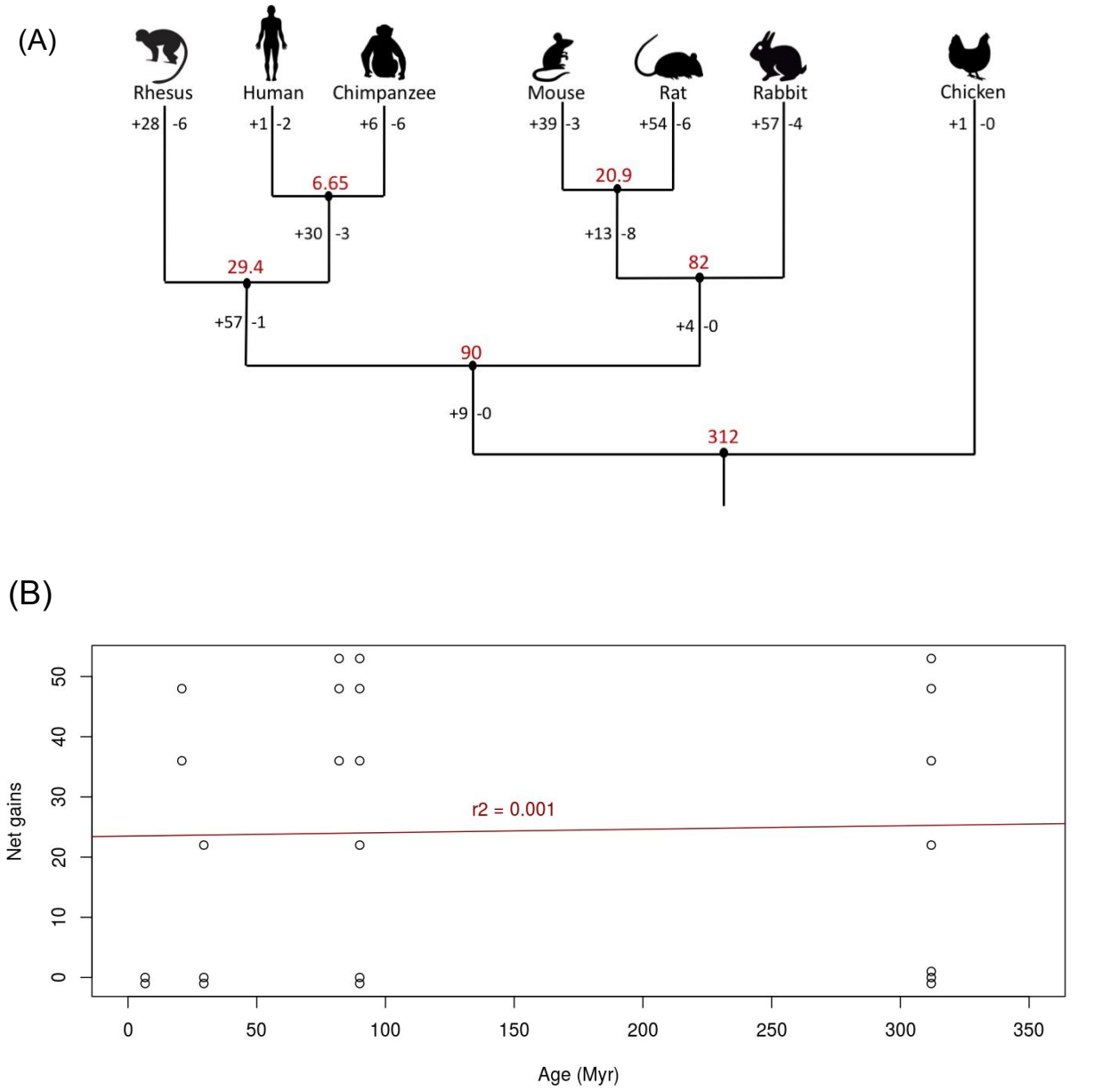

Figure 4.2 | Ancestral state of the regulation of Toll-like receptors by microRNAs: (A) seedVicious' predictions of the events of gains (+) and losses (-) of target sites of highly expressed and conserved microRNAs in Toll-like receptors. (B) Timeline of turnover by showing the net gains of microRNAs target sites in million years (Myr).

4.2. MicroRNA regulation of NOTCH signalling

The core Notch signalling network can be classified into three levels of signalling as ligands, receptors, and transcription factors. Once a ligand (attached to the signal-sending cell) binds to the extracellular domain of one of the transmembrane Notch receptors (on the signal-receiving cell), a series of biochemical interactions take place releasing the intracellular domain of the receptor (NICD) which then travels to the nucleus where it interacts with the transcription factors inside the cell nucleus regulating the expression of several genes (Bray 2016). In the next part, I study microRNA regulation of these different levels of the Notch signalling network to explore whether any of the levels is more strictly regulated which might be reflected in the importance of such levels.

In terms of microRNA-mediated regulation and in regards to the levels of Notch signalling, the transcription factors levels seems to be the highest-regulated level by highly-expressed microRNAs, followed by the receptors level and the ligands level respectively. However, while testing the significance of such pattern of regulation, no significance was detected. In addition, it worth noting that the small sample size per level (for instance, only two transcription factors) makes it deceptive to rely on such visualization (Figure 4.3).

Interestingly, the analysis of the gains and losses of the targets sites of microRNAs that are both highly expressed and highly conserved shows a relatively high number of losses on the mouse/rat lineage complemented by a relatively high number of gains on the primate one in a pattern similar to that observed in ErbB receptors (Figure 4.4). Such pattern might reflect the important

roles played by microRNAs in the regulation of such networks in primates in normal physiology and in disease.

These results, in addition to the result of microRNA regulation of the Toll-like network and in contrast to the results of chapter 3, shows that microRNA-mediated regulation might be consistent among different levels (i.e. the signalling level does not have an effect on the degree of microRNA-mediated regulation).

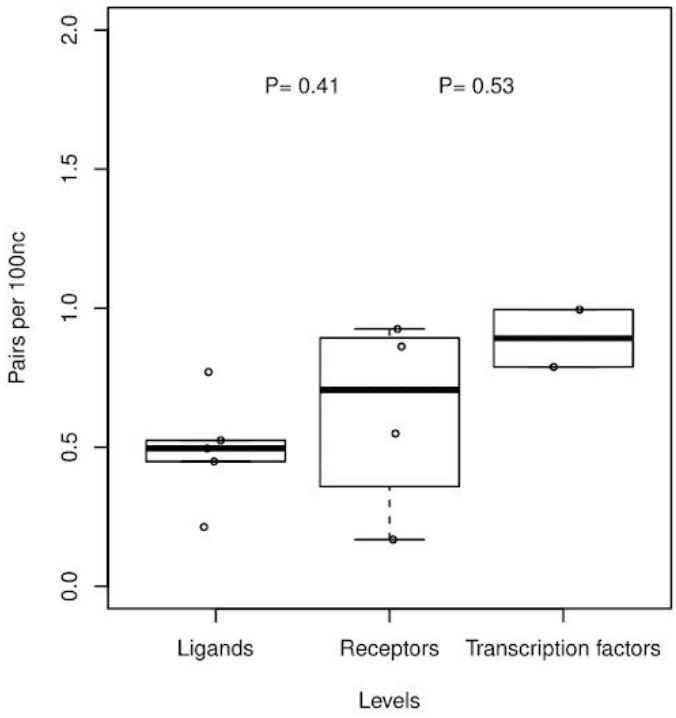

Figure 4.3 | MicroRNA-mediated regulation of Notch signalling network: Boxplot showing the regulation of highly-expressed microRNAs that have multiple target sites per each of the different levels of Notch signalling network.

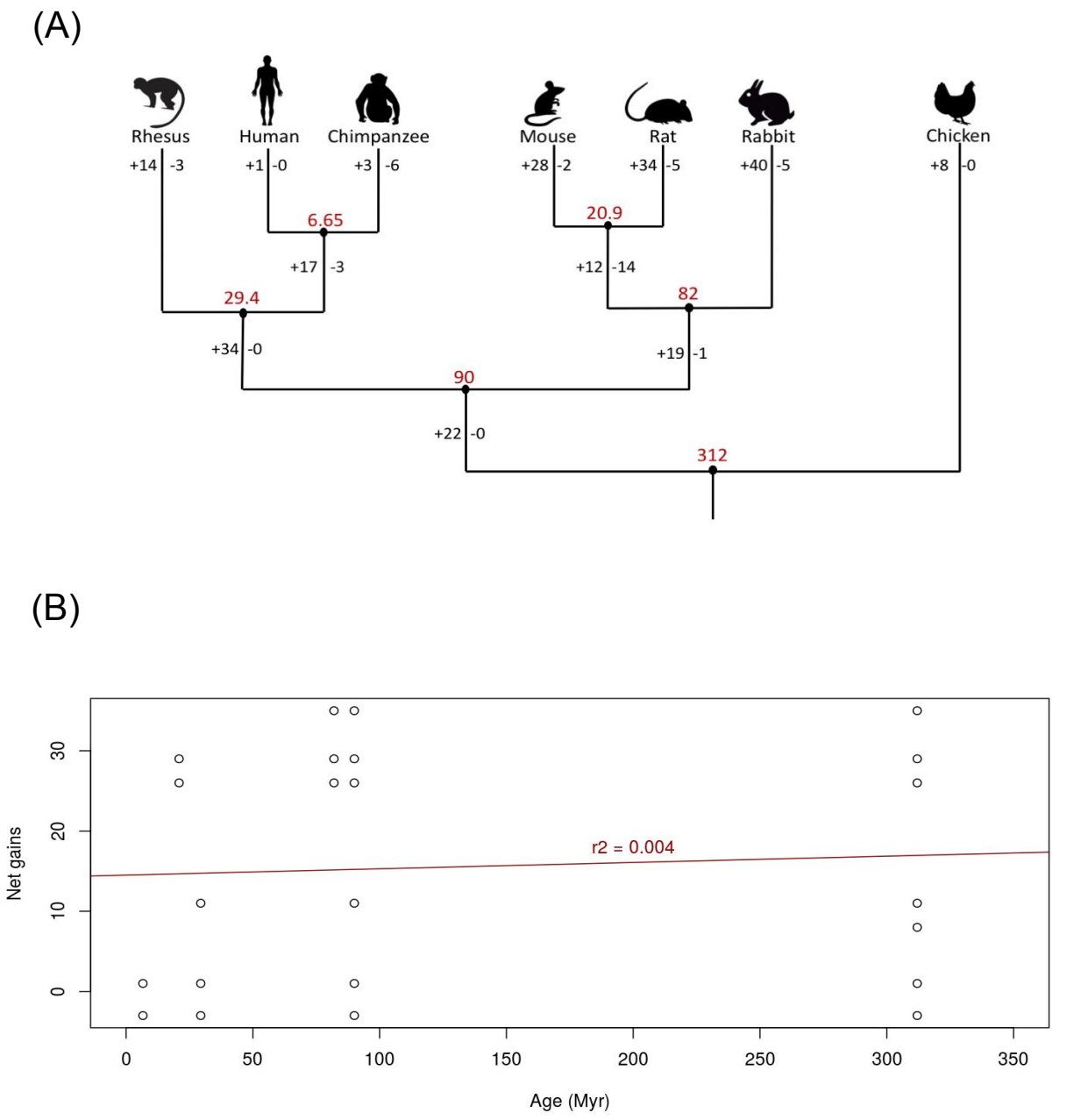

Figure 4.4 | Ancestral state of the regulation of Notch receptors by microRNAs: (A) seedVicious' predictions of the events of gains (+) and losses (-) of target sites of highly expressed and conserved microRNAs in Notch receptors. (B) Timeline of turnover by showing the net gains of microRNAs target sites in million years (Myr).

4.3. MicroRNA regulation of TRAIL signalling

TRAIL signalling (Tumor necrosis factor-related apoptosis-inducing ligand) is one of the most studied signalling pathways due to its immense importance in regulating apoptosis and thus cancer (Falschlehner et al., 2007). However, in regards with the signalling cascades, TRAIL does not maintain a signalling cascade like other networks studied in this work (ErbB, NOTCH, or Toll-like) as it first relies on one ligand only (TRAIL) and four receptors (*TRAILR1*, *TRAILR2*, *TRAILR3*, and *TRAILR4*). Interestingly, the conservation of some of the TRAIL receptors is restricted to primates (when compared to rodents and chicken). As a result, it was not possible to estimate gains and losses of microRNA target sites on such network.

4.4. The lengths of receptors 3'UTRs

Based on the results of the previous chapter, one of the questions that had to be answered in this chapter was whether receptors have special importance when it comes to microRNA-mediated regulation or not. In fact, unlike ErbBs, both Notch and Toll-like receptors are not favourably regulated by microRNAs. However, as microRNA-mediated regulation depends mainly on having microRNA target sites on the 3'UTR and the results of the analysis of ErbBs have shown a direct correlation between the length of 3'UTR and the density of target sites (i.e. the number of target sites per 100 nucleotides of length). Logically, a longer 3'UTR would mean more susceptibility of microRNA target sites on this 3'UTR and thus more microRNA regulation. In this part, I test whether receptors generally conserve longer 3'UTRs compared to non-receptors genes or not.

While comparing the lengths of the 3'UTRs of the receptors of the ErbB, Notch, and Toll-like networks to all other genes, only ErbB receptors have higher medians of 3'UTR lengths when compared to other genes in the network (figure 4.5). In fact, such bias in the median of the length of the 3'UTRs of ErbB receptors in mainly caused by two receptors (*ErbB1* and *ErbB4*). In addition, *TLR4* and *NOTCH2* stand out for conserving the longest 3'UTR among all other genes of the Toll-like and Notch networks, respectively. Interestingly, all of these receptors that conserve longer 3'UTRs are of significant biological importance. As mentioned previously, *ErbB1* and *ErbB4* are the only autonomous ErbB receptors (they can initiate the signalling independently, unlike *ErbB2* and *ErbB3*). In addition, *TLR4* is known for its pleiotropic nature as it is involved in the regulation of a broad range of processes and its mutations have been heavily

linked to a large number of diseases. These results might imply that some of the receptors might fall under strict pressure to conserve longer 3'UTRs accomodating more microRNA target sites allowing more strict regulation by microRNAs.

Based on these results and the established importance of receptors in cell signalling generally. I compare in the next analysis the lengths of 3'UTR of all human receptors to all other genes on the genome in order to test whether receptors are favourlibally regulated in general or not. In figure 4.6 while comparing the 3'UTRs of all the human receptors as reported by PANTHER (see Methods) to all other genes, there seems to be no significant difference between both groups which indicates that, unlike ErbBs, microRNAs do not necessarily favour the targeting of receptors. However, given the established importance of ErbB receptors the relatively-longer lengths of their 3'UTRs might indicate a special regulatory importance that resulted in longer 3'UTRs which will consequently accommodate more microRNAs target sites.

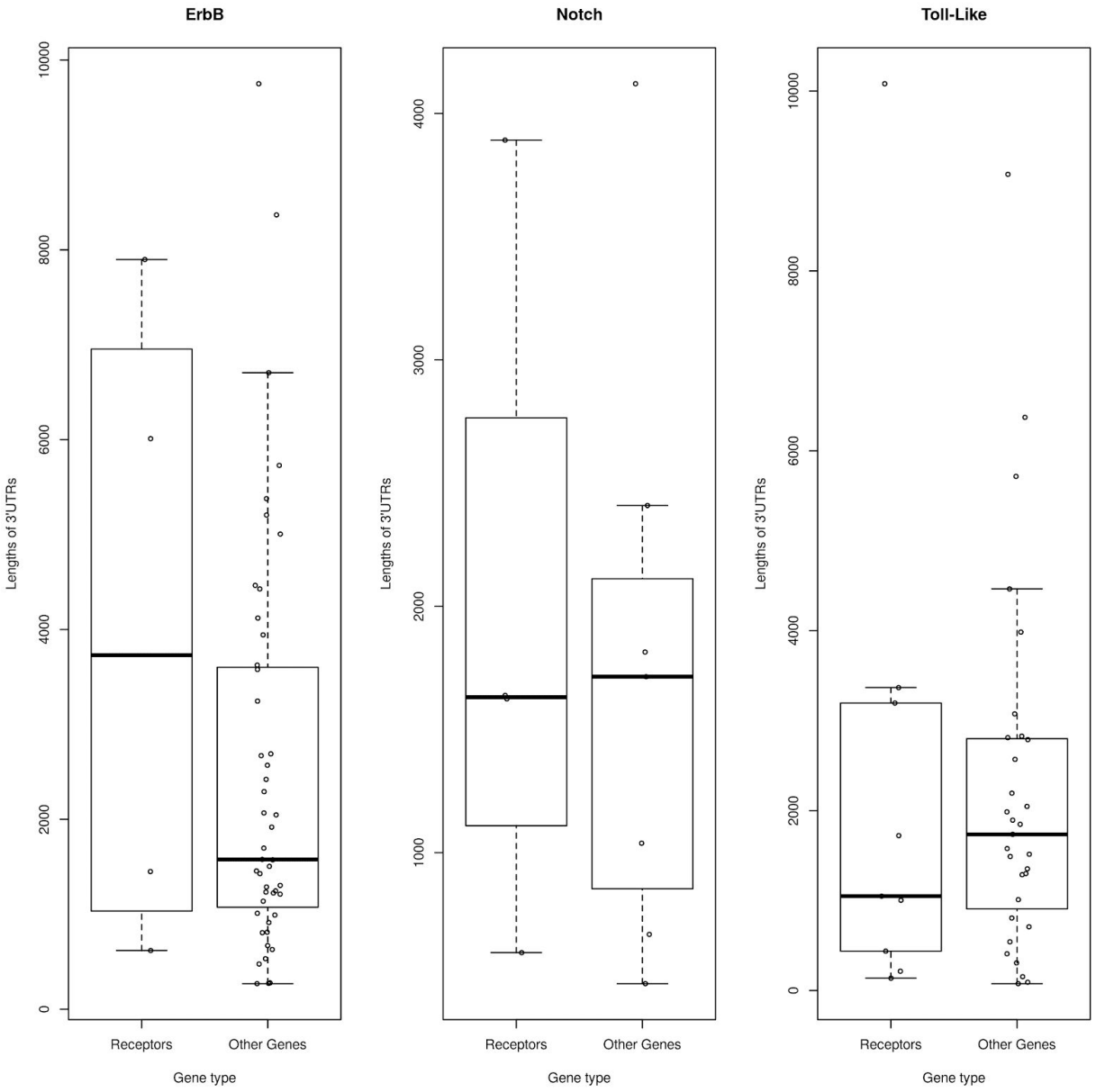

Figure 4.5 | The lengths of the 3'UTRs among the ErbB, Notch, adn Toll-like signalling networks: Boxplots showing the difference of the lengths between receptors and all other genes involved in the signalling of each network.

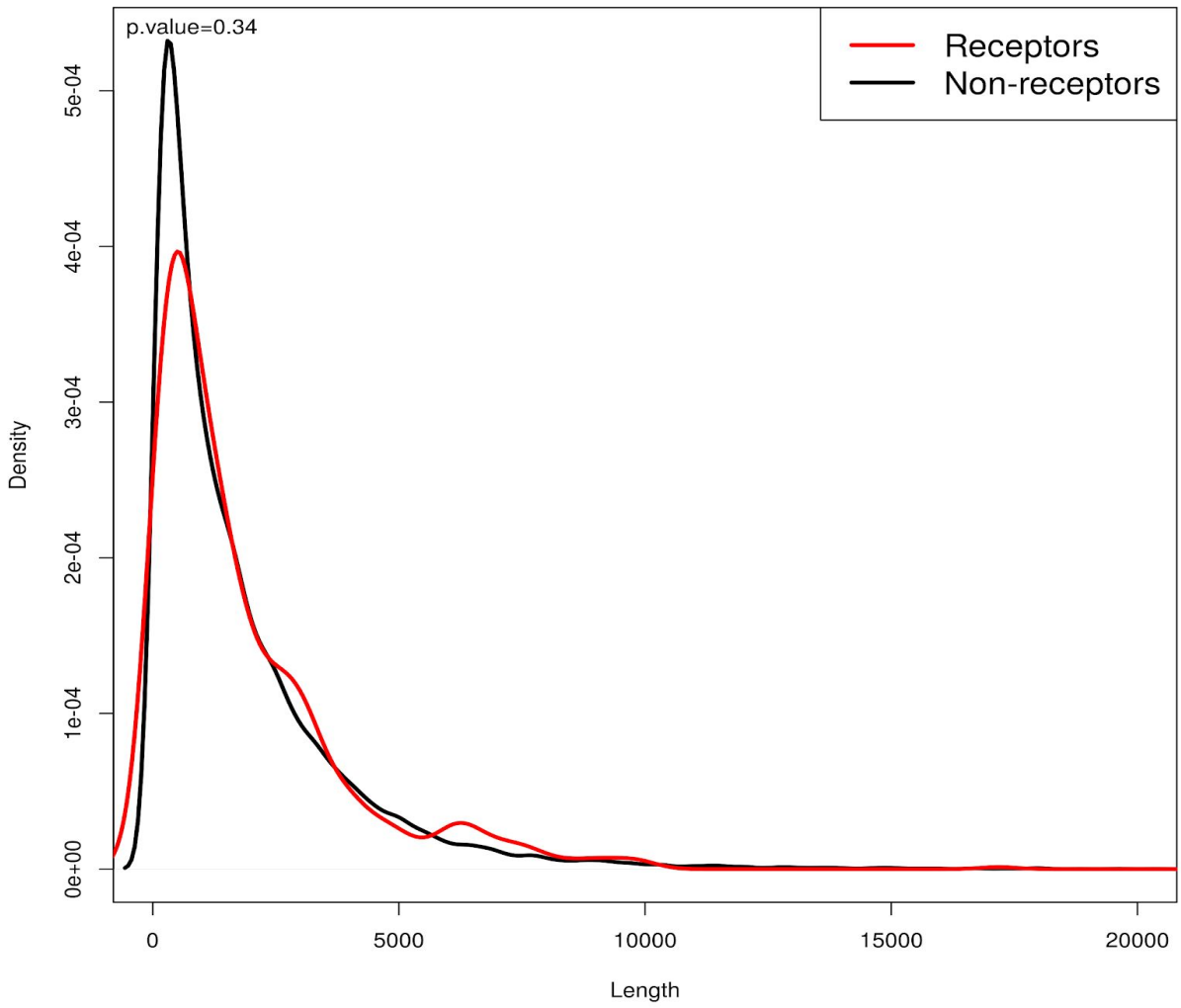

Figure 4.6 | Densities of 3'UTRs lengths: The differences in the densities of the lengths of the 3'UTRs of all human receptors (red) and all other human genes (black).

4.5. Discussion and conclusion

After exploring how microRNAs regulate the ErbB signalling network it was relevant to explore if the same patterns apply in other signalling networks or not. Hence, I chose three different networks that share some features with ErbB but are also unique in other ways.

One of the lessons learnt from assembling the ErbB network was the idea of levels of signalling, as based on Yarden and Sliwkowski (2001) results, signalling networks should actually be approached by the means of levels. For instance, if in a network like ErbB only two genes (out of 50 different genes) are under strict regulation by microRNAs, this might be considered neglectable. However, as it was the case in the previous analysis, these two receptors are the only autonomous ones. As a result, microRNAs do in fact -by regulating only these two genes- fine tune the whole network. One of the reasons of the choice of the Notch and Toll-Like networks was that each of them lack one of the signalling levels compared to ErbB (this being adaptors and ligands in Notch and Toll-Like respectively). It was interesting to explore whether the lack of one of these levels would put the layers up- or downstream under a higher regulatory pressure.

In the ErbB chapter, it was obvious that the very low precision caused by the use of all human microRNAs (stemming from the random complementarity) is suppressing the appearance of any regulatory patterns. However, upon limiting results to those who are highly expressed and have multiple target sites, some patterns appear (the favourable regulation of ErbB receptors for instance). However, applying a further filter (say, distances between targets and/or maximum energy levels) might as well be too strict.

To conclude, the analysis of the work of this chapter shows that microRNA-mediated regulation is not necessarily correlated with the level of signalling in a network. In addition, the relatively high levels of microRNAs target site turnover that was reflected in the gains and losses of microRNA target sites goes in-line with the results of the previous chapter. As a result, among different cell signalling networks, the regulatory interactions between microRNAs and their targets might not be as conserved as the microRNAs and the targets themselves. Finally, the analysis of the lengths of the 3'UTRs shows that only ErbB receptors conserve longer 3'UTRs when compared to other genes in their networks.

5. Tackling issues in microRNA target prediction

Aim: Since their discovery, a substantial amount of research has been done on microRNA which has led to their recognized importance presently. Accordingly, many microRNA target prediction software has been established, reviewed, enhanced, and optimised. However, there still seems to be much room for improvement. In this chapter, I provide a brief review of microRNA target prediction and I will, in line of my results, be providing new insights about the understanding of microRNA target prediction as well as highlighting some of the factors that need to be taken into consideration in microRNA target prediction.

MicroRNAs have come along way from being overlooked as one of the "junk" non-coding sequences to being regarded as one of the major regulators of gene expression that are believed to be able to regulate up to 60% of human protein-coding genes (Friedman et al., 2009). The relatively-recent interest in microRNAs, conservation, their roles, and their targets sparked a massive interest in microRNA research. The more microRNAs are researched, the more interesting they seem to be and the more the questions that need to be answered about them.

Target prediction lies in the heart of microRNA research (Peterson et al., 2014). While microRNAs themselves provide lots of insights on genomics and evolution, their ability to target (and possibly regulate) different genes, thus different biological functions, is always the fundamental aspect of microRNA research. Thus, efforts of establishing and continuously improving microRNA target prediction tools has been as old as microRNA discovery.

In animal microRNA research, targeting and regulating are not necessarily the same thing. In fact, having a seed-match is not enough neither necessary for a microRNA to be able to regulate its targets (Sætrom et al., 2007). On the one hand, a perfect seed-match on its own can not be considered a reliable indicator of regulation (Didiano and Hobert 2006). On the other hand, some interactions with G:U wobbles were reported to be able to regulate gene expression. Also, near-targets (target sites with a single nucleotide mismatch) may also be functional in regulating gene expression (Marco 2018).

Ultimately, a microRNA target interaction remains an indication of the regulation until it is experimentally validated. In fact, the relationship between microRNA target prediction and experimental validation is an ouroboros one, where computational prediction usually precedent experimental validation but also experimentally validated interactions are often used as a guide to train and optimize prediction software (Bradley and Moxon 2017).

Despite the huge advancement of the methods designed to predict microRNA targets, the false positive rate of most microRNA target prediction software remains greater than 0.3 (Zheng et al.,

2013). In fact, none of these software implies incorrect models, however, the nature of animal microRNA targeting implies that a sequence complementarity of as little as 6 nucleotides is, in principle, a possible regulatory interaction. This inevitably results in a large pool of predictions. Thus, in most microRNA target prediction software, the main improvement efforts are spent filtering for true positives. For this reason, microRNA target prediction usually needs to be supported by other features beyond the seed region complementarity. This factors include evolutionary conservation, hybridization energy, AU content of seed region, and structural accessibility of binding sites (Ding et al., 2016). While all these features are believed to be informative in predicting microRNA targets with limited number of false positives, there still seems to be a high number of false positives. In addition, strictly limiting predictions to those who possess such features, ultimately leads to a higher number of false negatives (see discussion chapter).

In addition to the aforementioned, two other features are known to be good predictors of a microRNA target prediction. However, other than seedVicious, these features are not widely used in microRNA target prediction software. Among such features is the number of target sites the microRNA has on the 3'UTR of its target mRNA. If a microRNA has more than one target site on the same 3'UTR, such target sites work cooperatively in order to either cleave or repress the translation of the target gene. Thus, microRNA with multiple target sites are believed to be robust regulators of gene expression (Agarwal et al., 2015). However, the effect of multiple target sites might be diminished in cases where the target sites are separated by large distances

on the 3'UTR (Sætrom et al., 2007). As a result, microRNA with multiple target sites that are in close proximity to each other are considered optimal for gene regulation.

To conclude, despite all the layers of checks used by microRNA target prediction algorithms, the high number of false positives remain a known issue that is holding back the progression of microRNA research in general (Zheng et al., 2013). Among the different measurements used to limit false positives, I here focus on two that have not been used in other software (than seedVicious) which are the number of target sites and the minimum distance between them. In this chapter, I use mirTaRBase (7.0) (Chou et al., 2017), the database of experimentally validated microRNA target interaction and seedVicious to explore the effect of the number of target sites and the distances between such targets on the precision of microRNA target prediction.

5.1. The use of non-functional transcripts among microRNAs target prediction software

The transcription of a DNA to an mRNA depends on a splicing machinery that transforms pre-mRNA to mature mRNA. Due to the presence of different combination of splicing variants, different mRNAs can be transcribed from the same gene in a process known as alternative splicing (Cooper 2004). In microRNA target prediction, complementarity is predicted between a given microRNA and the 3'UTR (or less-frequently, CDS or 5'UTR) of a transcript of the targeted gene. However, if more than one transcript exists per gene, microRNA target prediction software either use the most representative transcript which is decided by the presence of the highest number of 3P-seq tags, or the transcript with the longest 3'UTR to avoid missing possible target sites based on the logical assumption that the longer the 3'UTR the more the

target sites, or different transcripts per gene. The problem is, not all these transcripts are functional. In fact, a significant number of microRNA-target interactions presented by major pre-computed microRNA target prediction tools are based on the use of a non-functional type of transcripts known as nonsense-mediated decay (NMD). Nonsense-mediated decay is a surveillance quality-control mechanism that protects cells by selectively eliminating mRNA bearing premature termination codons (Lindeboom et al., 2016). In the following analysis, I highlight the presence of such transcripts among major microRNA target prediction software and the number of erroneous predictions caused by their use.

When counting for the presence of nonsense-mediated decay transcripts among TargetScan (7.2) (Agarwal et al., 2015), DIANA-microT-CDS (v5) (Paraskevopoulou et al., 2013), RNA22 (v.2) (Miranda et al., 2006), miRanda (Enright et al., 2003), miRDB (v5.0) (Wong and Wang 2014), and seedVicious (v1.1) (Marco 2018), there seems to be a large number of NMD transcripts especially among DIANA-microT-CDS and RNA22 software. In addition, since each of these transcripts, can (and usually will) be targeted by many microRNA, this results in a large number of interactions that are based on non-functional transcripts even among software that include less numbers of NMD transcripts like TargetScan and miRDB. In fact, depending on the software, some of these predictions might even be given a high prediction score (in case of low hybridization energy or high degree of conservation for instance) (Table 1).

In line with these results, I believe that the choice of a transcript in microRNA target prediction should always be given enough attention in order to avoid the use of non-functional ones as well

as to use the most relative transcript according to the tested cell-line, disease status, phenotype, and any other factor that may favour the use of one transcript over another. Finally, based on these results, only protein coding transcripts have been used in seedVicious predictions throughout this work.

Table 1 | NMD transcripts in microRNA target prediction software

Software	NMD Transcripts	Affected interactions
DIANA-microT-CDS	2807	1142030
RNA22	2484	9844520
TargetScan	23	8960
miRDB	14	385
miRanda	3	24
seedVicious	0	0

5.2. The effect of numbers of target sites and distances on the precision of microRNA target prediction

I here explore the effect of the number of target sites and the distances between them on the precision of microRNA target predictions compared to experimentally-validated interactions obtained from miRTarBase. For this, I calculate the precision as a ratio of true positives (experimentally validated interactions) and false positives (predictions that are not validated experimentally) (see Methods).

In figure 5.1, the precision of human microRNA target prediction increases as the number of target sites increases till it reach a peak point at 11 target sites per interaction. Precision starts to decline at higher numbers of target sites due to the low number of interactions. In addition, when counting for the distances between such pairs, the precision seems to be inversely proportional to the distances between pairs (i.e. the shorter the distance between target sites, the higher the precision). In fact, precision is clearly highest in cases of adjacent target sites (those who are 6 nucleotides apart, where the 1st and 2nd nucleotide of the second target sites overlaps with positions 7 and 8 of the first one). Finally, when combining both number of target sites and distances between pairs, the precision of targeting prediction seems to be optimal for interactions that have between 6 and 20 target sites and are located adjacently.

Likewise, while exploring the precision of the predictions of mouse microRNA targets. The precision again seems to be correlated with the number of target sites till it reach an optimal point at 7 target sites and it remains considerably high till 20 target sites. While for the distance,

the precision again seems to be highest at closer distances. However, unlike human, the optimal distance for precision is 8 nucleotides between target sites. However, combining number of target and distances, shows an optimal area of precision for microRNAs that have between 4 and 13 target sites with a less influence of distances (figure 5.2).

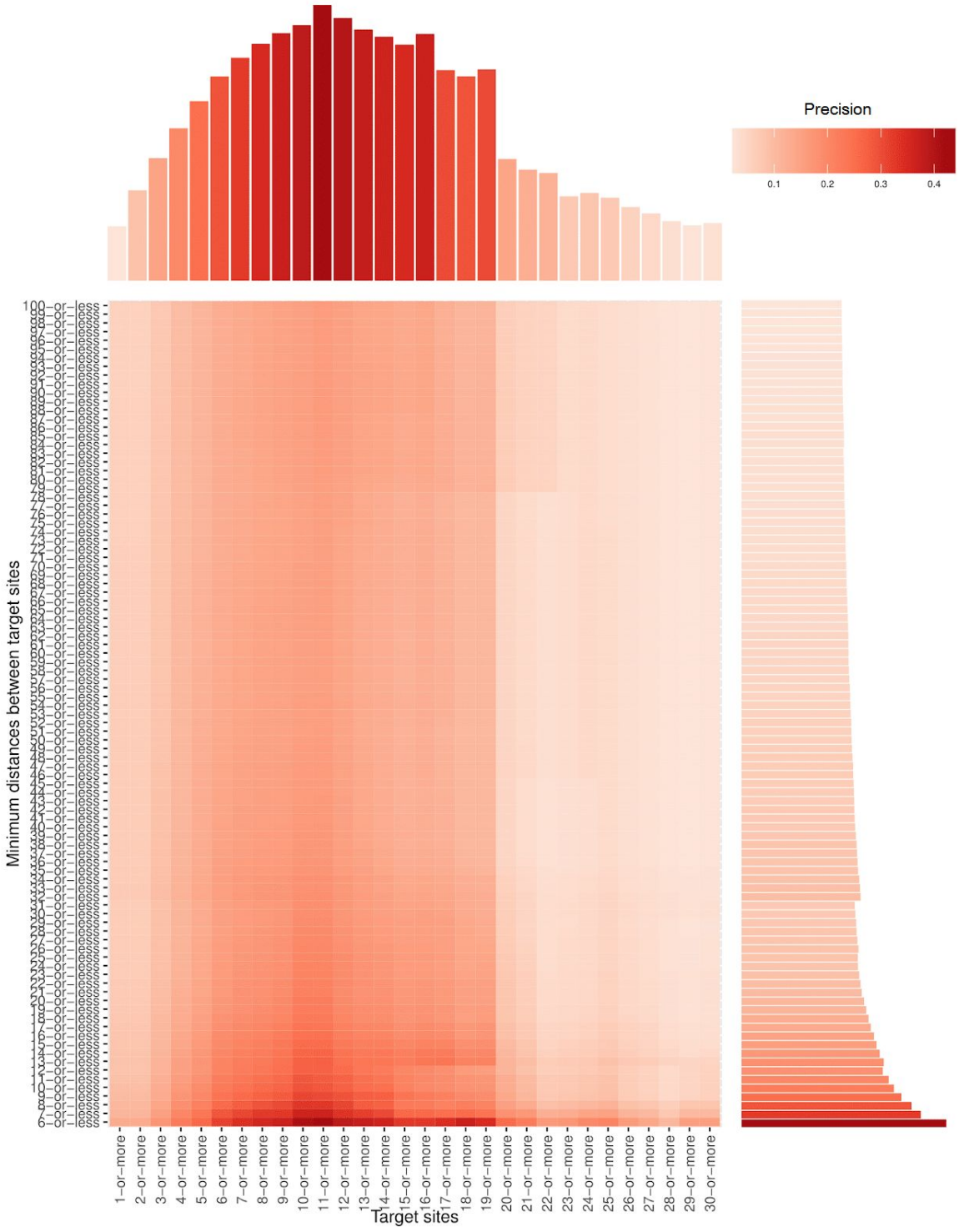

Figure 5.1 | Precision of human microRNA targets predictions: Heatmap showing the precision of the predictions of microRNA in regards with the number of target sites and the minimum distances between such targets. Top bar plot shows the precision based on the number of target sites only while side bar plot shows the precision based on minimum distance between targets.

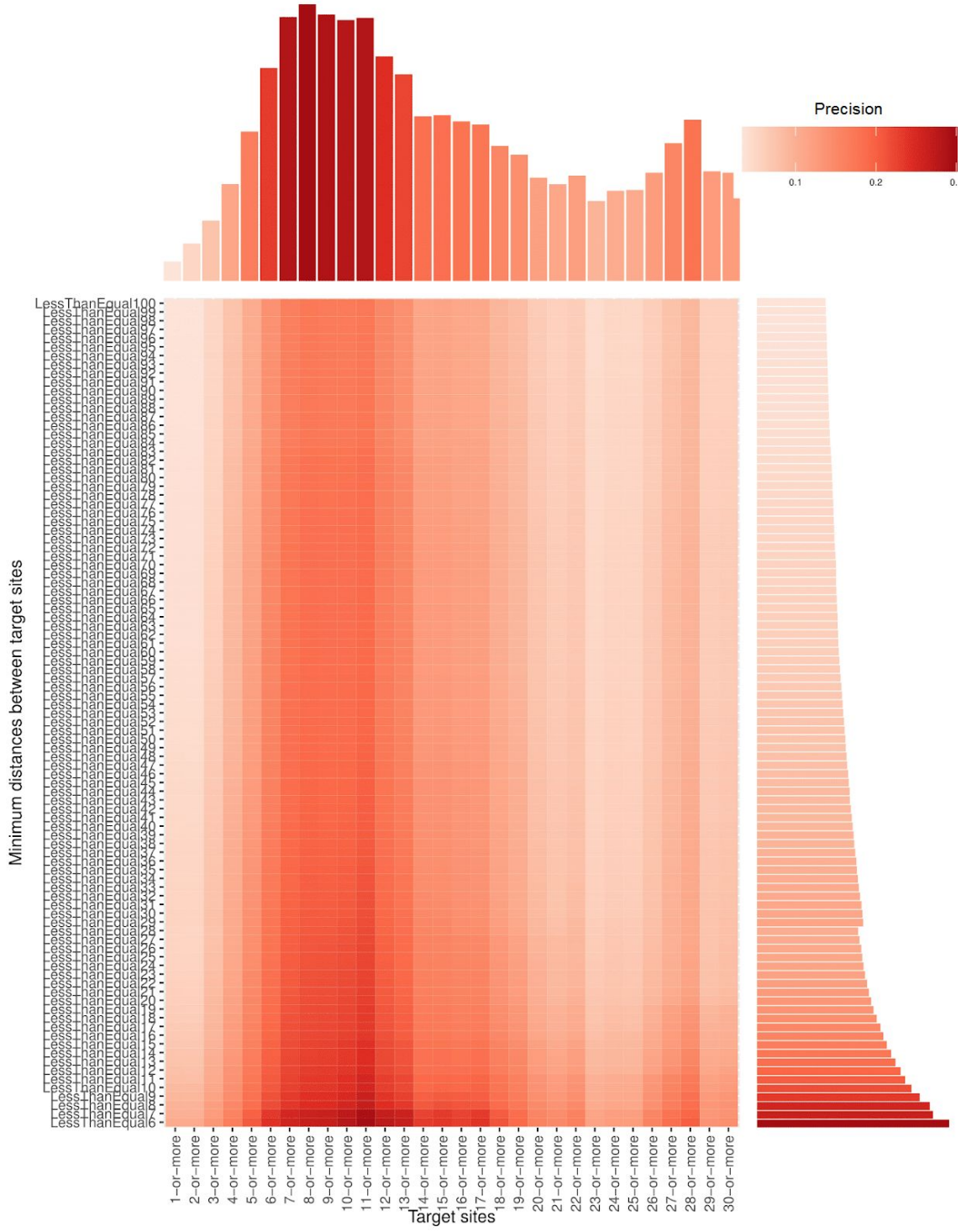

Figure 5.2 | Precision of mouse microRNA targets predictions: Heatmap showing the precision of the predictions of microRNA in regards with the number of target sites and the minimum distances between such targets. Top bar plot shows the precision based on the number of target sites only while side bar plot shows the precision based on minimum distance between targets.

5.3. The effect of seed region complexity on the precision of microRNA target prediction

It has previously been noted that the presence of low-complex sequences (those with repetitive dinucleotide motifs) among microRNA seeds regions affects the precision of microRNA target prediction as they are more likely to be detected by high throughput experiments like HITS-CLIP and PAR-CLIP (Chandradoss et al., 2015).

To test the effect of such low-complex microRNA on the precision of microRNA prediction, I first identify microRNAs whose seed-regions consists of only repetitive dinucleotides (for instance, the seed region of hsa-miR-190a-3p is UAUAUAU). In total, there was 29 low-complex human microRNAs and 15 mouse ones.

In order to explore the effect of these microRNAs, I calculate the precision of the three groups; low-complex, all microRNA and high-complex ones (all microRNA after excluding the low complex ones) in human and mouse. In figure 5.3A and 5.3B, the precision of low-complex microRNA is clearly higher in both human and mouse and even while counting for only microRNA with multiple target sites. Such result implies the huge effect played by such small number of low-complex microRNA on the precision of experimental validation of microRNA in general. Next, as low-complex microRNA are believed to be detected significantly in high throughput experiments, I repeat the analysis after excluding validated predictions that are based on HIT-CLIP and PAR-CLIP experiments. Interestingly, as shown in figure 5.3C and 5.3D, the precision of low-complex microRNA is much lower than high-complex ones.

5.4. The precision of high-complex microRNA target prediction

In order to avoid the experimental bias of low-complex microRNAs, I here explore the effects of target sites and distances using only microRNA that are highly complex. In figure 5.4, the effect of the number of target sites remains similar to that of all microRNA (increases till it reaches a peak precision at 11 target sites). In contrast, there is not any clear correlation between the precision and the distances between targets.

In figure 5.5, the patterns of precision of highly complex microRNA in mouse seems to be similar to those of all microRNA of optimum precision between 3 and 17 target sites and distances between 6 and 10 nucleotides. However, in both human and mouse, the general precision of high complex microRNA is less than that of all microRNA (as shown in figure 5.3). However, the patterns of target sites remains similar, while the distance seems to have minimal effect on the precision in human.

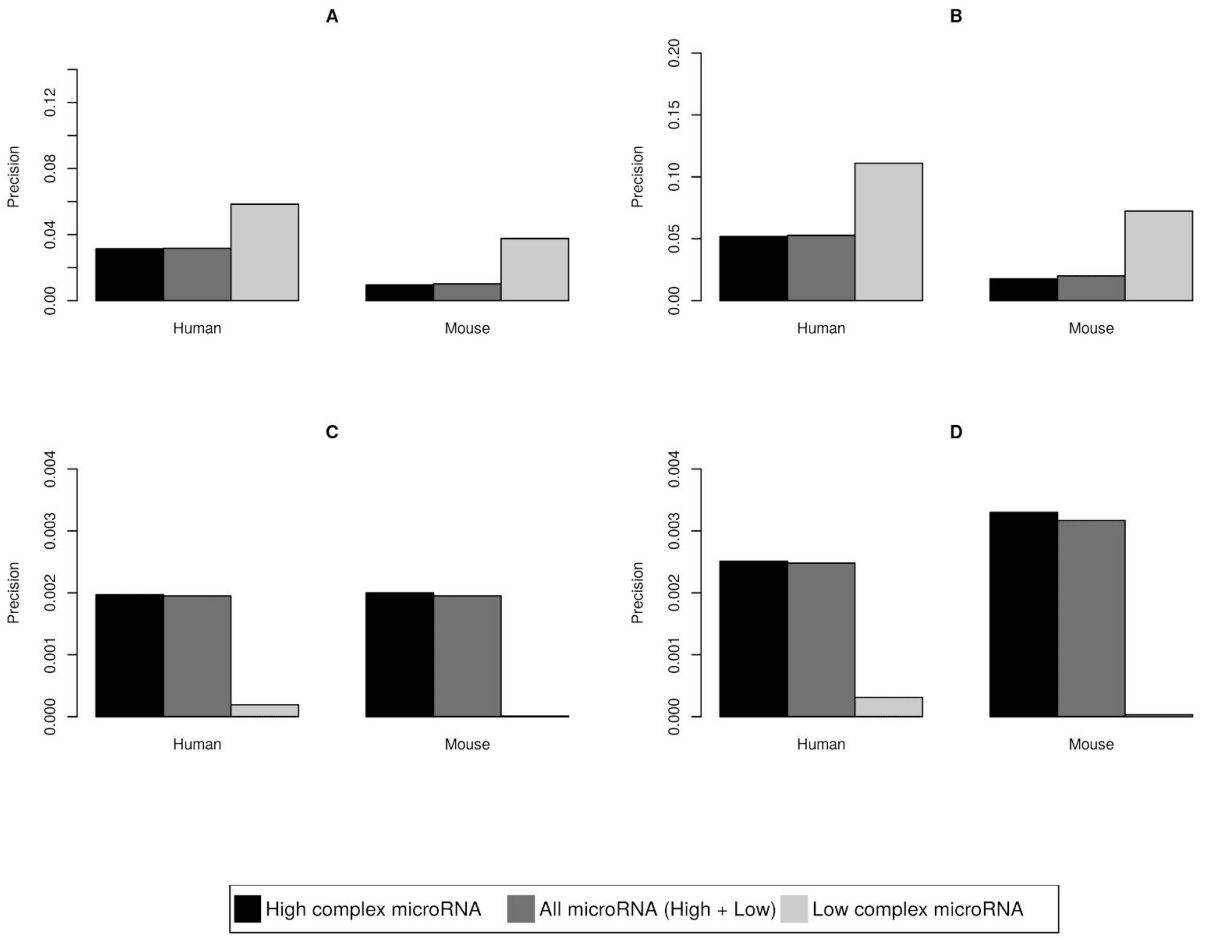

Figure 5.3 | The precision of high and low complex microRNA: (A) The difference in precision between high-complex microRNA, low-complex ones, and both (all microRNA). (B) The difference in precision after restricting to only microRNA with multiple target sites. (C) The precision of all, high, and low complex microRNA after excluding interactions that were validated using PAR-CLIP and HIT-CLIP. (D) The precision after limiting to only microRNAs that have multiple target sites and are not validated by PAR-CLIP and HIT-CLIP methods.

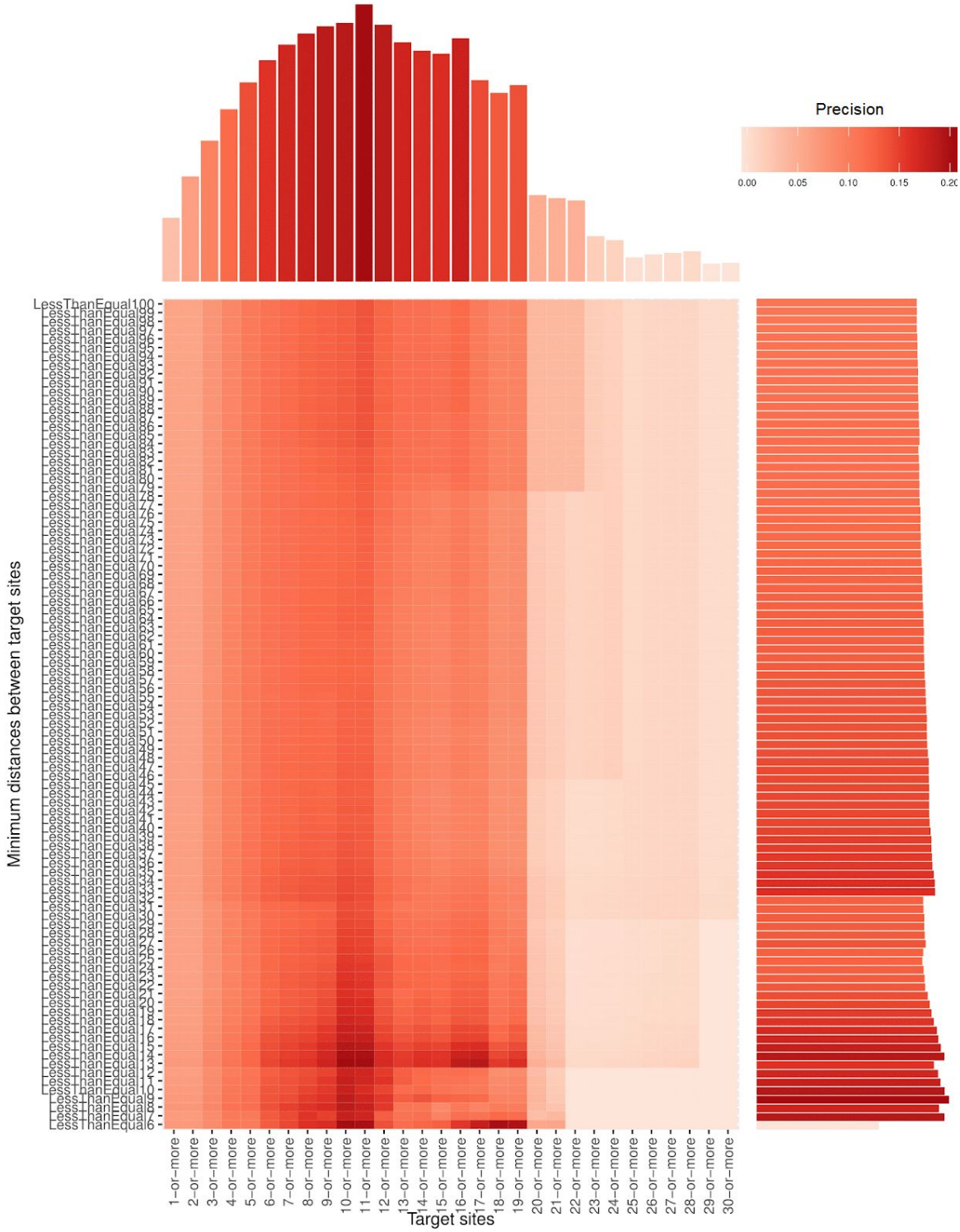

Figure 5.4 | Precision of human complex-microRNA targets predictions: Heatmap showing the precision of the predictions of microRNA in regards with the number of target sites and the minimum distances between such targets. Top bar plot shows the precision based on the number of target sites only while side bar plot shows the precision based on minimum distance between targets.

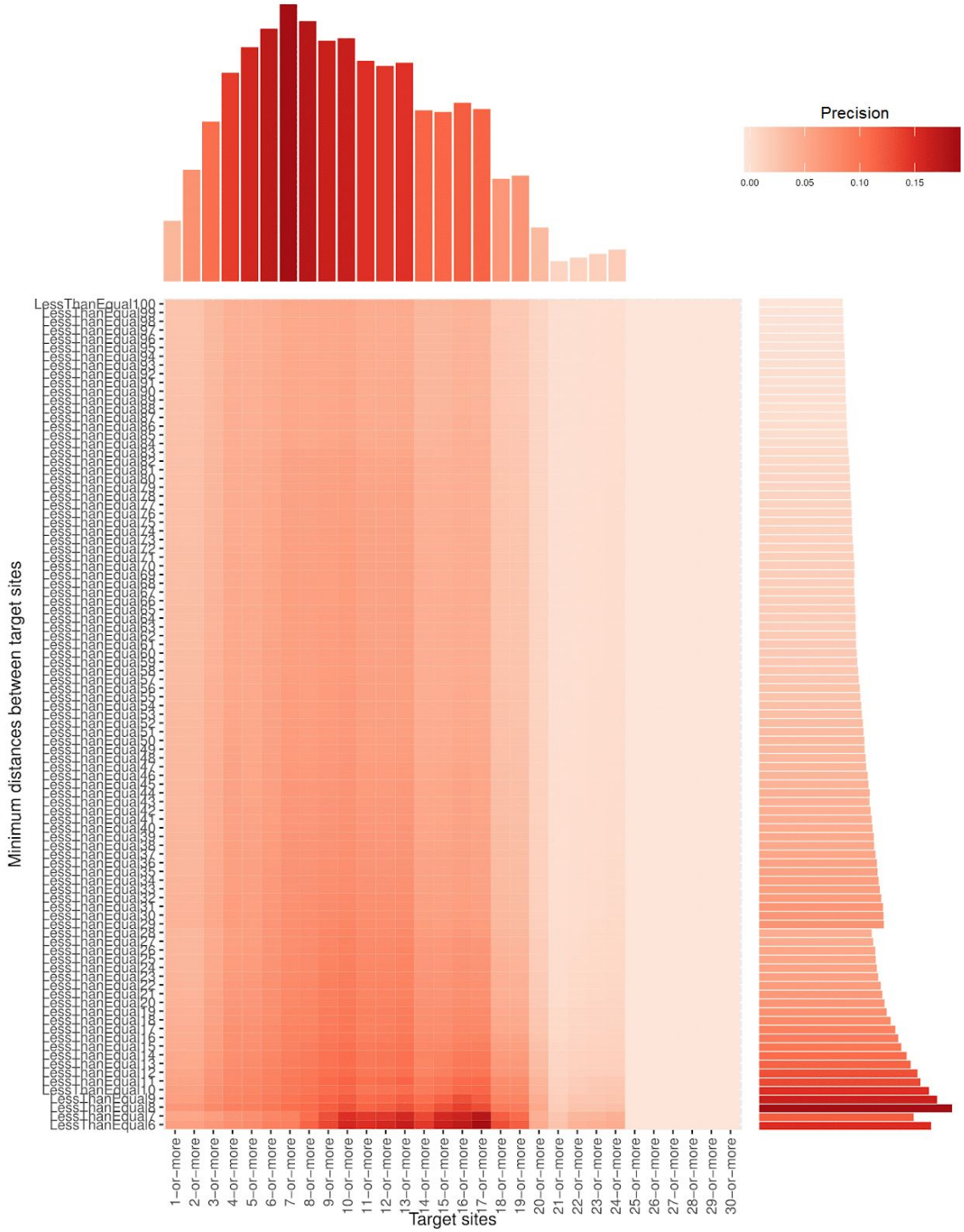

Figure 5.5 | Precision of mouse complex microRNA predictions: Heat map showing the precision of the predictions of microRNA in regards with the number of target sites and the minimum distances between such targets. Top bar plot shows the precision depending on target sites, side bar plot shows the precision based on minimum distance between targets.

5.5. The precision of highly-expressed microRNA target prediction

Many of the experimental methods used for the validation of microRNA target interaction insure that the microRNA in a physical contact with its candidate target. If the microRNA was able to repress the expression of the target gene, the targeting is considered valid. However, biologically, the story is different. In a living system, for the microRNA to be able to interact with and effectively regulate the expression of a gene, it must be expressed (and in high levels) at the same tissue. For this reason, in the following analysis I test the effect of numbers of target sites and distances between them on the precision of the prediction of microRNA that are highly highly expressed (484 microRNA in human and 485 in mouse).

The precision of the prediction of such highly-expressed microRNA increases as the number of target sites increases reaching a peak of 11 target sites. In addition, the precision is highest for microRNAs that have target sites that are 6-10 nucleotides apart (figure 5.6). However, in figure 5.7, the precision of highly-expressed mouse microRNA seems optimal for microRNA that have between 4 and 7 target sites. However, while compiling both target sites and distances, the precision seems to be highest for microRNA that have 13 or 17 target sites that are 6-10 nucleotides apart. Generally, the analysis of highly expressed microRNA might be affected by the rather small number of microRNA tested (in contrast to all microRNA or even high complex ones).

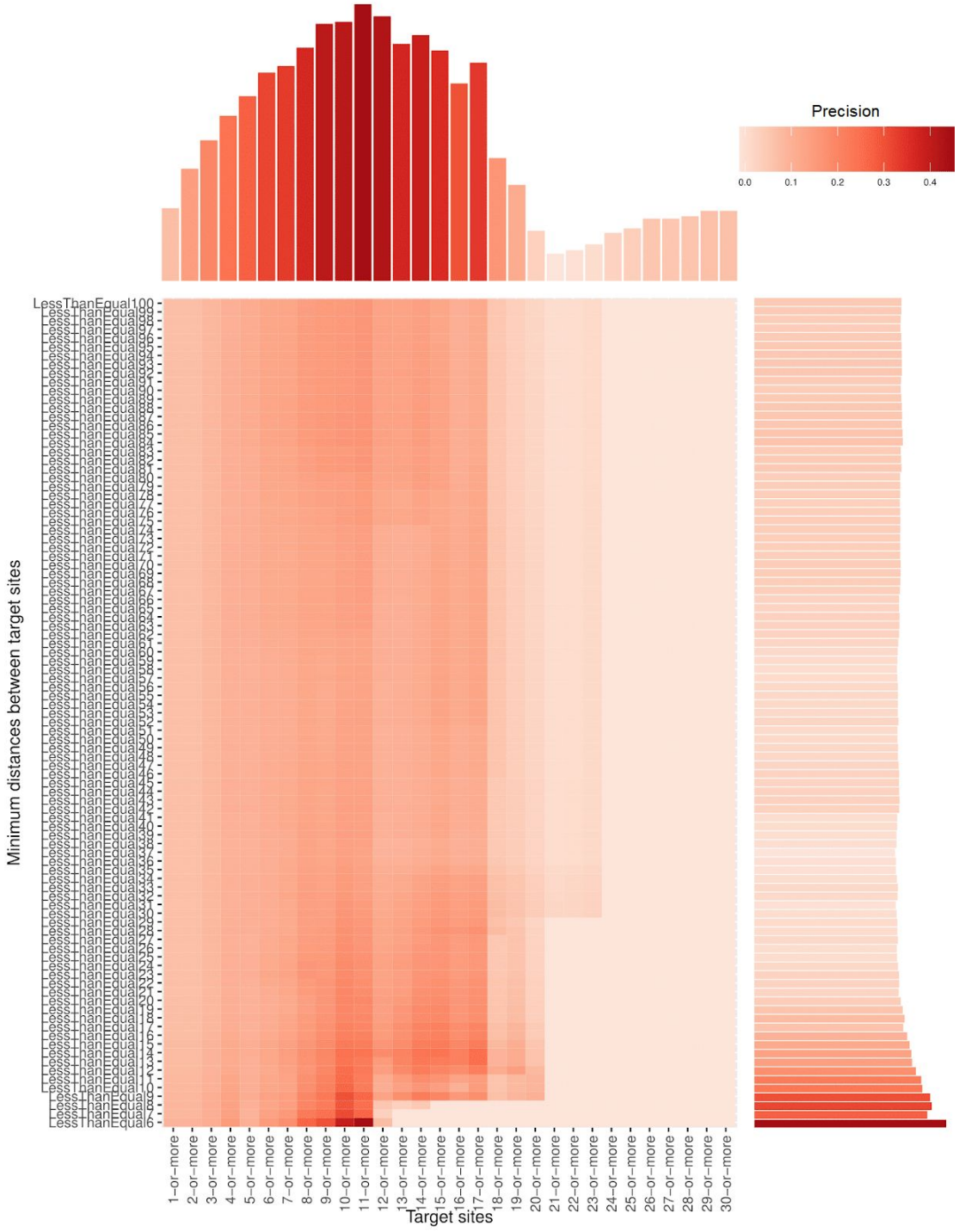

Figure 5.6 | Precision of human highly-expressed microRNA predictions: Heat map showing the precision of the predictions of microRNA in regards with the number of target sites and the minimum distances between such targets. Top bar plot shows the precision depending on target sites, side bar plot shows the precision based on minimum distance between targets.

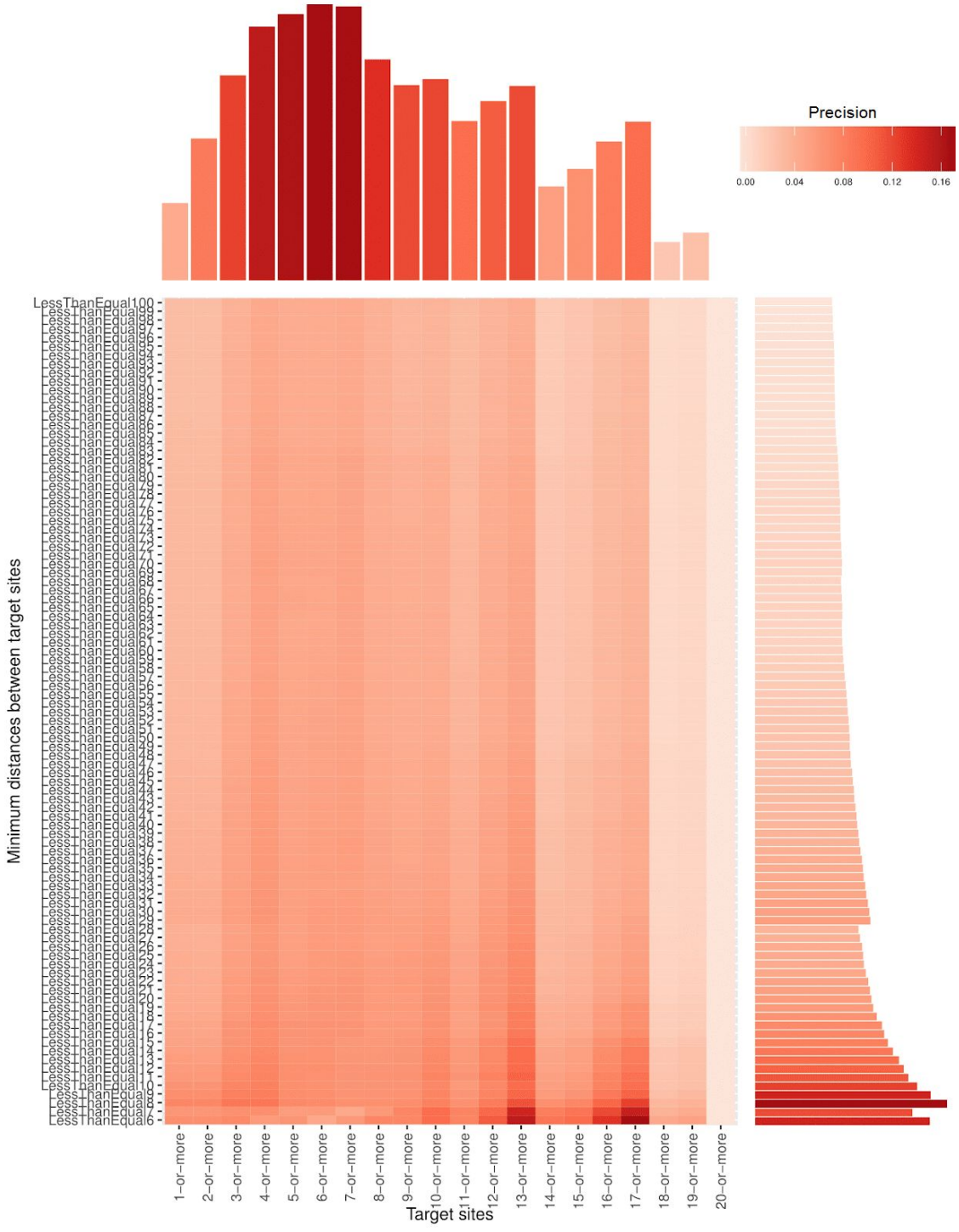

Figure 5.7 | Precision of mouse highly-expressed microRNA predictions: Heat map showing the precision of the predictions of microRNA in regards with the number of target sites and the minimum distances between such targets. Top bar plot shows the precision depending on target sites, side bar plot shows the precision based on minimum distance between targets.

5.6. The effect of free energy on the precision of microRNA target prediction

Thermodynamic stability is a fundamental aspect of all biological interactions. As a result, the minimum amount of free energy produced by the binding of the microRNA to its target is used to assess the strength of the microRNA:target duplex where the less the energy the more potent the regulation (Riffo-Campos et al., 2016).

In the next analysis, I use seedVicious to compute the free energies of such duplexes where a microRNA has only one target site (to avoid the bias caused by the number of target sites and/or distances between them on the precision). seedVicious' predictions were grouped into bins of 5 minimum free energy (Kcal/mol) in width ranging from -55 to 15. The precision of each bin was then calculated using the findings of miRTarBase. The precision of the prediction is inversely proportional with the amount of free energy produced by the interaction where the lower the energy the higher the precision. However, according to these results, the precision of such predictions seems to be optimal when the value of the minimum free energy produced is between -50 and -30 (Kcal/mol) (figure 5.8A). In fact, the common belief in microRNA target prediction is that the maximum value of free energy for a thermodynamically-stable interaction should not exceed -15 Kcal/mol (Ghoshal et al., 2015). As a result, I use the threshold of -15 (Kcal/mol) to split seedVicious' predictions to two groups, (i) less than or equal -15 (Kcal/mol), (ii) more than -15 (Kcal/mol). The precision for each bin was then calculated. In figure 5.8B, the precision is unsurprisingly higher for interactions with no more than -15 Kcal/mole of minimum free energies. However, the lack of significance between the two groups suggests that it might be useful to set a more strict threshold value for more precise microRNA target prediction.

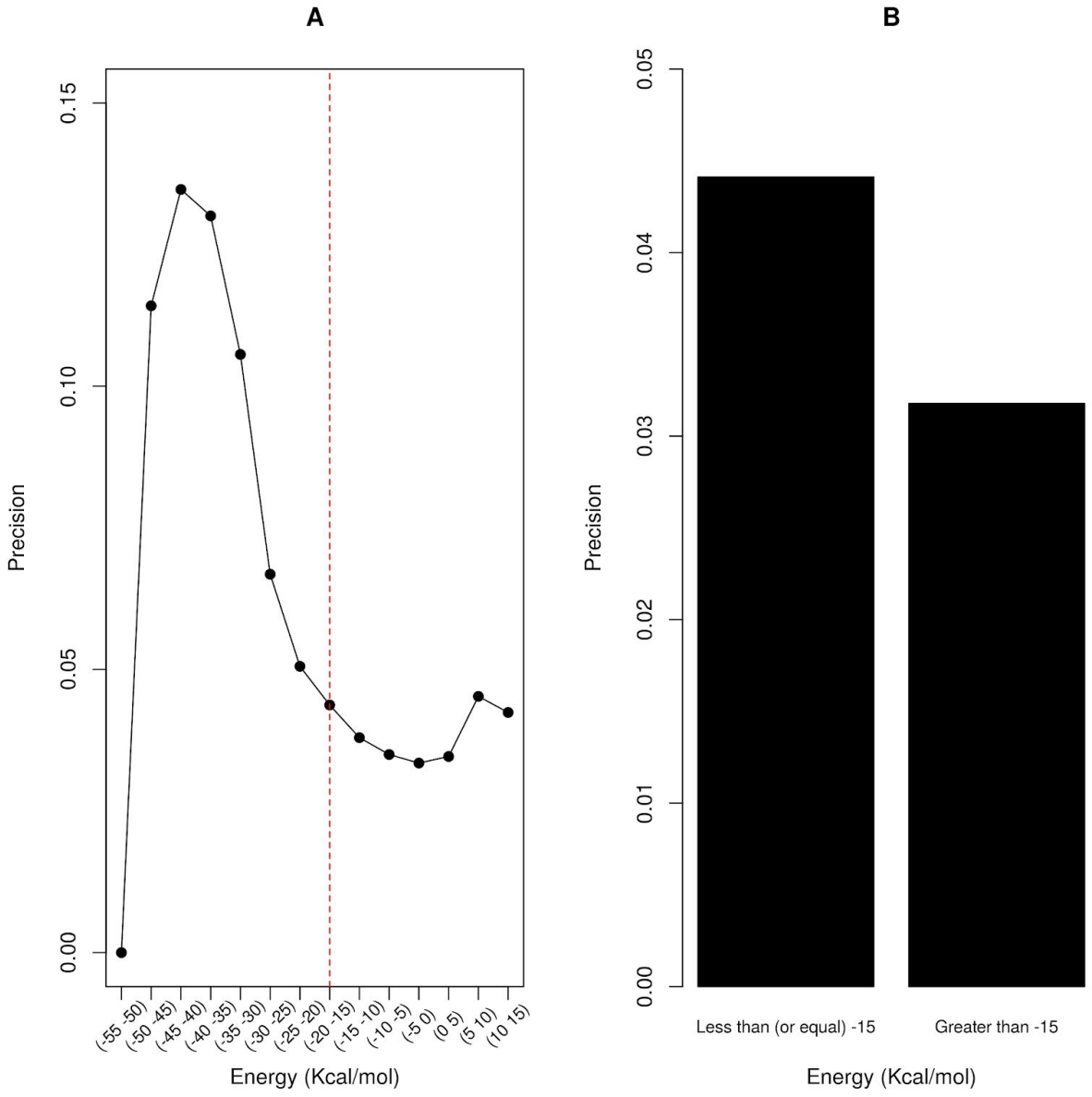

Figure 5.8 | The effect of free energy on the precision of microRNA target prediction: (A) The precision of microRNA target predictions per bin of 5 Kcal/mol of minimum free energy, the red line resembles the -15 Kcal/mol threshold of stable interactions. (B) The difference in precision between interactions that has less than (or equal to) -15 Kcal/mol and those that has higher than -15 Kcal/mol of minimum free energy.

5.7. Discussion and conclusion

Throughout this work, I have been extensively using different tools and features for microRNA target prediction. However, the high percentage of false positives is a known issue in microRNA research in general. In order to minimize the impact of such issue, I have compiled some further layers of checks in predicting microRNAs targets. Thus, in this chapter I highlight some of the issues in microRNA target prediction as well as providing possible answers to the precision problem that might help in minimising the occurrence of false positives using current technologies.

First, I highlight the use of non-functional NMD transcripts in major microRNA target prediction software which results in many erroneous predictions affecting many studies. Secondly, I highlight the impact of the number of target sites and the minimum distances between such targets. In addition, I show that the reliance on high-throughput methods of experimental validation of microRNA target interactions, despite being efficient, might be producing another layer of false positive results. However, in such cases the false positives are experimentally validated ones. Finally, comparing the precision in regards to the amount of free energy produced by the microRNA-mRNA duplex confirms the negative correlation proposed between the strength of the regulatory duplex and the minimum amount of free energy produced.

Producing the results in this work would have not been possible without the use of seedVicious as many of the factors studied here are only reported by seedVicious. On the other hand, the results of the work provided in this chapter has been taken into consideration in updating

seedVicious. For instance, in addition to including common features of microRNA target predictions software, seedVicious doesn't include any NMD transcripts. Furthemore, seedVicious restricts the occurrence of false positive predictions by providing counts of target sites per microRNA as well as the minimum distance between pairs of target sites. Also, seedVicious can provide a history of the gains and losses of microRNAs target sites. A feature that, in addition to providing the evolutionary insights, can be used to minimize the false negative predictions.

In conclusion, the importance of microRNAs as novel regulators of gene expression is beyond any doubt. For this reason, there exist many microRNA target prediction software that vary in nature, publicity, and quality. Unfortunately, none of these software is considered a trustable predictor of microRNA target sites. In fact, it is currently believed that for a microRNA target prediction to be considered robust, the same prediction must be reported by more than one software (Oliveira et al., 2017). I believe that, the advances in microRNA identification are yet to be met by a strong, reliable, gold-standard microRNA target prediction software. Alternatively, and more feasibly, it might be practical to shift into a more specialised microRNA target prediction paradigm in which some specific software might be optimized to predict microRNA targets in regards to a specific species, tissue, biological function, or disease.

6. Discussion

Some 600 million years ago, multicellular organisms evolved from their unicellular ancestors and the preservation of complex cell signalling mechanisms was at the heart of this rise in organismal complexity (Brunet and King 2017). After all, how can an organism function if its cells each act alone. However, cell signalling was not the only factor involved in the rise of multicellularity, but microRNAs were too. In fact, both a unicellular organism and a single cell of a multicellular one excel in what they do. A unicellular organism has, through years of evolution and accumulated mutations, been able to master the ability to conserve and consume energy to reproduce and thus to survive. However, one of the major differences between multiple unicellular organisms and multiple single cells in a multicellular organism is the ability of the later to differentiate. Here comes the obvious role of microRNAs. As cells differentiate they need to stay under control and microRNAs have profound roles in differentiation and cell-fate decisions (Ivey and Srivastava 2010) as well as in the regulation of gene expression. However, microRNAs also offer the advantage of being tiny in size which makes them more likely to mutate and evolve. As a result, the rise of microRNAs might have offered a rather vital role in the transition to multicellularity. In fact, despite one case (Zhao et al., 2007), microRNAs have only been detected in multicellular organisms thus far. Interestingly, microRNAs despite being conserved among the majority of multicellular organisms, have evolved independently in both animals and plants (Axtel et al., 2011). The work contained in this thesis aimed to discover how regulatory interactions between two of the fundamental aspects of life takes place.

"Cell signalling is the music of life", the title of a recently published article might illustrate both the importance and complexity of one of the principal topics of modern biology, cell signalling (Hancock 2008). Unlike microRNAs, cell signalling has been discovered, reviewed, and valued for so many years. However, the more cell signalling is studied, the more complex it appears to be and the more the known unknowns there seem to be (Hancock 2010).

In fact, cell signalling should not be a simple process. It is cell signalling that orchestrates the entire functioning of living organisms, one of the most complex machines that ever been created (Dawkins 1976). It was cell signalling that allowed multicellular organisms to evolve from their unicellular ancestors (Alberts et al., 2002). It is also up to cell signalling to ensure the development of functioning living organisms and, when things go wrong for these organisms, it is cell signalling that has to respond to pressure and/or suppress disease. However, cells are always bombarded by a wide range of signals, which ones do they respond to and which ones do they ignore is what decides the fate of such cells and the whole organism they are governed within. As the field of cell signalling developed, cell signalling started to be looked upon in terms of complex pathways and networks rather than the ancient simple approach (Hancock 2010). In fact, the terms cell signalling pathways and cell signalling network are sometimes used interchangeably. However, throughout this work the focus has been on studying how cell signalling take place in networks. Cell signalling networks are often a result of crosstalk between different pathways that result in diverse cellular actions to the same signal. Thus, in this work the focus has been on how microRNAs might regulate a network as a whole, rather than a

single pathway within a network that might turn redundant by the presence of another unregulated pathway.

The results of the study of ErbB, one of the most studied cell signalling networks, have revealed a few interesting features of microRNA regulation. First, the previously mentioned importance of ErbBs receptors have been reflected in having significantly longer 3'UTRs as well as being the level most-strictly regulated by microRNAs. In addition, despite the strong conservation of both microRNAs and ErbBs receptors and ligands, the interactions between them are not as conserved, a variation which is sometimes caused by a single nucleotide change. Furthermore, the analysis of the gains and losses of these interactions showed a clear bias towards the primates lineage. After studying ErbBs signalling, it was important to test whether the regulatory patterns of microRNAs apply to other cell signalling networks or are limited to ErbB. For such reasons, while incorporating the lessons learnt from ErbB analysis, microRNA regulation of the Notch and Toll-like signalling networks were studied. While the regulation of the networks by microRNAs in regards to the signalling levels did not show any significant patterns (unlike the receptors bias in ErbB), it was interesting to notice the conservation of the primates-bias in the Notch signalling network. However, the results of the analysis of both Notch and Toll-like signalling networks have rather emphasised the unique features of the ErbB network which indeed adds to its established importance and complexity.

All the conclusions of how microRNAs regulate cell signalling and how this is conserved, are actually based on microRNA target prediction and validation. In order to explore the regulatory

roles played by microRNAs, it was essential to predict the interactions between microRNAs and the genes involved in these networks. Thus, the robustness of these predictions is vital to our understanding of the results of this work and any other work that involves microRNA regulation. For this reason, throughout this work, several layers of filters have been applied in order to keep microRNA target prediction as precise and realistic as possible. First, microRNAs expression levels have been taken into consideration where microRNAs that have low (or non) expression levels being omitted in some analysis. Furthermore, the numbers of microRNA target sites and the distances between them has been considered and only functional mRNA transcripts have been used.

As cells proliferate, they become at risk of acquiring mutations that might turn uncontrollably-proliferating cells to malignancies (Cooper 2000). Thus, theoretically, more morphologically-complex animals (i.e. those with higher number of cells and more distinct cell types) should be at higher risk of developing cancers. In addition, assuming that mutations accumulate over time, animals with longer life spans should develop more cancers as an inevitable result of the accumulation of such mutations. Surprisingly, the exact opposite is observed. In fact, there appears to be absolutely no correlation between the size and longevity of an animal and its susceptibility to developing cancers as smaller animals (with shorter lifespans) develop much more cancers than larger, long-living ones (Caulin and Maley 2011). Such a striking observation is referred to as Peto's Paradox as it was first noted by epidemiologist Richard Peto in 1975 (Peto et al., 1975). Since then, much research has been done trying to resolve Peto's paradox. Among the possible explanations, one theory relates this observation to

only one gene, *TP53*. *TP53* is a tumor suppressor gene that induces apoptosis (Hollstein et al., 1991). While comparing human and elephant genomes, it has been reported that elephants conserve 20 copies of *TP53* compared to only one copy in humans which might provide an explanation to why humans are at higher risk of developing cancer compared to elephants (Abegglen et al., 2015). However, the fact that whales -which are larger and develop fewer cancers than elephants- also conserve one copy of *TP53* makes it appears more like an elephant-specific way of suppressing cancers (Tollis et al., 2017). Another interesting explanation to the paradox is the recently-suggested idea of hyper-tumours. Hyper-tumors are described as benign tumors that grow on top of a malignant ones and destroy it. It is thought that such hyper-tumors will be favoured by natural selection among the cancer cells population and as tumor in larger organisms usually take more time to develop, this gives the chance to hyper-tumor to be formed only in larger animals (Nagy et al., 2007). However, this hypothesis still has to be investigated more comprehensively including more morphologically-complex organisms. While many studies provide interesting possible explanations to Peto's paradox, they remain insufficient as the paradox persists. Most of the studies mentioned above have focused on trying to explain how can larger animals protect themselves. However, another question that has to be asked is why do they do so.

Interestingly, a very recent study has suggested a new solution to the paradox. It was suggested that only larger organisms would be able to afford such tumor suppressor mechanisms. In most populations, higher fitness is usually linked to large body sizes allowing the organism better chances of mating, accessing resources, and avoiding predators. However, there will usually be

trade-offs between this and that. On the other hand, it might not make much sense for smaller, short-lived animals to invest in cancer defending mechanisms where they will probably die because of other extrinsic reasons (predators for instance).

In fact, this links up very interestingly with some of the findings of this work. As mentioned previously, one of the findings of this work was the lack of conservation of the regulatory interactions between microRNAs and signalling networks (despite them both being highly-conserved individually). Out of the signalling networks studied in this work, ErbB, Notch, and Toll-like are all among the major cancer-regulating signalling networks which resulted in some of the individual genes of these networks (mainly receptors) being targets to many of the recently-developed cancer therapies. Nonetheless, such therapies have never met their therapeutic expectations. In the case of ErbBs especially, there was a clear direction of microRNAs target sites turnover in favour of the primates linage (the more morphologically-complex group in this analysis) when compared to the mouse and rat group. As a result, it might be the case that -under selective pressure- some regulatory features had to be conserved or lost according to the life history of species. However, in case of tumor-suppressor interactions, the lack of conservation might mean that morphologically-complex species might exclusively conserve tumor suppressor regulatory interactions and not just genes as in the case of elephants' *TP53*.

In addition to the possible impact of the gains and losses of microRNA target sites in understanding Peto's paradox, these results also imply that regulatory factors including, but not

limited to, microRNAs should be taken into consideration when choosing an animal model. In fact, the use of animals to substitute humans in laboratories stems back to almost 2500 years ago (Ericsson et al., 2013). Since then, the use of animal models has been a fundamental part of shaping our understanding in many fields like basic biology, genetics, and cancer.

While the choice of animal models classically relied on factors such as the costs, endemicity and morphological characteristics. The rise of the genomics era made it necessarily to include more genetic factors in the process of choosing an animal model. For instance, the analysis of microRNA regulation of some of the genes in this work unveiled some species or lineage specific interactions, making it clear that regulatory interactions needs to be taken into consideration in the process of an animal model choice. In a previous study, Webster et al (2009) combined a series of computational and biological techniques to examine the regulation of *ErbB1* by one of the most studied tumor-suppressor microRNAs, hsa-miR-7. Their results show that hsa-miR-7 effectively down regulate *ErbB1* expression and thus, induce cell death in cancer cells. Interestingly, in a previous study (Helmy and Marco, unpublished), we found that the interaction between the same microRNA (hsa-miR-7) and *ErbB1* is not conserved among any of the different animal models studied. Furthermore, as mentioned previously, several genes and pathways are known to crosstalk with ErbB signalling. Among such genes is *IGF1R* which is known to activate the same signalling pathways activated by ErbB1 (Riedemann et al., 2007). Such crosstalk has later been related to the known issue of resistance to ErbB targeted therapies as, through *IGF1R*, *ErbB1* inhibition might be by-passed (Oliveira et al., 2009). Interestingly, in the study mentioned above, we found no regulation of *IGF1R* by the mentioned

tumor-suppressor microRNA (hsa-miR-7-5p). Such results imply that a downregulation of a given gene by a microRNA might be restricted to one species and might as well be surpassed by crosstalking. As a result, this (and other factors) has to always be taken into consideration while studying microRNA regulation.

In the year 2015, one study referred to the discovery of microRNAs as "the usher of a new era of molecular biology" (Hammond 2015). Just two years later, the roles of microRNAs has been referred to as "an overestimation" (Pinzón et al., 2017). Such disagreement is more than just different opinions. In fact, the role of microRNAs as important regulators of gene expression is beyond any doubt. However, especially in the case of animal microRNAs, it is vital to differentiate between targeting and regulating. Usually, an animal microRNA will have many targets among any genome (i.e. complementarity between its seed region and 3'UTRs) and from there stems the huge expectations of microRNAs and the assumption that they regulate the majority of human genes. However, as mentioned earlier, such complementarity does not mean that regulation will necessarily take place. In addition, as shown in the results of this work, many of the pre-computed microRNA target interactions are based on the use of NMD transcripts which have resulted in millions of inaccurate predictions. Furthermore, experimentally-validated microRNA target interactions have long been considered the gold-standard of microRNA targeting predictions. Unfortunately, many such interactions might not be as trustworthy as once believed.

The validation of a microRNA-target interaction is either carried out using traditional laboratory techniques (like luciferase reporter assay and western blotting) or by the means of the more efficient high-throughput methods (like CLIP) (Thomson et al., 2011). According to the results of this work, the results of the experimental validation of some specific microRNAs may have been significantly inflated by the low-complex nature of their seed regions which questions the reliability of such results. In the case of traditional experimental validation methods, despite being time and cost consuming, they do assume that the microRNA and its targets are in physical contact. An assumption that will probably not take take place unless the microRNA is expressed, and in high quantities, in the same tissue as its target. As a result, I believe that the expression levels of microRNAs and their target sites must be taken into consideration before suggesting that regulation would take place. Based on this, throughout this work highly-expressed microRNAs have been identified and used as more robust regulators of gene expression.

In addition to all of the above, the fact that microRNA target predictions are carried out using reference genomes is an oversight of the different genomic variants and mutations. According to the results of this work, a relatively-large number of genetic variations exists in either the target sites of the ErbB genes or the seed regions of the microRNAs targeting them. As mentioned earlier, the very delicate nature of animal microRNA target interactions makes a single indel or SNP in the seed region of a microRNA (or its target site) a turning point which may result in the destruction of an established microRNA-target interaction or the arising of a new one.

While the incorporation of machine learning models has been useful in the identification of microRNA sequences (Saçar and Allmer 2014), the use of machine learning techniques in microRNA target prediction was not as successful (Amirkhah et al., 2017). Only very recently, the not very successful use of machine learning in microRNA target prediction has resulted in a paradigm shift to the use of deep learning methodologies that, unlike machine learning, does not rely on already-known features of microRNA target prediction but rather rely purely on data (Pla et al., 2018; Wen et al., 2018). However, while the preliminary results of these deep learning based techniques show promise, it is still too early to estimate the extent of their effects on the field of microRNA target prediction.

To conclude, trying to win the endless battle against false positive microRNA target prediction by adding several layers of precision filters will inevitably result in false negative predictions. As a result, I believe that microRNA target prediction should take place in a more case-oriented manner (i.e. specific disease, tissues, species, or population depending) rather than the current global way of predicting microRNA targets regardless of the context that result in this huge gap in the appreciation of microRNAs powers as regulators of gene expression.

.

7. Conclusions

The work of this thesis explores the evolution of the regulatory relationships between two of the most studied topics in today's biology; cell signalling and microRNA. Looking at cell signalling in the context of networks has uniquely provided few insights. First, in ErbB signalling, receptors fall under a strict microRNA-mediated regulation when compared to other signalling levels. In fact, this observation stems from two results, first, the 3'UTRs of ErbB receptors are significantly longer when compared to the 3'UTRs of all other levels which increases the chance of accommodating more microRNA target sites. Furthermore, and regardless of the length, microRNAs that have multiple target sites on the same 3'UTR which are believed to be more robust regulators, show higher density of targeting in the receptors level. It was also evident that the density of target sites for microRNAs that have multiple target sites increases as the length of the 3'UTR increases among the ErbB network. In addition, given the established conservation of both ErbBs receptors and ligands on one hand and microRNAs on the other hand it might be assumed that their interactions are as conserved. Surprisingly, there appears to be a relatively high rate of turnover in microRNA target sites between the species studied. In addition, a surprisingly high number of mutations (SNPs and indels) has been found in either the microRNA target sites on ErbBs or in the seed regions of such microRNAs. Together, these results provide a novel perception of the way microRNA regulated cell signalling. Such perception might be considered in developing our understanding of the functioning of the ErbB network and its effects on diseases, therapies, and animal models.

MicroRNA-mediated regulation does not follow the same pattern in neither Toll-like nor Notch networks. In fact, the density of microRNA targets seems constant among different levels of signalling. However, the relatively high number of microRNA target sites turnover among both Notch and Toll-like networks confirms that the interactions between microRNAs and their targets might under selective pressures. While the lengths of the 3'UTRs of all human receptors are not significantly longer when compared to non-receptors, it was rather surprising that out of all the genes studied in this work, the 3'UTRs of specific receptors (namely *ErbB1*, *ErbB4*, *NOTCH2*, and *TLR4*) that are known for their fundamental roles in regulating diseases were clearly longer which can lead to more strict levels of microRNA regulation.

Target prediction is one of the major aspects in microRNA research. In addition, MicroRNA target prediction was fundamental for the completion of the work in this thesis. For this reason, chapter five was devoted to trying to improve our understanding of microRNA target prediction and for tackling the known issue of low precision among different microRNA target prediction methods. In fact, a large number of pre-computed microRNA target interactions are based on the use of non-functional mRNA transcripts (NMD ones). Accordingly, it is advised that attention should be given to the transcript used in microRNA target prediction. Likewise, the presence of microRNAs with low-complex seed regions might be inflating the validations of microRNA target interaction done using high-throughput methods. Finally, a consideration of the number of microRNA target sites together with a consideration of the distances between such targets might increase the precision of microRNA target interactions.

To conclude, the results of this work provide a novel insight into the evolution of microRNA-mediated regulation among different cell signalling networks as well as tackling some of the issues that is affecting the progression of the field of microRNA research.

8. Future work

Throughout this work, the focus was on canonical microRNA target sites. However, I do acknowledge the existence and importance of other types (non-canonical and near target sites). Likewise, only 3'UTRs were scanned for microRNA target sites despite the few reported regulatory interactions that can take place through binding at the 5'UTR or coding regions of their target mRNAs. The evolutionary analysis of microRNA target interactions in this work has been limited by the availability of reliable genomic data for some species. As the genomes of more and more species gets fully sequenced, more genomic data should be available that would make it possible to provide a wider view on the conservation of microRNA regulatory interactions generally and between the primates and rodents groups specifically. In addition, more data should be available on the expression levels of both microRNAs and their targets. Consequently, the advances in the field of microRNA identification promises the identification of microRNAs from more species which might also be added to this analysis.

As mentioned earlier, networks represents the most complex part of cell signalling. In this work, three of the major cell signalling networks have been studied. An interesting addition to this work would be to explore microRNA regulation to other networks as well as less-complex pathways. Also, due to the life-history of this work, the analysis of Notch and Toll-like networks were influenced by the findings of the ErbB network. It would also be interesting to blindly study these and other networks.

Finally, the nature of the work in this thesis has been pure computational. However, it would also be interesting to validate some of the results of this work (especially of chapter 5) using experimental procedures.

9. References

Aalaei-andabili, S. H., & Rezaei, N. (2013). Toll like receptor (TLR)-induced differential expression of microRNAs (MiRs) and immune response against infection: A systematic review.*Journal of Infection*, *67*(4), 251-264.

Abegglen, L. M., Caulin, A. F., Chan, A., Lee, K., Robinson, R., Campbell, M. S., et al. (2015). Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *Jama*, *314*(17), 1850-1860.

Adem, B. F., Bastos, N. R. A., Dias, F., Teixeira, A. L., & Medeiros, R. (2016). miRNAs: mediators of ErbB family targeted therapy resistance. *Pharmacogenomics*, *17*(10), 1175-1187.

Agarwal, V., Bell, G. W., Nam, J. W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *elife*, *4*.

Agrawal, N., Dasaradhi, P. V. N., Mohmmed, A., Malhotra, P., Bhatnagar, R. K., & Mukherjee, S. K. (2003). RNA interference: biology, mechanism, and applications. *Microbiology and molecular biology reviews*, 67(4), 657-685.

Akira, S. (2006). TLR signaling. *From Innate Immunity to Immunological Memory*. 1-16.

Akira, S., & Takeda, K. (2004). Toll-like receptor signalling. *Nature reviews immunology*, *4*(7), 499-511.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J. D. (1983). *Molecular biology of the cell* (1st ed.). New York: Garland Science.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J. D. (1989). *Molecular biology of the cell* (2nd ed.). New York: Garland Science.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J. D. (1994). *Molecular biology of the cell* (3rd ed.). New York: Garland Science.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular biology of the cell* (4th ed.). New York: Garland Science.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). *Molecular biology of the cell* (5th ed.). New York: Garland Science.

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., et al. (2014). *Molecular biology of the cell* (6th ed.). New York: Garland Science.

Alexiou, P., Maragkakis, M., Papadopoulos, G. L., Reczko, M., & Hatzigeorgiou, A. G. (2009). Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics*, *25*(23), 3049-3055.

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., & Matzke, M. (2003). A uniform system for microRNA annotation. *RNA*, 9(3), 277-279.

Ameres, S. L., & Zamore, P. D. (2013). Diversifying microRNA sequence and function. *Nature reviews Molecular cell biology,* 14(8), 475-488.

Amirkhah, R., Meshkin, H. N., Farazmand, A., Rasko, J. E., & Schmitz, U. (2017). Computational and Experimental Identification of Tissue-Specific MicroRNA Targets in *MicroRNA Detection and Target Identification*. 127-147.

Anklesaria, P., Teixido, J., Laiho, M., Pierce, J. H., Greenberger, J. S., & Massague, J. (1990). Cell-cell adhesion mediated by binding of membrane-anchored transforming growth factor alpha to epidermal growth factor receptors promotes cell proliferation. *Proceedings of the National Academy of Sciences*, *87*(9), 3289-3293.

Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., et al. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular cell,* 31(6), 785-799.

Arteaga, C. L., & Engelman, J. A. (2014). ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer cell*, *25*(3), 282-303.

Ausubel, F. M. (2005). Are innate immune signaling pathways in plants and animals conserved?. *Nature immunology*, *6*(10), 973.

Axtell, M. J., Westholm, J. O., & Lai, E. C. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome biology*, *12*(4), 221.

Barker, A., Giles, K. M., Epis, M. R., Zhang, P. M., Kalinowski, F., & Leedman, P. J. (2010). Regulation of ErbB receptor signalling in cancer cells by microRNA. *Current opinion in pharmacology*, *10*(6), 655-661.

Bartel, D. P., & Chen, C. Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nature Reviews Genetics*, *5*(5), 396.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell,* 136(2), 215-233.

Bazzoni, F., Rossato, M., Fabbri, M., Gaudiosi, D., Mirolo, M., Mori, L., et al. (2009). Induction and regulatory function of miR-9 in human monocytes and neutrophils exposed to proinflammatory signals. *Proceedings of the National Academy of Sciences*, *106*(13), 5282-5287.

Ben-Ali, M., Barbouche, M. R., Bousnina, S., Chabbou, A., & Dellagi, K. (2004). Toll-like receptor 2 Arg677Trp polymorphism is associated with susceptibility to tuberculosis in Tunisian patients. *Clinical and diagnostic laboratory immunology*, *11*(3), 625-626.

Beniaminov, A., Westhof, E., & Krol, A. (2008). Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA,* 14(7), 1270-1275.

Bentwich, I. (2005). Prediction and validation of microRNAs and their targets. *FEBS letters,* 579(26), 5904-5910.

Berezikov, E., Thuemmler, F., van Laake, L. W., Kondova, I., Bontrop, R., Cuppen, E., et al. (2006). Diversity of microRNAs in human and chimpanzee brain. *Nature genetics,* 38(12), 1375-1377.

Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. *Nature Reviews Genetics,* 12(12), 846-860.

Beveridge, N. J., Gardiner, E., Carroll, A. P., Tooney, P. A., & Cairns, M. J. (2010). Schizophrenia is associated with an increase in cortical microRNA biogenesis. *Molecular psychiatry,* 15(12), 1176-1189.

Bhattacharya, A., Ziebarth, J. D., & Cui, Y. (2013). PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic acids research*, *42*(D1), D86-D91.

Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I., & Filipowicz, W. (2006). Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell*, *125*(6), 1111-1124.

Bigas, A. and Espinosa, L. (2012). Hematopoietic stem cells: to be or Notch to be. *Blood*, 119(14), 3226-3235.

Blokzijl, A., Dahlqvist, C., Reissmann, E., Falk, A., Moliner, A., Lendahl, U., et al. (2003). Cross-talk between the Notch and TGF-β signaling pathways mediated by interaction of the Notch intracellular domain with Smad3. *The Journal of cell biology*, *163*(4), 723-728.

Bossaller, L., Christ, A., Pelka, K., Nündel, K., Chiang, P. I., Pang, C., et al. (2016). TLR9 deficiency leads to accelerated renal disease and myeloid lineage abnormalities in pristane-induced murine lupus. *The Journal of Immunology*, 1044-1053.

Bostjancic, E., & Glavac, D. (2008). Importance of microRNAs in skin morphogenesis and diseases. *Acta dermatovenerologica Alpina, Pannonica, et Adriatica,* 17(3), 95-102.

Bradley, T., & Moxon, S. (2017). An Assessment of the Next Generation of Animal miRNA Target Prediction Algorithms. *MicroRNA Detection and Target Identification: Methods and Protocols*, 175-191.

Bray, S. J. (2006). Notch signalling: a simple pathway becomes complex. *Nature reviews Molecular cell biology*, *7*(9), 678-689.

Bray, S. J. (2016). Notch signalling in context. *Nature reviews Molecular cell biology*, *17*(11), 722-735.

Bray, D., & Lay, S. (1994). Computer simulated evolution of a network of cell-signaling molecules. *Biophysical journal*, *66*(4), 972-977.

Brookes, A. J. (1999). The essence of SNPs. *Gene*, *234*(2), 177-186.

Brunet, T., & King, N. (2017). The origin of animal multicellularity and cell differentiation. *Developmental cell*, *43*(2), 124-140.

Burgess, A. W., Cho, H. S., Eigenbrot, C., Ferguson, K. M., Garrett, T. P., Leahy, D. J., et al. (2003). An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Molecular cell*, *12*(3), 541-552.

Calin, G. A., Dumitru, C. D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., et al. (2002). Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences,* 99(24), 15524-15529.

Campo‐Paysaa, F., Sémon, M., Cameron, R. A., Peterson, K. J., & Schubert, M. (2011). microRNA complements in deuterostomes: origin and evolution of microRNAs. *Evolution & development,* 13(1), 15-27.

Caulin, A. F., & Maley, C. C. (2011). Peto's Paradox: evolution's prescription for cancer prevention. *Trends in ecology & evolution*, *26*(4), 175-182.

Chakraborty, C., Sharma, A. R., Patra, B. C., Bhattacharya, M., Sharma, G., & Lee, S. S. (2016). MicroRNAs mediated regulation of MAPK signaling pathways in chronic myeloid leukemia. *Oncotarget*, *7*(27), 42683.

Chan, L., F Wang, F., & CS Cho, W. (2012). Genomic sequence analysis of EGFR regulation by microRNAs in lung cancer. *Current topics in medicinal chemistry*, *12*(8), 920-926.

Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J., & Joo, C. (2015). A dynamic search process underlies microRNA targeting. *Cell*, *162*(1), 96-107.

Chen, J. (2015). MicroRNAs, signaling pathways and diseases. *Annals of translational medicine*, *3*(21), 329.

Chen, J. F., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T. E., Hammond, et al. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature genetics*, *38*(2), 228-233.

Cheng, C., Bhardwaj, N., & Gerstein, M. (2009). The relationship between the evolution of microRNA targets and the length of their UTRs. *BMC genomics*, *10*(1), 431-437.

Chitnis, A., Henrique, D., Lewis, J., Ish-Horowicz, D., & Kintner, C. (1995). Primary neurogenesis in Xenopus embryos regulated by a homologue of the Drosophila neurogenic gene Delta. *Nature*, *375*(6534), 761-766.

Chou, C. H., Shrestha, S., Yang, C. D., Chang, N. W., Lin, Y. L., Liao, K. W., et al. (2017). miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic acids research*, *46*(D1), 296-302.

Citri, A., & Yarden, Y. (2006). EGF–ERBB signalling: towards the systems level. *Nature reviews Molecular cell biology*, *7*(7), 505-516.

Cohen, S. (1962). Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. *Journal of Biological Chemistry*, *237*(5), 1555-1562.

Collu, G. M., Hidalgo-Sastre, A., & Brennan, K. (2014). Wnt–Notch signalling crosstalk in development and disease.*Cellular and molecular life sciences*, *71*(18), 3553-3567.

Cooper, G. M., & Hausman, R. E. (2004). *The cell: Molecular approach*. Medicinska naklada.

Cooper, G. M. (2000). *The cell: A molecular approach*. Washington: ASM Press.

Croce, C. M., & Calin, G. A. (2005). miRNAs, cancer, and stem cell division. *Cell,* 122(1), 6-7.

Dawkins, R. (1976). The selfish gene Oxford university press. New York.

De Angelis, M. H., Mclntyre II, J., & Gossler, A. (1997). Maintenance of somite borders in mice requires the Delta homologue Dll1. *Nature*, *386*(6626), 717-721.

De la Pompa, J. L., & Epstein, J. A. (2012). Coordinating tissue interactions: Notch signaling in cardiac development and disease. *Developmental cell*, *22*(2), 244-254.

Didiano, D., & Hobert, O. (2006). Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nature Structural and Molecular Biology*, *13*(9), 849-851.

Dunwoodie, S. L. (2009). The role of Notch in patterning the human vertebral column. *Current opinion in genetics & development*, *19*(4), 329-337.

Eichhorn, S. W., Guo, H., McGeary, S. E., Rodriguez-Mias, R. A., Shin, C., Baek, D., et al. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Molecular cell*, *56*(1), 104-115.

Eisenberg, I., Eran, A., Nishino, I., Moggio, M., Lamperti, C., Amato, A. A., et al. (2007). Distinctive patterns of microRNA expression in primary muscular disorders. *Proceedings of the National Academy of Sciences,* 104(43), 17016-17021.

Epis, M. R., Giles, K. M., Barker, A., Kendrick, T. S., & Leedman, P. J. (2009). miR-331-3p regulates ERBB-2 expression and androgen receptor signaling in prostate cancer. *Journal of Biological Chemistry*, 284(37), 24696-24704

Esteller, M. (2011). Non-coding RNAs in human disease. *Nature Reviews Genetics,* 12(12), 861-874.

Eungdamrong, N. J., & Iyengar, R. (2004). Modeling cell signaling networks. *Biology of the Cell*, *96*(5), 355-362.

Fabris, L., Berton, S., Citron, F., D'Andrea, S., Segatto, I., Nicoloso, M. S., et al. (2016). Radiotherapy-induced miR-223 prevents relapse of breast cancer by targeting the EGF pathway. *Oncogene*, *35*(37), 4914-4926.

Falschlehner, C., Emmerich, C. H., Gerlach, B., & Walczak, H. (2007). TRAIL signalling: decisions between life and death.*The international journal of biochemistry & cell biology*, *39*(7-8), 1462-1475.

Fichtlscherer, S., Zeiher, A. M., & Dimmeler, S. (2011). Circulating microRNAs: biomarkers or mediators of cardiovascular diseases?. *Arteriosclerosis, thrombosis, and vascular biology*, *31*(11), 2383-2390.

Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?. *Nature reviews genetics*, *9*(2), 102.

Freeman, M., & Gurdon, J. B. (2002). Regulatory principles of developmental signaling. *Annual review of cell and developmental biology*, *18*(1), 515-539.

Friedman, R. C., Farh, K. K. H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome research*, *19*(1), 92-105.

Gazave, E., Lapébie, P., Richards, G. S., Brunet, F., Ereskovsky, A. V., Degnan, B. M., et al. (2009). Origin and evolution of the Notch signalling pathway: an overview from eukaryotic genomes. *BMC evolutionary biology*, *9*(1), 249-286.

Ghoshal, A., Shankar, R., Bagchi, S., Grama, A., & Chaterji, S. (2015). MicroRNA target prediction using thermodynamic and sequence curves. *BMC genomics*, *16*(1), 999.

Giles, K. M., Barker, A., Zhang, P. M., Epis, M. R., & Leedman, P. J. (2011). MicroRNA regulation of growth factor receptor signaling in human cancer cells. In *MicroRNA and Cancer,* 676, 147-163.

Giraldez, A. J., Cinalli, R. M., Glasner, M. E., Enright, A. J., Thomson, J. M., Baskerville, S., Hammond, S. M., Bartel, D. P. & Schier, A. F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. *Science*, 308(5723), 833-838.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N., et al. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*, 455(7217), 1193-1197.

Gu, S., & Kay, M. A. (2010) How do miRNAs mediate translational repression?. *Silence* 1(1), 11.

Guy, P. M., Platko, J. V., Cantley, L. C., Cerione, R. A., & Carraway, K. L. (1994). Insect cell-expressed p180erbB3 possesses an impaired tyrosine kinase activity. *Proceedings of the National Academy of Sciences*, *91*(17), 8132-8136.

Hammond, S. M. (2015). An overview of microRNAs. *Advanced drug delivery reviews*, *87*, 3-14.

Hancock, J. (2010). *Cell signalling*. Oxford: Oxford Univ. Press.

Hancock, J. T. (2008). Cell signalling is the music of life. *British journal of biomedical science*, *65*(4), 205-208.

Hatfield, S. D., Shcherbata, H. R., Fischer, K. A., Nakahara, K., Carthew, R. W., & RuoholaBaker, H. (2005). Stem cell division is regulated by the microRNA pathway. *Nature*, 435(7044), 974-978.

He, X., Jing, Z., & Cheng, G. (2014). MicroRNAs: new regulators of Toll-like receptor signalling pathways. *BioMed research international*, *2014*.

Hollstein, M., Sidransky, D., Vogelstein, B., & Harris, C. C. (1991). p53 mutations in human cancers. *Science*, *253*(5015), 49-53.

Huang, S., Yuan, S., Guo, L., Yu, Y., Li, J., Wu, T., et al. (2008). Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome research*.

Ichimura, A., Ruike, Y., Terasawa, K., & Tsujimoto, G. (2011). miRNAs and regulation of cell signaling. *The FEBS journal*, *278*(10), 1610-1618.

Ivey, K. N., & Srivastava, D. (2010). MicroRNAs as regulators of differentiation and cell fate decisions. *Cell stem cell*, *7*(1), 36-41.

Jansson, M. D., & Lund, A. H. (2012). MicroRNA and cancer. Molecular oncology, 6(6), 590-610.

Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., et al. (2005). RAS is regulated by the let-7 microRNA family. *Cell*, *120*(5), 635-647.

Jordan, J. D., Landau, E. M., & Iyengar, R. (2000). Signaling networks: the origins of cellular multitasking. *Cell*, *103*(2), 193-200.

Josse, C., Bouznad, N., Geurts, P., Irrthum, A., Servais, L., Hego, A., et al. (2014). Identification of a microRNA landscape targeting the PI3K/Akt signaling pathway in inflammation-induced colorectal carcinogenesis. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *306*(3), 229-243.

Joutel, A., Corpechot, C., Ducros, A., Vahedi, K., Chabriat, H., Mouton, P., et al. (1996). Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature*, *383*(6602), 707-710

Kang, T. J., Lee, S. B., & Chae, G. T. (2002). A polymorphism in the toll-like receptor 2 is associated with IL-12 production from monocyte in lepromatous leprosy. *Cytokine*, *20*(2), 56-62.

Kedmi, M., Sas-Chen, A., & Yarden, Y. (2015). MicroRNAs and growth factors: an alliance propelling tumor progression. *Journal of clinical medicine*, *4*(8), 1578-1599.

Kiechl, S., Lorenz, E., Reindl, M., Wiedermann, C. J., Oberhollenzer, F., Bonora, E., et al. (2002). Toll-like receptor 4 polymorphisms and atherogenesis. *New England Journal of Medicine*, *347*(3), 185-192.

Kim, S., & Misra, A. (2007). SNP genotyping: technologies and biomedical applications. *Annu. Rev. Biomed. Eng.*, *9*, 289-320.

Kim, V. N., & Nam, J. W. (2006). Genomics of microRNA. *Trends in Genetics,* 22(3), 165-173.

Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature*, *267*(5608), 275.

Kohanbash, G., & Okada, H. (2012). MicroRNAs and STAT interplay. *Seminars In Cancer Biology*, *22*(1), 70-75.

Kozomara, A., & Griffiths-Jones, S. (2013). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, *42*(D1), 68-73.

Kuhn, D. E., Martin, M. M., Feldman, D. S., Terry Jr, A. V., Nuovo, G. J., & Elton, T. S. (2008). Experimental validation of miRNA targets. *Methods*, *44*(1), 47-54.

Kuhn, D. E., Nuovo, G. J., Terry, A. V., Martin, M. M., Malana, G. E., Sansom, S. E., et al. (2010). Chromosome 21- derived microRNAs provide an etiological basis for aberrant protein expression in human Down syndrome brains. *Journal of Biological Chemistry,* 285(2), 1529-1543.

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, *34*(7), 1812-1819.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., & Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. *Current Biology,* 12(9), 735-739.

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell,* 75(5), 843-854.

Lee, S., & Vasudevan, S. (2012). Post-transcriptional Stimulation of Gene Expression by MicroRNAs. *Advances In Experimental Medicine And Biology*, 97-126.

Li, L., Krantz, I. D., Deng, Y., Genin, A., Banta, A. B., Collins, C. C., et al. (1997). Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nature genetics*, *16*(3), 243-251.

Li, S. D., Chono, S., & Huang, L. (2008). Efficient oncogene silencing and metastasis inhibition via systemic delivery of siRNA. *Molecular Therapy,* 16(5), 942-946.

Li, Y., & Kowdley, K. V. (2012). MicroRNAs in common human diseases. *Genomics, proteomics & bioinformatics,* 10(5), 246-253.

Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., & Cui, Q. (2013). HMDD v2. 0: a database for experimentally supported human microRNA and disease associations. *Nucleic acids research*, *42*(D1), 1070-1074.

Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B. & Bartel, D. P. (2003). The microRNAs of Caenorhabditis elegans. *Genes & development,* 17(8), 991-1008.

Lindeboom, R. G., Supek, F., & Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nature genetics*, *48*(10), 1112.

Lindsey, M. & Griffiths-Jones, S. (2013). *The role of non-coding RNAs in biology*. London: Portland Press Limited.

Ling, B., Wang, G. X., Long, G., Qiu, J. H., & Hu, Z. L. (2012). Tumor suppressor miR-22 suppresses lung cancer cell progression through post-transcriptional regulation of ErbB3. *Journal of cancer research and clinical oncology*, *138*(8), 1355-1361.

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., et al. (2005). MicroRNA expression profiles classify human cancers. *Nature*, 435(7043), 834-838.

Maragkakis, M., Alexiou, P., Papadopoulos, G. L., Reczko, M., Dalamagas, T., Giannopoulos, et al. (2009). Accurate microRNA target prediction correlates with protein repression levels. *BMC bioinformatics*, *10*(1), 295.

Marco, A., MacPherson, J. I., Ronshaugen, M., & Griffiths-Jones, S. (2012). MicroRNAs from the same precursor have different targeting properties. Silence, 3(1), 8-8.

Marco, A. (2018). SeedVicious: Analysis of microRNA target and near-target sites. *PloS one*, *13*(4), e0195532.

Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Human molecular genetics*, *15*(suppl_1), 17-29.

Maziere, P., & Enright, A. J. (2007). Prediction of microRNA targets. *Drug discovery today,* 12(11), 452-458.

McDaniell, R., Warthen, D. M., Sanchez-Lara, P. A., Pai, A., Krantz, I. D., Piccoli, D. A., & Spinner, N. B. (2006). NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *The American Journal of Human Genetics*, *79*(1), 169-173.

Meader, S., Ponting, C. P., & Lunter, G. (2010). Massive turnover of functional sequence in human and other mammalian genomes. *Genome research,* 20(10), 1335-1343.

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature protocols*, *8*(8), 1551-1566.

Millar, A. A., & Waterhouse, P. M. (2005). Plant and animal microRNAs: similarities and differences. *Functional & integrative genomics*, *5*(3), 129-135.

Mills, R. E., Bennett, E. A., Iskow, R. C., & Devine, S. E. (2007). Which transposable elements are active in the human genome?. *Trends in Genetics,* 23(4), 183-191.

Miranda, K. C., Huynh, T., Tay, Y., Ang, Y. S., Tam, W. L., Thomson, A. M., et al. (2006). A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, *126*(6), 1203-1217.

Mockenhaupt, F. P., Hamann, L., von Gaertner, C., Bedu-Addo, G., von Kleinsorgen, C., Schumann, R. R., et al. (2006). Common polymorphisms of toll-like receptors 4 and 9 are associated with the clinical manifestation of malaria during pregnancy. *The Journal of infectious diseases*, *194*(2), 184-188.

Molnár, V., Ersek, B., Wiener, Z., Tömböl, Z., Szabó, P. M., Igaz, P., et al. (2012). MicroRNA-132 targets HB-EGF upon IgE-mediated activation in murine and human mast cells. *Cellular and Molecular Life Sciences*, *69*(5), 793-808.

Morimoto, A., Kannari, M., Tsuchida, Y., Sasaki, S., Saito, C., Matsuta, T., et al. (2017). An HNF4α-microRNA-194/192 Signaling Axis Maintains Hepatic Cell Function. *Journal of Biological Chemistry*, 292(25), 10574-10585.

Moschos, S. A., Williams, A. E., Perry, M. M., Birrell, M. A., Belvisi, M. G., & Lindsay, M. A. (2007). Expression profiling in vivo demonstrates rapid changes in lung microRNA levels following lipopolysaccharide-induced inflammation but not in the anti-inflammatory action of glucocorticoids. *BMC genomics*, *8*(1), 240.

Moszyńska, A., Gebert, M., Collawn, J. F., & Bartoszewski, R. (2017). SNPs in microRNA target sites and their potential role in human disease. *Open biology*, *7*(4), 170019.

Murchison, E. P., & Hannon, G. J. (2004). miRNAs on the move: miRNA biogenesis and the RNAi machinery. *Current opinion in cell biology,* 16(3), 223-229.

Nagy, J. D., Victor, E. M., & Cropper, J. H. (2007). Why don't all whales have cancer? A novel hypothesis resolving Peto's paradox. *Integrative and comparative biology*, *47*(2), 317-328.

Nakasa, T., Miyaki, S., Okubo, A., Hashimoto, M., Nishida, K., Ochi, M., & Asahara, H. (2008). Expression of microRNA‐146 in rheumatoid arthritis synovial tissue. *Arthritis & Rheumatism,* 58(5), 1284-1292.

Nam, S., Kim, B., Shin, S., & Lee, S. (2008). miRGator: an integrated system for functional annotation of microRNAs. *Nucleic acids research*, 36(suppl 1), 159-164.

Nelson, P. T., Wang, W. X., & Rajeev, B. W. (2008). MicroRNAs (miRNAs) in neurodegenerative diseases. *Brain Pathology,* 18(1), 130-138.

O'neill, L. A., Golenbock, D., & Bowie, A. G. (2013). The history of Toll-like receptors—redefining innate immunity. *Nature Reviews Immunology,* 13(6), 453.

O'neill, L. A., Sheedy, F. J., & McCoy, C. E. (2011). MicroRNAs: the fine-tuners of Toll-like receptor signalling. *Nature Reviews Immunology*, *11*(3), 163.

Oliveira, A. C., Bovolenta, L. A., Nachtigall, P. G., Herkenhoff, M. E., Lemke, N., & Pinhal, D. (2017). Combining results from distinct MicroRNA target prediction tools enhances the performance of analyses. *Frontiers in genetics*, *8*, 59-69.

Oliveira, S., Schiffelers, R. M., Storm, G., Henegouwen, P. M. P., & Roovers, R. C. (2009). Crosstalk between epidermal growth factor receptor-and insulin-like growth factor-1 receptor signaling: implications for cancer therapy. *Current cancer drug targets*, *9*(6), 748-760.

Paraskevopoulou, M. D., Georgakilas, G., Kostoulas, N., Vlachos, I. S., Vergoulis, T., Reczko, M., et al. (2013). DIANA-microT web server v5. 0: service integration into miRNA functional analysis workflows. *Nucleic acids research*, 41(1), 169-173.

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808), 86-89.

Peterson, S. M., Thompson, J. A., Ufkin, M. L., Sathyanarayana, P., Liaw, L., & Congdon, C. B. (2014). Common features of microRNA target prediction tools. *Frontiers in genetics*, *5*, 23-24.

Peto, R., Roe, F. J., Lee, P. N., Levy, L., & Clack, J. (1975). Cancer and ageing in mice and men. *British Journal of Cancer*, *32*(4), 411-426.

Pinzón, N., Li, B., Martinez, L., Sergeeva, A., Presumey, J., Apparailly, F., & Seitz, H. (2017). microRNA target prediction programs predict many false positives. *Genome research*, *27*(2), 234-245.

Powar, C. B. (2010), Cell Signalling, Himalaya Publishing House, Mumbai.

Prickett, T. D., Agrawal, N. S., Wei, X., Yates, K. E., Lin, J. C., Wunderlich, J. R., et al. (2009). Analysis of the tyrosine kinome in melanoma reveals recurrent mutations in ERBB4. *Nature genetics*, 41(10), 1127.

Qian, F. H., Zhang, Q., Zhou, L. F., Jin, G. F., Bai, J. L., & Yin, K. S. (2010). 9 Polymorphisms in the Toll-like Receptor 2 Subfamily and Risk of Asthma: A Case-control Analysis in a Chinese Population. *Journal of investigational allergology & clinical immunology*, 20(4), 340-346.

Raby, B. A., Klimecki, W. T., Laprise, C., Renaud, Y., Faith, J., Lemire, M., et al. (2002). Polymorphisms in toll-like receptor 4 are not associated with asthma or atopy-related phenotypes. *American journal of respiratory and critical care medicine*, *166*(11), 1449-1456.

Radstake, T. R., Franke, B., Hanssen, S., Netea, M. G., Welsing, P., Barrera, P., et al. (2004). The Toll-like receptor 4 Asp299Gly functional variant is associated with decreased rheumatoid arthritis disease susceptibility but does not influence disease severity and/or outcome. *Arthritis & Rheumatism*, *50*(3), 999-1001.

Ramírez-Weber, F. A., & Kornberg, T. B. (1999). Cytonemes: cellular processes that project to the principal signaling center in Drosophila imaginal discs. *Cell*, *97*(5), 599-607.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., et al. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403(6772), 901-906.

Riedemann, J., Takiguchi, M., Sohail, M., & Macaulay, V. M. (2007). The EGF receptor interacts with the type 1 IGF receptor and regulates its stability. *Biochemical and biophysical research communications*, *355*(3), 707-714.

Riffo-Campos, Á. L., Riquelme, I., & Brebi-Mieville, P. (2016). Tools for sequence-based miRNA target prediction: What to choose?. *International journal of molecular sciences*, *17*(12), 1987.

Robey, E., Chang, D., Itano, A., Cado, D., Alexander, H., Lans, D., et al. (1996). An activated form of Notch influences the choice between CD4 and CD8 T cell lineages. *Cell*, *87*(3), 483-492.

Roca, C., & Adams, R. H. (2007). Regulation of vascular morphogenesis by Notch signaling. *Genes & development*, 21(20), 2511-2524.

Samaga, R., Saez-Rodriguez, J., Alexopoulos, L. G., Sorger, P. K., & Klamt, S. (2009). The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS computational biology*, 5(8), e1000438.

Sætrom, P., Heale, B. S., Snøve Jr, O., Aagaard, L., Alluin, J., & Rossi, J. J. (2007). Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic acids research*, *35*(7), 2333-2342.

Schmelzle, T., & Hall, M. N. (2000). TOR, a central controller of cell growth. *Cell*, *103*(2), 253-262.

Shan, X., Wen, W., Zhu, D., Yan, T., Cheng, W., Huang, Z., et al. (2017). miR 1296-5p inhibits the migration and invasion of gastric cancer cells by repressing ERBB2 expression. *PloS one*, *12*(1), e0170298.

Sheedy, F. J., Palsson-McDermott, E., Hennessy, E. J., Martin, C., O'leary, J. J., Ruan, Q., et al. (2010). Negative regulation of TLR4 via targeting of the proinflammatory tumor suppressor PDCD4 by the microRNA miR-21. *Nature immunology*, *11*(2), 141-147.

Simpson, M. A., Irving, M. D., Asilmaz, E., Gray, M. J., Dafou, D., Elmslie, F. V., et al. (2011). Mutations in NOTCH2 cause Hajdu-Cheney syndrome, a disorder of severe and progressive bone loss. *Nature genetics*, *43*(4), 303-308.

Smirnova, L., Gräfe, A., Seiler, A., Schumacher, S., Nitsch, R., & Wulczyn, F. G. (2005). Regulation of miRNA expression during neural cell specification. *European Journal of Neuroscience,* 21(6), 1469-1477.

Sood, P., Krek, A., Zavolan, M., Macino, G., & Rajewsky, N. (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. *Proceedings of the National Academy of Sciences*, *103*(8), 2746-2751.

Soung, Y. H., Lee, J. W., Kim, S. Y., Wang, Y. P., Jo, K. H., Moon, S. W., et al. (2006). Somatic mutations of the ERBB4 kinase domain in human cancers. *International journal of cancer*, 118(6), 1426-1429.

Sporn, M. B., & Roberts, A. B. (1985). Autocrine growth factors and cancer. *Nature*, *313*(6005), 745-774.

Stark, A., Brennecke, J., Bushati, N., Russell, R. B., & Cohen, S. M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′ UTR evolution. *Cell*, 123(6), 1133-1146.

Stark, A., Brennecke, J., Russell, R. B., & Cohen, S. M. (2003). Identification of Drosophila microRNA targets. *PLoS biology*, *1*(3), e60.

Stein, R. A., & Staros, J. V. (2000). Evolutionary analysis of the ErbB receptor and ligand families. *Journal of molecular evolution*, *50*(5), 397-412.

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research*, *34*(suppl_2), 609-612.

Taganov, K. D., Boldin, M. P., Chang, K. J., & Baltimore, D. (2006). NF-κB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proceedings of the National Academy of Sciences*, 103(33), 12481-12486.

Taniguchi, C. M., Emanuelli, B., & Kahn, C. R. (2006). Critical nodes in signalling pathways: insights into insulin action. *Nature reviews Molecular cell biology*, *7*(2), 85-96.

Tatusova, T. A., & Madden, T. L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS microbiology letters*, *174*(2), 247-250.

Team, R. C. (2014). R: A language and environment for statistical computing.

Tebbutt, N., Pedersen, M. W., & Johns, T. G. (2013). Targeting the ERBB family in cancer: couples therapy. *Nature Reviews Cancer*, *13*(9), 663-673.

Thomson, D. W., Bracken, C. P., & Goodall, G. J. (2011). Experimental strategies for microRNA target identification. *Nucleic acids research*, 39(16), 6845-6853.

Thomson, T., & Lin, H. (2009). The biogenesis and function PIWI proteins and piRNAs: progress and prospect. *Annual review of cell and developmental biology,* 25, 355-376.

Tollis, M., Boddy, A. M., & Maley, C. C. (2017). Peto's Paradox: how has evolution solved the problem of cancer prevention?. *BMC biology*, *15*(1), 60-65.

Toomey, S., Madden, S. F., Furney, S. J., Fan, Y., McCormack, M., Stapleton, C., et al. (2016). The impact of ERBB-family germline single nucleotide polymorphisms on survival response to adjuvant trastuzumab treatment in HER2-positive breast cancer. *Oncotarget*, *7*(46), 75518-75525.

Torok, H. P., Glas, J., Tonenchi, L., Mussack, T., & Folwaczny, C. (2004). Polymorphisms of the lipopolysaccharide-signaling complex in inflammatory bowel disease: association of a mutation in the Toll-like receptor 4 gene with ulcerative colitis. *Clinical immunology*, *112*(1), 85-91.

Tung, S. L., Huang, W. C., Hsu, F. C., Yang, Z. P., Jang, T. H., Chang, J. W., et al. (2017). miRNA-34c-5p inhibits amphiregulin-induced ovarian cancer stemness and drug resistance via downregulation of the AREG-EGFR-ERK pathway. *Oncogenesis*, *6*(4), 326-339.

Vasudevan, S. (2012). Posttranscriptional upregulation by microRNAs. *Wiley Interdisciplinary Reviews: RNA*, *3*(3), 311-330.

Vlachos, I. S., Paraskevopoulou, M. D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., et al. (2015). DIANA-TarBase v7. 0: indexing more than half a million experimentally supported miRNA: mRNA interactions. *Nucleic acids research,* 43(1), 153-159.

Voogdt, C. G., & van Putten, J. P. (2016). The Evolution of the Toll-Like Receptor System. In *The Evolution of the Immune System,* 311-330.

Wang, Q., & Greene, M. I. (2008). Mechanisms of resistance to ErbB-targeted cancer therapeutics. *The Journal of clinical investigation*, *118*(7), 2389-2392.

Wang, W. X., Rajeev, B. W., Stromberg, A. J., Ren, N., Tang, G., Huang, Q., et al. (2008). The expression of microRNA miR-107 decreases early in Alzheimer's disease and may accelerate disease progression through regulation of β-site amyloid precursor protein-cleaving enzyme 1. The Journal of Neuroscience, 28(5), 1213-122.

Ward, A. E., & Rosenthal, B. M. (2014). Evolutionary responses of innate immunity to adaptive immunity. *Infection, Genetics and Evolution*, 21, 492-496.

Watanabe, Y., Tomita, M., & Kanai, A. (2007). Computational methods for microRNA target prediction. *Methods in enzymology*, 427, 65-86.

Webster, R. J., Giles, K. M., Price, K. J., Zhang, P. M., Mattick, J. S., & Leedman, P. J. (2009). Regulation of epidermal growth factor receptor signaling in human cancer cells by microRNA-7. *Journal of Biological Chemistry*, *284*(9), 5731-5741.

Wei, J., & Hemmings, G. P. (2000). The NOTCH4 locus is associated with susceptibility to schizophrenia. *Nature genetics*, *25*(4), 376.

Weinberg, R. (2013). The biology of cancer. Garland Science.

Weng, A. P., Ferrando, A. A., Lee, W., Morris, J. P., Silverman, L. B., Sanchez-Irizarry, C., et al. (2004). Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science*, *306*(5694), 269-271.

Weyrich, P., Staiger, H., Stančáková, A., Machicao, F., Machann, J., Schick, F., et al. (2010). The D299G/T399I Toll-like receptor 4 variant associates with body and liver fat: results from the TULIP and METSIM Studies. *PloS one*, *5*(11), e13980.

Wieduwilt, M. J., & Moasser, M. M. (2008). The epidermal growth factor receptor family: biology driving targeted therapeutics. *Cellular and Molecular Life Sciences*, *65*(10), 1566-1584.

Wightman, B., Ha, I., & Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, 75(5), 855-862.

Williams, C. S., Bernard, J. K., Demory Beckler, M., Almohazey, D., Washington, M. K., Smith, J. J., & Frey, M. R. (2015). ERBB4 is over-expressed in human colon cancer and enhances cellular transformation. *Carcinogenesis*, *36*(7), 710-718.

Wong, N., & Wang, X. (2014). miRDB: an online resource for microRNA target prediction and functional annotations.*Nucleic acids research*, *43*(1), 146-152.

Wu, H., Liu, Y., Shu, X. O., & Cai, Q. (2016). MiR-374a suppresses lung adenocarcinoma cell proliferation and invasion by targeting TGFA gene expression. *Carcinogenesis*, *37*(6), 567-575.

Wu, L., & Belasco, J. G. (2008). Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs. *Molecular cell*, *29*(1), 1-7.

Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., & Li, T. (2009). miRecords: an integrated resource for microRNA–target interactions. *Nucleic acids research,* 37(suppl 1), D105-D110.

Xu, W., San Lucas, A., Wang, Z., & Liu, Y. (2014). Identifying microRNA targets in different gene regions. *BMC bioinformatics*, *15*(7), S4.

Yang, B., Lin, H., Xiao, J., Lu, Y., Luo, X., Li, B., et al. (2007). The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. *Nature medicine,* 13(4), 486-491.

Yarden, Y., & Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nature reviews Molecular cell biology*, *2*(2), 127.

Yin, F., Zhang, J. N., Wang, S. W., Zhou, C. H., Zhao, M. M., Fan, W. H., et al. (2015). MiR-125a-3p regulates glioma apoptosis and invasion by regulating Nrg1. *PloS one*, *10*(1), e0116759.

Yin, S., & Gao, B. (2010). Toll-like receptor 3 in liver diseases. *Gastroenterology research and practice*, *2010*.

Yu, J., Wang, F., Yang, G. H., Wang, F. L., Ma, Y. N., Du, Z. W., & Zhang, J. W. (2006). Human microRNA clusters: genomic organization and expression profile in leukemia cell lines. *Biochemical and biophysical research communications*, *349*(1), 59-68.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G. J., Wang, X. J., & Qi, Y. (2007). A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii. *Genes & development*, *21*(10), 1190-1203.

Zheng, H., Fu, R., Wang, J. T., Liu, Q., Chen, H., & Jiang, S. W. (2013). Advances in the techniques for the prediction of microRNA targets. *International journal of molecular sciences*, *14*(4), 8179-8187.

Zhu, Y., Yu, J., Yin, L., Zhou, Y., Sun, Z., Jia, H., et al. (2016). MicroRNA‐146b, a Sensitive Indicator of Mesenchymal Stem Cell Repair of Acute Renal Injury. *Stem cells translational medicine*, *5*(10), 1406-1415.

Zorc, M., Skok, D. J., Godnic, I., Calin, G. A., Horvat, S., Jiang, Z., et al. (2012). Catalog of microRNA seed polymorphisms in vertebrates. *PLoS One,* 7(1), e30737.

10.1. Appendix 1

The ErbB signalling network

Level of signalling	Gene
Ligands	*EGF*
	TGFA
	BTC
	AREG
	EREG
	HBEGF
	NRG1
	NRG2
	NRG3
	NRG4
	EPGN
Receptors	*ErbB1*
	ErbB2
	ErbB3
	ErbB4
Adapters	*GAB1*
	GRB2
	PLCG1
	HRAS
	KRAS
	NRAS
	GAB1
	GRB2
	PLCG1
	HRAS
	KRAS
	NRAS
	PIK3CA
	PIK3CB
	PIK3CD
	PIK3R2
	MAP2K1
	MAP2K2
	MAP2K3
	MAP2K4
	MAP2K6
	MAP2K7
	SOS1
	SOS2
	AKT1
	CBL
	PRKCA
	RAF1
	SHC1
	SRC
	MAPK1
	MAPK8
	PTPN6
	PTPN11

Transcription factors	*JUN* *FOS* *EGR1* *MYC* *ELK1* *SP1* *STAT1* *STAT3*

Toll-like signalling network

Level of signalling	Gene
Receptors	*TLR1*
	TLR2
	TLR3
	TLR4
	TLR5
	TLR6
	TLR7
	TLR8
	TLR9
Adapters	*MYD88*
	MAL
	TICAM1
	TRAM1
	IRAK1
	IRAK2
	IRAK4
	TRAF3
	TRAF6
	RIPK1
	TAB2
	TAB3
	MAP3K7
	MAP2K3
	MAP2K4
	MAP2K6
	MAP2K7
	MAPK8
	MAPK14
	CHUK
Transcription Factors	*IKBKB*
	CREB1
	JUN
	FOS
	NFKB1
	NFKB2
	RELA
	RELB
	REL
	IRF3
	IRF7

Notch signalling network

Level of signalling	Gene
Ligands	*DLL1*
	DLL3
	DLL4
	JAG1
	JAG2
Receptors	*NOTCH1*
	NOTCH2
	NOTCH3
	NOTCH4
Transcription factors	*MAML1*
	RBPJ

10.2. Appendix 2

The scripts used for the generation and visualization of the results of this work can be accessed online via the following link:

https://tinyurl.com/y9gdnhk9