Genetic variation in human miRNAs: functional consequences and involvement in Cancer

Ignasi Torruella Loran

TESI DOCTORAL UPF / 2016

DIRECTOR DE LA TESI

Dra. Yolanda Espinosa Parrilla

INSTITUT DE BIOLOGIA EVOLUTIVA (CSIC-UPF)

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

A la meva família;
en especial al meu germà, als meus pares i a l'Ana.
Sense vosaltres res d'això tindria sentit.

Agraïments

Aquesta feina ha estat possible gràcies a la Universitat Pompeu Fabra (UPF), al Departament de Ciències experimentals i la Salut (DCEXS), al finançament del *"ministerio de ciencia i innovación"* (micinn), al *"Servicio publico de empleo"* (sepe) i la direcció de la Dra. Yolanda Espinosa Parrilla. En especial vull agrair la oportunitat que em va donar la Yolanda, cinc anys enrere, de formar part del seu grup de recerca, així com la confiança que ha mantingut en mi en tot moment i la seva disposició i cordialitat al llarg dels anys, moltes gràcies.

Molta més gent m'ha ajudat al llarg d'aquest procés i mereix ser mencionada en aquestes pàgines, a tots ells el meu sincer agraïment i etern record:

Als meus companys de programa de doctorat, el talent i la generositat dels quals m'ha sorprès moltíssim, que m'han ajudat cada dia amb un somriure; teniu el meu respecte i admiració. En especial a la Jessica i el Javi, que mereixerien una pàgina d'agraïments cadascun.

A les meves companyes de grup i de despatx amb qui he passat més hores en els últims anys. Neus, Alícia, Maria, Koldo i Carla, gràcies per la vostra estima, ajuda, consell i alegria, serà difícil trobar millors companys.

Moltes gràcies també als visitants que veníeu per descansar i ajudar-me sempre que ho he necessitat, a la Núria, la Mònica i la Lara. Moltes gràcies també a la Judit per la seva vocació d'ajut, la seva professionalitat i la seva alegria.

Gràcies al Marc, al Javi i a la síndic de greuges de la universitat, per ser-hi quan els vaig necessitar.

Gràcies als amics que he fet aquests anys i que mai oblidaré:

Als companys de futbol, basquet i voleibol, Diego, Nino, Juan, Joan Pau, Marc P., Laura, Irene, Marco, el Max, la Marina i tants altres; a tots moltes gràcies.
Als meus companys de cartes, cafès i tuppers. A l'Irene H., al Ramón, al Jose, al Javi, a la Jessica, al Guillem, a l'Irene L., la Raquel, la Claudia, al Marc dM, al Lukas, a l'Aitor, a tots vosaltres que sempre m'heu fet riure.
Al Guillem i al Javi, perquè un birrabox sol no és el mateix.

A la meva família, l'estima i l'orgull dels quals és la meva força de voluntat i motivació diària. Ells han fet possible aquesta oportunitat i m'han donat suport sempre; teniu el meu amor i agraïment més sentit.

A l'Ana, que és un exemple i una guia que no m'ha abandonat mai, res del que digui li farà justícia.

Abstract

MicroRNAs are important gene regulators and major contributors to phenotypic diversity that are clearly involved in human disease. In this thesis we analysed genetic variation in human microRNAs and how microRNA variants could modify their biological effect and role in Cancer. We compared genetic variation in distinct microRNA regions and found that mature and seed regions accumulate less high frequency variants than the rest of the microRNA gene, probably as a result of purifying selection. Also, we analyzed the genetic variation of different microRNA regions in human populations and identified consistently high population fixation indexes in the seed compared with other regions suggesting the existence of local adaptation in this important microRNA region Then we studied the functional consequences of Cancer genetically associated variants, including one that we found associated with gastric Cancer in European populations, and found that these variants could alter microRNA expression and the repertoire of target gene and networks. Finally we could show allele dependent regulation of three Cancer related genes by the microRNAs carrying these variants.

Resum

Els microRNAs són importants reguladors de gens i destacats contribuïdors a la diversitat fenotípica clarament involucrats en malaltia humana. Hem analitzat la variabilitat genètica en microRNAs humans i como aquesta variabilitat pot afectar a la biologia i funció dels microRNAs en la patologia del càncer. Hem comparat la variació genètica a les diferents regions dels microRNAs i hem trobat que les regions madura i "llavor" acumulen menys variants comunes que la resta del microRNA, probablement com a resultat de l'acció de la selecció purificadora. També hem trobat que les variants genètiques més diferents en freqüència entre poblacions tendeixen a acumular-se més a la regió "llavor" que a la resta de la molècula del microRNA. Seguidament vam estudiar les conseqüències funcionals de variants associades genèticament amb càncer, incloent una variant que el nostre grup va trobar associada a càncer gàstric, i vam observar com aquestes eren capaces d'alterar l'expressió dels propis microRNAs així com el repertori de gens i xarxes gèniques regulades per aquests microRNAs. Finalment, vam poder mostrar l'existència de regulació específica d'al·lel de tres gens involucrats en càncer per part dels microRNAs que contenen les variants associades genèticament amb càncer.

Preface

Thanks to our current understanding of the human genome thousands of genetic variants have been associated with a given phenotype. This knowledge can help us, not only to understand but also to occasionally modify a given phenotype. In the case of disease, we can identify biological targets for clinical treatment but also for prevention. Some of these variants alter the biological functions of a gene in a predictable manner and, thus, their effects are well known and furthermore used for a medical purpose. Many diseases caused by such variation are now under better treatment and prevention. However, many other genetic variants that are found to be involved in a given pathology have a complex effect, which requires further study in order to be understood. This is often the case of genetic variation occurring in non-protein coding regions such as microRNAs. These molecules represent a promising target for disease treatment since some variants affecting microRNAs have been genetically associated with disease and variation in such non-coding elements is well known to alter important biological pathways.

In this regard, this thesis represents an effort to study human genetic variation in miRNAs and to further analyse the biological effect of Cancer genetically associated variants. Hopefully, this effort could contribute to a better understanding of the molecular mechanism though which microRNA variants are involved in Cancer.

Table of contents

1. INTRODUCTION

1.1 MiRNAs overview

MicroRNAs (miRNAs) are small RNAs that regulate gene expression through post-transcriptional repression. The first miRNA identified, lin-4, was discovered in a study of genes involved in the control of temporal development in *C. elegans* in 1993 (Lee et al. 1993). Twenty-five years later, thousands of new miRNAs have been found in different species all across Metazoans and the number of known miRNAs has grown to a point where these molecules have become a research field on its own (**Figure 1**). Nowadays, all the information about known miRNAs is stored in a public database called miRBase (www.mirbase.org), which is regularly updated. According to miRBase release 21 (Kozomara and Griffiths-Jones 2014), 2588 mature miRNAs are known for humans, 1915 for mouse and 466 for fly, among the repertoire of other species. These numbers keep varying as new discovered miRNAs are incorporated and, sometimes, the already annotated ones are discarded due to experimental evidences or renamed according to the established rules (See BOX1 miRNA nomenclature).

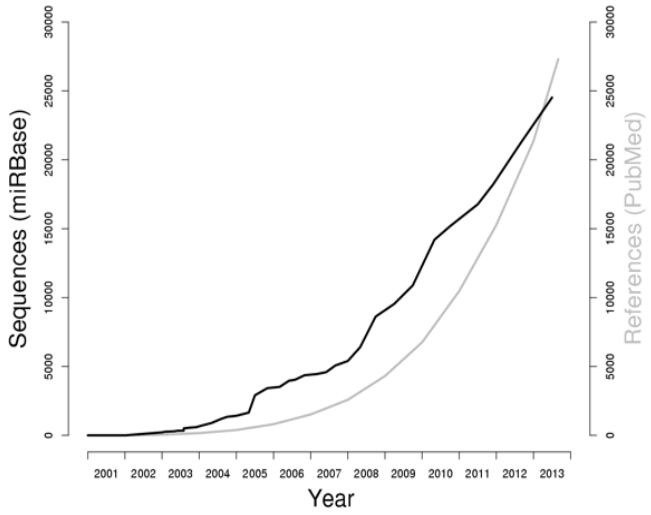

Figure 1. Variation in the number of miRNA sequences on MiRBase and in the number of references in Pubmed with the word *miRNA* across time (in years). Adapted from www.mirbase.org

Classical definition of a miRNA requires it to be expressed, to fold in secondary hairpin structure and to be able to load into DICER and RISC (See Section 1.2 miRNA biogenesis). However, not all the identified new miRNAs fulfil these criteria (Berezikov et al. 2006), and thus, revising the current repertories of miRNAs is very necessary.

MiRNAs can originate in a genome through several processes mainly based in the appearance of an RNA product able to fold in a hairpin-like structure (Berezikov 2011a). Gene duplication is a common process through which new paralogous miRNAs can be acquired and to further differentiate from the original miRNA. This process generates families of miRNAs very similar in sequence, which can even share the same seed region (Berezikov 2011a).

Paralogous miRNAs can be located in the same genomic region or in distal regions, even in different chromosomes. MiRNAs can also appear de novo in introns, as those are genomic regions already transcribed that do not require a specific promoter and that frequently form hairpin structures. Similarly to pseudogenes or other small ncRNAs (sncRNAs), miRNAs appear easily in transcribed genomic regions (Berezikov 2011a). Transposable elements (TEs) are considered an important source of genomic novelty and in fact, many Transposable Elements derived miRNAs have been identified in the human genome (Piriyapongsa et al. 2007; Borchert et al. 2011; Ahn et al. 2013). Finally, another important source of *de novo* miRNAs, are miRNAs themselves. Antisense miRNAs may be originated from the transcription of the opposite DNA branch of an existing miRNA. Those DNA regions are usually imperfect palindroms that often fold in hairpin structures like the ones of miRNA genes (Berezikov 2011a).

miRNAs do not appear gradually along evolution. Contrary, miRNAs suffer several bursts and expansions along evolution. The main ones coincide with the appearance of vertebrates, placental mammals and lately primates (Hertel et al. 2006; Iwama et al. 2013; Hertel and Stadler 2015). For this reason some authors claim that miRNAs have expanded in parallel to biological complexity, and thus their role and evolutionary impact must be studied.

BOX1 miRNA Nomenclature: In order to classify miRNAs easily and to be distinguished from other types of molecules, miRNAs are named using letters and numbers following some criteria:

As several miRNAs are conserved along species, the proper nomenclature starts with three letters that indicate the organism where that specific miRNA has been found (Example: the prefix *"hsa"* is used for humans, *"ptr"* for chimpanzees, *"cel"* for *C. elegans*). Usually when the prefix is missing, is the mature miRNA that we are identifying, whereas precursor miRNAs are usually referred to using the species prefix.

Then, after the species letters, the number that identifies that particular miRNA follows the letters *"miR"* (for the mature) or "mir" (for the precursor) in order to identify them as miRNAs (Example: hsa-mir-938 or miR-938). This number is basically sequential, meaning that if the last published miRNA in mouse is miR-352, the next miRNA identified should own the number 353. Nonetheless, there are some exceptions. For instance, if a miRNA is discovered in *Drosophila melanogaster* with a sequence identical to human miRNA miR-121, this miRNA might be named dme-miR-121.

Genes that encode for a miRNA are written in capital italics letters, like all human genes, with the code "MIR" ahead (Example: *MIR938*).

The mature miRNA names are followed by either a 3p or a 5p depending on the branch of the miRNA. As previously mentioned usually the species letters are not used for mature miRNAs (Example: miR-92-3p, miR-92-5p). Originally the 3p branch of the miRNA was referred as passenger miRNA and was annotated with an asterisk at the end of the miRNA code (miRNA* or miR-92*). Due to the low expression levels observed, passenger miRNAs were considered miRNA biogenesis products without function or relevant effect. However, several examples are now known where passenger miRNAs bind to mRNAs and repress its expression and therefore this nomenclature is not used anymore.

Letters at the end of the miRNA name, following miRNA number, indicate mature miRNA very similar in sequence in the same organism (Example: hsa-miR-121a, hsa-miR-121b). In a similar way, numbers at the end of the miRNA denote distinct miRNA genomic loci that produce identical

mature miRNA sequence, meaning miRNA paralogous genes (Example: hsa-miR-196a-1, hsa-miR-196a-2).
All these conventions follow the criteria described in (Ambros et al. 2003) but, despite being the general criteria, and more importantly, the one used by miRBase, some publications might follow different rules. As those are the most used, and the ones chosen by miRBase, I will use these conventions all along this thesis.

a) miRNAs as non-coding RNAs

Since the discovery of the central dogma of molecular biology the interest of molecular genetics was focussed in finding those regions of the genome of a given organism that encoded for each of the known proteins of that species. However, after the publication of the human genome and the finding of an unexpectedly low number of protein-coding genes, the study of the genome without protein-coding ability has increased exponentially. Those regions of the genome, that were even referred to as "Junk DNA", are now known as non-coding genome or non-coding DNA (ncDNA) and is transcribed into a very diverse group of RNAs with different functions and origins, the non-coding RNAs (ncRNAs) (Mattick 2011; Clark et al. 2013; Morris and Mattick 2014). Thanks to the expansion on high-throughput methodologies such a tilling arrays and RNA-seq, and consortia such as ENCODE or the RIKEN project (Riken et al. 2001; Hayashizaki 2003; Consortium 2004; Qu and Fang 2013) we know that up to 90% of the eukaryotic genome is transcribed. However, as previously

mentioned, a great part of it has no coding capacity. Lately several members of this group of RNAs have been found to perform a regulatory function. In fact, these RNAs regulate a wide variety of biological processes and thus they are classified in two main groups according to length: Long non-coding RNAs (lncRNAs) and sncRNAs (Mattick 2011; Clark et al. 2013; Morris and Mattick 2014).

lncRNAs are a functionally heterogeneous group of RNA transcripts, which can vary from few hundreds to thousands of nucleotides (ntds) in length (Morris and Mattick 2014). These ncRNA subgroup emerged thanks to the efforts of several genomic consortia and the recent employ of deep RNA sequencing and tilling array technologies which revealed that a large fraction of mammalian transcriptomes are comprised by long molecules of transcripts with few or none peptide coding capacity (Morris and Mattick 2014). Although many efforts have been made in understanding the role of these molecules, still we can tell little about the function of most of the known lncRNAs. However, some of their functions are slowly discovered and many examples of individual lncRNAs have been deeply studied and successfully understood (Orom et al. 2010; Clark and Mattick 2011). This is the case of Homeobox transcript antisense intergenic RNA (HOTAIR) whose expression has been found to increase in many tumors compared with healthy adjacent tissue and for which many

aspects of its role in tumorigenesis are now known (Zhou et al. 2014).

On the other hand, sncRNAs are a class of non-coding elements comprised by different kinds of RNAs, usually smaller than 200 ntds. The most studied of these RNAS, miRNAs and small interfering RNAs (siRNAs), are involved in gene regulation through processes like the RNA interference (RNAi) pathway (Clark et al. 2013; Morris and Mattick 2014). RNAi pathway is based in the interaction between the regulator RNA, the targeted mRNA and a complex of proteins, among which the Argonaute protein family stands out. Other types of sncRNAs are piwi-interacting RNAs (piRNAs), small nuclear RNAs (snoRNAs), whose degradation can produce other types of sncRNAs (sno-derived RNAs (sdRNAs)), and small nuclear RNAs (snRNAs) that participate of the RNA splicing, among others (Mattick 2011).

Within non-coding RNAs, miRNAs are a large family of small molecules (22 ntds) that recognise mRNAs by base pairing and consequently regulating their expression. In the following sections (1.2 miRNAs biogenesis, 1.3 miRNAs function, 2.2 Variation in miRNA and 3. miRNAs in human disease), I will describe the most relevant aspects of their synthesis and role in cellular biology.

b) miRNAs biogenesis

MiRNAs are the final product of a maturation process that starts in the nucleus with the transcription of the miRNA gene by RNA polymerase II or III (**Figure 2**). The first miRNA transcript is called pri-miRNA and can consist of the primary transcripts of several miRNA genes (Bartel 2004; Krol et al. 2010). For this reason, pri-miRNAs can have a wide variety of lengths and be larger than what we usually call miRNA genes. Duplication events can generate groups of miRNAs located in the same genomic region called miRNA clusters, although this is not the most common way in which clusters originate (Marco et al. 2013). Similarly to individual miRNAs, clustered miRNAs can either locate in intergenic regions or within introns in a gene. Individual and clustered miRNAs located within a gene are transcribed together with its hosting gene, whereas intergenic miRNAs can have their own Transcription Starting Sites (TSS). Usually intergenic miRNAs TSSs are located within 2 kb of their miRNA gene (Saini et al. 2007). A significant part of these groups of miRNAs have lengths between 3-4 kbs (Bartel 2004; Saini et al. 2007).

It is worth noticing that more than 90% of the human miRNAs are found clustered together or in the same transcriptional unit (Muinos-Gimeno et al. 2010). In fact, as a single miRNA can regulate hundreds of genes, and a gene can be repressed by a similar number of different miRNAs, some

authors postulate that clustered miRNAs could be regulating genes of the same biological pathway (Muinos-Gimeno et al. 2010).

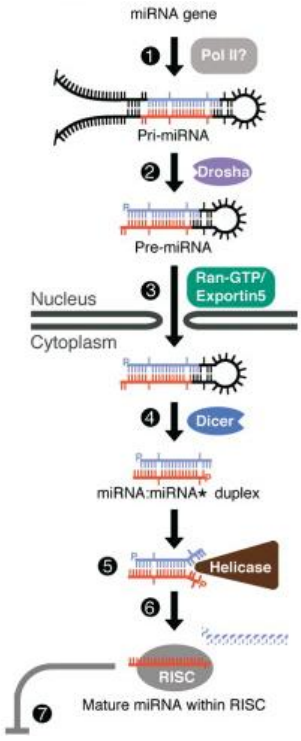

Figure 2. MiRNAs biogenesis. Adapted from Bartel *et al* 2004

The Drosha RNAse III endonuclease (Drosha) cuts the pri-miRNA into a 70 ntds RNA precursor molecule (pre-miRNA) folded forming individual hairpin structures. Some miRNA genes are located in short introns that are transcribed together with the mRNA of the gene and are processed by the splicing machinery, instead of Drosha, directly producing a pre-miRNA. The pre-miRNA leaves the nucleus through Exportin5 and Ran-GTP, to the cytoplasm where the rest of the maturation

process happens. Once in the cytoplasm pre-miRNAs are recognized and cleaved by Dicer into an RNA duplex composed by two potential mature miRNAs (20-22 ntds long) imperfectly paired (the 5' and the 3' branches of the hairpin)(Bartel 2004; Filipowicz et al. 2008; Krol et al. 2010) (**Figure 3**).

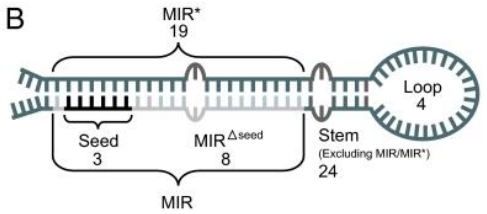

Figure 3. MiRNA hairpin structure. Adapted from Saunders *et al* 2007

The RNA oligonucleotide originated from the 3' branch of the hairpin precursor was originally referred to as passenger miRNA (miRNA*) (See **BOX1**). Because its low expression levels detected compared with its complementary branch, it was initially believed to lack a function and thus to be discarded. However, several authors demonstrated the expression an function for both molecules in the case of several miRNAs and we know now that although less expressed, passenger miRNAs, might also be functional and should be considered (Bartel 2004; Krol et al. 2010).

c) miRNAs action

Each of the two mature miRNAs that result from the cleavage of the miRNA hairpin is a potential gene regulator with a different target gene spectrum, able to target hundreds of genes (Friedman et al. 2009). In a similar manner, mRNAs regulated by miRNAs are usually targeted by thousands of them. Hence, miRNAs should be considered not only as gene regulators, but also as gene networks tuners (Lewis et al. 2003). As previously mentioned, miRNAs require partial complementarity between them and the mRNA they regulate in order to repress their expression. However, the region comprised between the second and the eighth ntds of the mature miRNA (starting from the 5' end), what is called seed region, requires a perfect complementarity with the target site in the mRNA. This knowledge, together with the ability to predict the stability and folding of RNA molecules, allow predicting interactions between miRNAs and mRNAs in order to identify potential target genes for given miRNAs. Several algorithms have been built in order to fulfill this objective, among them Targetscan, miRanda and PITA are widely used (See BOX2).

Once miRNAs bind to the target site of their targeted genes, usually at the 3'UTR, they prevent protein synthesis either by translational repression of that specific transcript or the deadenylation and further degradation of it (Hutvágner and

Zamore 2002; Krol et al. 2010). When the binding between the target site of the mRNA and the miRNA has enough complementarity, the mRNA will be cleaved. Interestingly, whereas the mRNA is destroyed, the miRNA is still active, allowing it to additionally repress other mRNAs (Hutvágner and Zamore 2002). On the other hand, if the complementarity between miRNA and target mRNA is insufficient, the mRNA won't be cleaved, instead, its translation will be repressed (Bartel 2004). The last is the most common scenario in animals, whereas plant miRNAs have higher complementary between miRNAs and their targets and thus their regulatory action is commonly performed though mRNA degradation.

BOX2: The miRNA target prediction tools. Nowadays there is a wide set of tools available to work with miRNAs, from target prediction tools to artificial molecules that mimic miRNA behaviour in functional analysis. In many ways, to identify a miRNA target is the best way to understand the role of a given miRNA in a specific scenario. However, an individual miRNA can target thousands of mRNAs and validating these interactions, despite being necessary, it is also costly and time consuming.

For this reason target prediction algorithms are among the best and more used miRNA tools. Many software applications are available to the user that helps to predict miRNA-mRNA interactions. All of them might consider the four main principles to predict miRNA targets: seed match, site conservation, free energy of the duplex and site accessibility (Peterson et al. 2014).

As previously described, the seed is the region comprised between the second and eighth ntds (starting from the 5' end) of the mature miRNA molecule. This region requires a proper Watson-Crick match with its complementary site in the target mRNA. Nevertheless, as our understanding of seed

is still incomplete there are several types of seed considered by the distinct target prediction applications. Those differ in the number of ntds considered (six, seven or eight) and weather they consider or not the first nucleotide of the mature miRNA molecule.

Conservation refers to the existence of the target site across species. This is considered prove of functionality as a functional interaction is more likely to be conserved than false interactions.

In addition, Gibbs free energy is a measure of stability in biological systems. The lower the free energy, the more stable a biological system is. Thus, if a miRNA-mRNA interaction is predicted to have low values of free energy, that interaction will be more stable and as a consequence more likely to be functional.

Site accessibility is the last of the most considered features by target prediction algorithms. After transcription, mRNA molecules fold themselves in secondary structures that can interfere with the miRNA-mRNA interaction.

Other features can be considered when trying to predict if a miRNA binds to a specific mRNA. Among them the target site abundance is of particular interest as most miRNA-mRNA interaction are known to happen more than once in the same 3'UTR.

Several reviews have addressed the question on which is the best target prediction algorithm available.(ReyesHerrera and Ficarra 2012; Dweep et al. 2013; Vlachos and Hatzigeorgiou 2013; Peterson et al. 2014) Although there is a wide variety of prediction software available, the major conclusion seems to be that there is not an option more reliable than the rest, but a more reasonable algorithm depending on the user needs. Therefore the user should identify his needs before choosing a given target prediction algorithm. However it is recommendable to consider issues such us version update, maintenance and user adjustability besides the features considered in the prediction. The more used and recommended might include Targetscan, miRanda, PITA and DIANA.

d) miRNA function

Although miRNAs are expressed in many tissues, and their involvement in nearly all biological processes is quite clear, regulation of neuronal processes, animal development, apoptosis, cell cycle control are among the most studied roles of miRNAs.

Brain represents a complex organ with many different cell types in which miRNAs have been found differentially expressed. For instance, miR-132 is found in cortical neurons, whereas mir-134 is localized more precisely in the synaptic sites, where it controls the development of dendritic spines. Also, miR-124a, the most expressed miRNA in mammalian brain, controls the maturation of neurons (Kloosterman and Plasterk 2006).

Many aspects of animal metabolism depend on miRNAs as well. For instance, miR-375, which is expressed in the pancreatic islands, supresses insulin secretion through the regulation of Myotrophin (*Mtpn*)(Poy et al. 2004). Also, flies lacking miR-14 have been found obese and present high levels of triglycerides (Xu et al. 2003).

miRNAs have also a role in cell death control, miR-14 is a perfect example of the multiple functions that miRNAs may

have. Often Cancer appears as a result of a failure in the apoptotic pathway. Cancer is one of the most studied pathologies in which miRNAs have been found to be involved (Section 3.2 describes the role of miRNAs in Cancer). Thus, there is plenty of information about the role that miRNAs play in apoptosis, and the study of apoptosis involved miRNAs is very relevant for the treatment of Cancer patients. For instance, the mRNA encoded by *BCL2* (B cell lymphoma 2), whose protein controls the release of cytochrome c in the firsts steps of apoptosis, is a direct target of miR-15 and miR-16. Interestingly, these two miRNAs are found down-regulated in the majority of chronic lymphocytic leukaemia (CLL), and their regulation of *BCL2* induces apoptosis (Cimmino et al. 2005). The products of other genes from the apoptotic pathway such as *APAF1*, *BAK*, *BAX* or *FASL* are also targeted by miRNAs, and an alteration of this regulation can lead to different types of tumours (Su et al. 2015).

Many other examples of miRNAs implicated in development, cell signalling and cycle, and other aspects of the metabolism are known (Kloosterman and Plasterk 2006). However, the understanding of miRNA function might be easier through the study of their role in disease (See section 3 MiRNAs in Disease).

BOX3: The miRNA tools for *in vitro* and *in vivo* analysis.
To fully characterize interactions predicted by target prediction algorithms, functional experiments should be

performed. Besides the usual molecular biology tools such as microarrays or RNA sequencing, miRNA emergence has brought a set of new/adapted tools to aid the study of miRNAs. Antimirs, miRNA mimics and expression vectors are of particular interest.

The most common approach to validate the interaction between mRNA and a miRNA the use of luciferase reporter constructs. The 3'UTR is cloned next to the luciferase reporter gene, and constructs are then transfected into cells that will process them. Luciferase expression appear reduced in the presence of any miRNA able to interact with the 3'UTR (Van Rooij 2011). The changes in luciferase expression can be easily detected through its luminescent activity. This approach requires the use of synthetic miRNAs, when the used cells don't express the miRNA of interest to confirm the specific interaction between the miRNA and the mRNA. However, systems to inhibit miRNAs are also an option when the studied miRNA is expressed in the cell line. Some of these tools are described in this section.

miRNA mimics are synthetic RNA molecules modified to improve stability and recognition by the cell machinery (Van Rooij 2011). In order to prevent the opposite strand to develop a function, this RNA molecule has specific alterations that restrict RISC incorporation. The main advantages of using miRNA mimics, is that that they are very easy to use, however this is not the natural scenario in which miRNAs act and a lot of information might be lost in the process, such us the effect of the opposite branch or the way the precursor molecule interacts with the cell machinery. MiRNA expression vectors on the other hand, allow the user to mimic a more natural behaviour of the miRNA by cloning a miRNA gene in a vector that contains sequences in the regions flanking the cloning site that force the RNA molecule to fold in a hairpin and thus promotes the interaction with the cell machinery. Despite being more difficult to work with and that these molecules are still synthetic, their behaviour resembles much more the natural one, and allow to study the effect of both strands of the miRNA and their interaction with the proteins that affect the miRNA maturation process. Another way of studying miRNA function is by inhibiting the function of endogenous miRNA. To do so different RNA

molecules can be used. On one hand antimirs are antisense RNA molecules modified to increase its affinity for a given miRNA to which they will bind in order to repress its function. In a different manner miRNA sponges are RNA molecules harbouring miRNA target sites to block its function. These molecules can inhibit the function of several miRNAs at the same time (Van Rooij 2011).

Besides the usual transcriptome microarrays, miRNA arrays have become a common tool to study the presence and expression level of miRNAs in different types of samples (Shingara et al. 2005; Yin et al. 2008). This tool is based on the hybridization of labelled miRNAs with a pool of miRNA probes in the microarray. RNA is extracted from the sample of interest (Cells, tissue biopsy, etc.) and then, small RNAs are labelled with fluorophore molecules using RNA ligase. Such analyses allow the parallel study of many miRNAs.

Similarly, next generation sequencing techniques can also be used to detect and quantify miRNAs (Landgraf et al. 2007; Morin et al. 2008). Moreover, deep sequencing offers several advantages over microarrays. As well as microarrays, sequencing allows the analysis of many miRNAs at the same time, but contrary to microarrays, it is not subject to probes design, and thus, it may allows for a more sensitive measure, the detection of new miRNAs as well as polymorphic variants of already known miRNAs. It is also worth noticing that contrary to microarrays, deep sequencing measures the absolute abundance of miRNAs in a sample.

Nevertheless, real-time quantitative Polymerase Chain Reaction (RT-qPCR) is still the most common approach to measure miRNA abundance. The main challenges that this approach presents are miRNAs short length and their lack of a polyA tail. However, these limitations can be resolved with the use of polyadenylathion enzymes. The design of primers is crucial in order to allow sensible measurements. RT-qPCR also allows the analysis of several miRNAs at the same time, but not many.

Altogether, these tools allow us to study miRNAs and to better understand their role in biology, and even in human disease (Van Rooij 2011).

2. Human Genetic Variation:

Genetic variation could be defined as the collection of changes that occur in the DNA sequence that constitutes the genome. Human genetic diversity has long been studied, both to infer human evolutionary history and to understand the biology behind disease. Thanks to the growth of sequencing and array technologies there is a burst in the knowledge of genetic diversity, which is precisely annotated in different databases, most of them available to the public. Some examples are dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/index.html) and 1000 genomes Project (http://www.1000genomes.org) (Abecasis et al. 2012; Consortium 2015).

Polymorphic changes in the DNA sequence can occur in many ways. From all the variation observed in the human genome, 90% are point mutation described as Single Nucleotide Polymorphisms (SNPs), which are ntds of the genome that are found to vary among two, three or even the four ntds that form the DNA genetic code. Similarly some ntds can be either added or eliminated in a single point in the genome. This variation is produced by single nucleotide insertions or deletions called indels (Abecasis et al. 2012; Consortium 2015).

On the other hand, other types of variation affect more than a single nucleotide; this variation is called structural variation

(SV). Despite being less abundant, structural variation affect larger fragments of the genome and can also affect the individual in many ways. This type of variation can consist of inversions, deletions, duplications or copy number polymorphisms of sequences with a length that can vary from few ntds to megabases or even large chromosome fragments.

Lately, the best attempts to study human genetic diversity are performed by big consortiums similar to *the 1000 genomes project* (Abecasis et al. 2012), which catalogues different kinds of genetic variation across human populations. This variation includes SNPs, indels and SVs. In fact, of the 100 million variants available on the public dbSNP catalogue, 80 million have been provided or validated by *the 1000 genomes project*. According to this consortium, a typical human genome, which is estimated to be 3.2 Gb long, differs from the reference genome at 4.1 to 5.0 million sites. Although most of these variants (>99.9%) are SNPs and short indels, structural variants, are able to affect larger fragments of the genome. From those 4.1-5.0 million variable sites in a human genome, only 149-182 are protein truncating variants and between 10,000 and 12,000 variants are able to alter protein sequence. On the other hand, around half a million variants affect regions of the genome with known regulatory function but no coding capacity (Consortium 2015).

Overall, it is well known that most of the human genetic variability is present in populations from the African continent. In fact according to *the 1000 genomes project* consortium, 63% of novel variants were of African origin, whereas only the 33% of them were found in Europe. In fact, those variants found to highly differ in frequency across human population of distinct geographical origin are usually considered to be very interesting in population genetics and might have important biological implications. In order to measure these differences the index F_{ST} is widely used. The F_{ST} index summarizes genetic differentiation among populations by estimating the fraction of inter-individual variance that can be justified by the intergroup variance (Weir and Hill 2002). Variants with high F_{ST} values are enriched for "non-synonymous" mutations, which lead to structural, and mainly functional, changes in the proteins encoded by them (Lewontin and Krakauer 1973; Grossman et al. 2010).

a) Genetic variation in non-coding elements

When comparing the amount of variation found in ncRNAs with the variation present in Coding Sequences (CDSs), it is clear that CDSs accumulate fewer variants than non-coding elements. In fact, in CEU population, SNP diversity is double in ncRNAs than in CDSs and one fifth of this of neutral reference sequences (Consortium 2015). However, the average

constrain on ncRNAs is higher than CDSs for indels (Consortium 2015). Additionally, there is an excess of low frequency variants in ncRNAs compared to CDSs. It is also worth noticing that among ncRNAs there are important differences between different types of regions in terms of conservation, and, for instance, those regions with higher expression levels tend to be more conserved than those with lower transcription activity (Mu et al. 2011).

b) Genetic variation in miRNAs

Since their discovery, several authors have studied the global amount of variation in human miRNA regions. Many miRNAs are known to be conserved across species and thus they are expected to develop an important biological role. However, it is also important to understand how recent evolution may have shaped miRNA genes to comprehend the importance of their role in human disease.

Some authors have shown that miRNA genes have less SNP density that their surrounding regions, but also that purifying selection has globally constrained the amount of variation in miRNA regions, especially in the mature region (Saunders et al. 2007; Quach et al. 2009; Carbonell et al. 2012). Indeed, 1000 genomes data analysis reveals that mature miRNA regions have less SNP diversity than regions surrounding them. In addition

the seed region of the miRNA was reported to have 53% less SNP diversity than the mature miRNA (Mu et al. 2011).

The deep analysis of 1000 genomes data has revealed that miRNAs with larger number of predicted target sites, have less SNP diversity than those miRNAs with fewer target sites (Mu et al. 2011). These results suggest that miRNAs with large number of predicted binding sites show signatures of positive selection, while fewer predicted target sites miRNAs, seem to be affected by purifying selection. Furthermore, there are examples of miRNA-containing regions affected by positive selection in specific human populations that contribute to the adaptation of the mentioned population to their environment (Liu et al. 2012).

Besides miRNA regions, it is worth noticing that target sites in the 3' UTR of genes predicted to be target by miRNAs are also more conserved than the rest of conserved sites in the same genomic region (Chen and Rajewsky 2006). Moreover some authors have identified SNPs in miRNA target sites, with high values of integrated haplotype scores (iHS), what indicates that a given SNP might be affected by recent positive selection in a given population (Saunders et al. 2007).

Structural variants are also able to affect miRNAs and alter their function. In fact, this type of variation is able both to increase and decrease miRNA expression. Moreover, CNVs are

capable of altering the miRNA seed and thus affect miRNA genes in a different manner (**Figure 4**) than to protein coding genes and many examples of functionally relevant CNV affecting miRNA are being investigated (Marcinkowska et al. 2011).

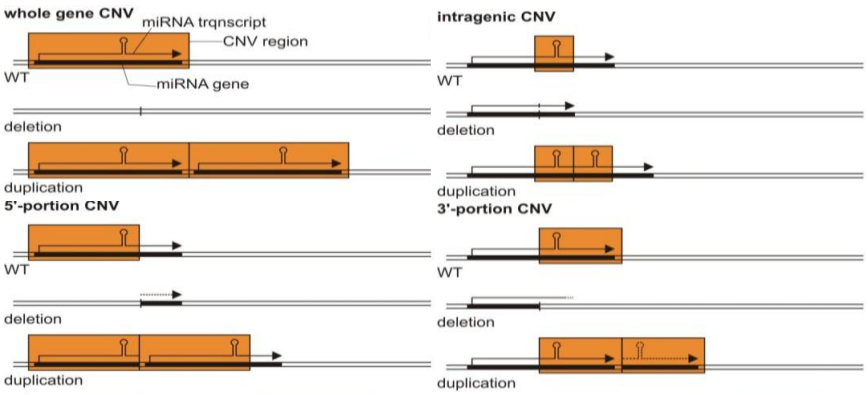

Figure 4. CNV in miRNAs Adapted from Marcinkowska *et al* 2011

Another source of variation in miRNAs are isomirs. Isomirs are different miRNA variants that differ in length or sequence, originated from the same miRNA locus (Neilsen et al. 2012). Some authors classify isomirs in three different categories depending on the type of variation: 5' isomirs are those isomirs that change in length in the 5' extreme of the canonical miRNA, 3' isomirs are those miRNA variants that differ from the canonical miRNA in the length of the 3' end and polymorphic isomirs are those isomirs that differ in sequence (Neilsen et al. 2012). The most frequent isomer type observed in animals and plants is 3' isomirs. Cell and tissue specificity is

observed in terms of isomiRs expression. Moreover, there is evidence that suggest that isomiRs could change the expression in response to biological changes. However, the context-dependent-variation is very reduced. 3' and 5' isomiRs mainly originate from a variable cleavage site of DROSHA or DICER, among other enzymes responsible for the miRNA biology, in the precursor miRNA. This phenomenon has reduced or no implications in the biological function of the miRNA unless it produces another phenomenon called seed shifting, which consists of a potential change of the seed region (Neilsen et al. 2012).

RNA editing can also affect miRNAs. RNA editing can be defined as the site-specific modification of an RNA sequence that results in a molecule different from the encoded in the DNA. Usually, RNA editing in humans is based in the conversion from adenosione to inosine (A-to-I), which is expressed as A-to-G change, since inosine is recognized as guanosine by the cell machinery. Editing in protein coding genes can modify the function of the encoded protein (Maas et al. 2001). Nevertheless, the biological role of RNA editing still remains under debate. Because A-to-G editing is catalysed by Adenosine Deaminases (ADARs), which act on double strand RNAs, editing can also affect miRNAs as they fold in hairpin structures (Blow et al. 2006; Kawahara et al. 2008). Some miRNA editing events have been described in certain species. miRNA editing can interrupt the biogenesis from the primary

forms to the mature ones. Also miRNA editing can alter miRNA function. For instance, the edition of miRNA has been related to cell development of CD4+ T cells (Negi et al. 2015).

c) Human Genetic variation and disease

Genetic variants are able to influence the phenotype of their hosts, being able to affect its health. Whereas some variants determine that an individual will develop a disease, others are able to protect it against a given pathology, favour a disease or affect the susceptibility to a given treatment.

There are many ways in which a genetic variant can be involved in disease. Most of the classical examples of genetic variation that cause disease correspond to protein truncating variants. Perfect examples are the mutations in the genes *TYR*, *OCA2* and *TYRP1* among others of the melanin biosynthesis pathway, that cause oculocutaneous albinism (Kamaraj and Purohit 2014). Similarly, a repetitive CAG sequence in the protein encoded by the Huntington's disease (HD) gene, is the responsible for this pathology (HD collaborative and research group 1993). Interestingly, the length of this repetitive element determines many aspects of the disease onset (Seong et al. 2005).

Fundamentally, most mutations that have a high disease risk are rare because of the effect of natural selection on them. As a consequence, these diseases causing variants can have a reduced geographic distribution, probably due to a recent origin. Moreover, the resulting disease might show a high risk within the members of the same family. The classic genetic studies of human variants that cause disease are mainly focussed on Mendelian disorders, pathologies that follow the first rules of genetic heredity described by Gregor Mendel, and usually affect members of the same family. Mutations causing Mendelian disorders have a major effect on disease risk and thus, few genes allow the organisms to survive with those mutations. However, many disorders don't follow Mendelian rules, despite a clear, though less evident, genetic component. These "complex diseases" are thought to be influenced by several genetic components as well as environmental factors and their study requires powerful tools such us Genome Wide Association Studies (GWAS). GWAS studies use commercial SNP chips to capture most of the variability present in an individual trying to relate it with the phenotype of a group of individuals shearing the same phenotype. This approximation relies in the hypothesis called "common disease common variance" which defends that common disease can be explained by genetic variants present in more than 1% of the population (Manolio et al. 2009). Although GWAS have been very useful to identify variants that explain phenotypic traits and human diseases, still many heritable phenotypes can not be explain by

the common variations usually found in GWAS (Manolio et al. 2009). Moreover, due to the on-going human demographic explosion that started with the agricultural development, our species is suffering an excess of rare variation. Until recently our technology could not analyse properly this excess of low frequent variants with an excess of functional substitutions and hence, potentially involved in disease. For these reasons, the use of new sequencing platforms, that allow the sequencing of whole-genome at high quality, in studies with thousands of individuals, together with the deep study of the functional consequences of non-protein coding variation (especially in regulatory regions such as miRNAs) will represent a necessary challenge in the interpretation of personalized genomic risk profiles, which are the future of modern medicine (Casals and Bertranpetit 2012).

3. MiRNAs in human Disease

a) MiRNAs in disease

Considering the degree of conservation in miRNAs it is likely that loss-of-function studies with miRNAs would have resulted in severe developmental phenotypes on the model organisms. However, this has not been the case for the majority of miRNAs in mammals (Mendell and Olson 2012). Nevertheless, knock-down of key enzymes for the miRNA maturation

processes are known to cause severe phenotypes (Chen et al. 2008a; Tao et al. 2011). These findings suggest that although miRNAs are key regulators of many biological processes, their function is usually redundant and thus deletion of a single miRNA is rarely lethal but can influence disease.

In fact, many examples of miRNAs involved in disease are now known thanks to a wide variety of studies and their deficiency or excess has been correlated with numerous pathologies. Moreover, numerous efforts are being performed in this area, as miRNA are considered great targets for therapeutic applications, a topic that will be introduced at the end of this section.

Changes in the expression levels of miRNAs can lead to disease either through an overregulation of their targets, or an unsuccessful control of them. For instance, altered levels of miR-106b have been found to be involved in prostate Cancer through the regulation of caspase-7 (Hudson et al. 2013). These changes in the expression levels of miRNAs can appear as a result of genetic variation in the precursor miRNA molecule or in their promoters. Genetic variants in the precursor miRNA molecule can alter the stability of the hairpin and thus the recognition of this structure by the miRNA biosynthesis machinery. For example a polymorphism in miR-196a-2 has been found to alter the expression levels of one of its mature forms, being able to influence in the genes and networks

regulated by this miRNA. Moreover this variant has been found to associate with Breast Cancer and other types of Cancer (Hoffman et al. 2009).

Not only nucleotide variants affect miRNA biology leading to disease, structural variants are also able to affect the expression of miRNAs causing a pathogenic phenotype. For instance, the microdeletion in 22q11.2 affects the gene *DGCR8* located in this region, which is essential for the maturation process of a set of miRNAs. Moreover, this deletion also affects the miRNA miR-185 located in the same genomic region. Microdeletion in 22q11.2 has been related to several psychiatric disorders, especially with Schizophrenia (SCZ). Mice modelling this microdeletion showed less expression of miR-185 and present abnormalities in the morphology of dendrites of the hippocampus neurons (Xu et al. 2013). Other structural variations that alter miRNA expression have also been related to SCZ and other neuronal disorders (Xu et al. 2010b).

Genetic variation in miRNAs can not only alter miRNA expression, but also modify their seed leading to changes in the regulation of their target genes and thus to disease. This is the case of the first identified miRNA variant found to cause a mendelian disease. Several SNPs in the seed region of miR-96 were found to be responsible of deafness (Mencía et al. 2009). This finding was later confirmed in a mice model (Lewis et al. 2009).

Moreover, several groups have attempted to catalogue the miRNAs involved in diseases and comprehend the association between miRNAs and disease in humans (Lu et al. 2008). Although many groups keep finding new associations, the databases that store miRNA disease associations such as the Hunan MicroRNA Disease Database (HMDD, http://www.cuilab.cn/hmdd) or miR2Disease (http://www.mir2disease.org) are far from completeness, moreover, the complexity in pointing to a precise diagnosis in modern medicine contributes to slow this process.

MicroRNAs in cardiovascular Disease

MiRNAs are essential actors in the cardiovascular system both at the developmental and functional levels. For instance, down-regulation of miR-29 is known to induce fibrosis in the heart as several of the genes this miRNA targets are involved in the formation of the extracellular matrix which is one of the first steps of fibrosis (Mendell and Olson 2012). As with other pathologies, cardiovascular disease related miRNAs can either protect or induce disease. An example of this phenomenon is illustrated on cardiac hypertrophy. Expression of miR-23a can induce hypertrophy in the heart and its down-regulation through antimirs can protect against it. Moreover, miR-133 has a protective role against cardiac hypertrophy, and it's knock-down in mice is shown to induce this pathology (Small and Olson 2011).

MicroRNAs in Central Nervous System diseases

Another system where miRNAs are being deeply studied is the central nervous system (CNS), which is deeply affected by miRNA regulation both at development and differentiation. As explained previously, many studies have proven that the inhibition of enzymes that participate in the miRNA maturation process produce strong phenotypes in the brain. Besides altering the general biogenesis of miRNA, abnormal expression of individual miRNAs is also known to affect neural physiology. For instance, miR-134 decreases spine volume and controls neural morphology (Xu et al. 2010b). Also, miR-124a, which is the miRNA more expressed in mammal brains, participates in the maturation process of neurons (Xu et al. 2010b).

An example of the effect that the microdeletion in 22q11.2 has in altering miRNAs expression in SCZ has already been described in this introduction. Interestingly, the same region has been found duplicated, in patients with Autism (Xu et al. 2010b). However, the role of miRNA in SCZ seems to be of a larger relevance. In this sense, association studies have been able to identify SNPs that affect miRNAs such us miR-206 and miR-198, which are expressed in human brain and whose predicted targets interact with other genes previously associated with SCZ. Moreover, the analysis of post-mortem brains

reveals different miRNA expression patterns between individuals with SCZ and controls (Hansen et al. 2007).

The analysis of miRNA expression in postmortem brains of patients with neurodegenerative disorders such us Parkinson disease (PD), Alzheimer disease (AD), Huntington's disease (HD) or Amytrophic lateral sclerosis (ALS) have also been able to relate miRNA abnormal behaviour with these pathologies (Junn and Mouradian 2012). However, it is difficult to establish causality in such findings and more detailed studies are needed in order to further understand the role of miRNAs in neurodegenerative disorders.

Analysis of post-mortem brains and cohort studies have been able to relate genetic variants affecting miRNA sites or miRNA deregulation to other neural disorders such us Rett syndrome, Tourette syndrome, Fragile X syndrome, epilepsy and Down syndrome among others (Xu et al. 2010b).

b) MiRNAs in Cancer

All the disorders that are referred to using the term "Cancer" have in common an abnormal cell growth and the potential to spread to other parts of the body through tissue invasion and tumour formation. These alterations may damage different body organs and eventually lead to death, especially without

treatment and when there is tissue invasion. Usually, the origin of the disease is a cell injury that might have a physical, a chemical or a biological origin. Several cell genes, because its physiological role might help the disease to spread (oncogenes), whereas others are responsible of preventing or stopping this pathology (tumour suppressors).

Over 14 million new cases of Cancer are estimated to have occurred in 2012 and more than eight million people died from this disease during the same year (Ferlay et al. 2015). Lung Cancer is the most mortal and abundant type of Cancer, while the second most common, Breast Cancer, is the fifth cause of death among Cancer diseases. These two types, together with Colorectal Cancer, Prostate Cancer, Stomach Cancer and Liver Cancer where responsible for more than half of Cancer cases in 2012 (Ferlay et al. 2015). Incidence and mortality vary a lot according to country, sex and age. Also, there is a great variety of treatments, with new drugs being developed every year, although surgery is still the most common strategy to face the disease. In this context, miRNAs might not only be part of the cause, but also part of the solution.

There is recurrent evidence that miRNAs are involved in Cancer. The first proves of this involvement came from studies of the proteins that participate in the miRNA maturation process, such as DICER, that produce severe tumours when knocked-out in mice (Esquela-Kerscher and Slack 2006).

Moreover, reduced and high levels of DICER correlate with different types of Cancer in humans (Jansson and Lund 2012). Also, miRNAs are usually located in genomic regions which are found altered in cells of different types of Cancer (Calin et al. 2004).

Once miRNAs were identified as relevant actors in Cancer biology; oligonucleotide miRNA microarrays were used to study the differences in miRNA expression between tumour and normal tissues. These studies lead to the identification of hundreds of miRNAs, whose expression was found deregulated in Cancer. For instance, miR-21 has been found up-regulated in many Cancer cells and thus it is considered one of the first described miRNAs acting as oncogens, oncomiRs (Volinia et al. 2006). Nowadays nearly every type of tumour analysed has a miRNA expression profile distinct to healthy tissue. Interestingly, miRNA expression levels are globally lower in carcinogenic tissue compared with normal cells (Lu et al. 2005).

Nevertheless, finding deregulation of miRNA genes when comparing tumours with adjacent healthy tissue does not necessarily implicate these miRNAs in the aetiology of Cancer. The subsequent study of the function of these miRNAs identified several genes among their targets involved in carcinogenesis. For instance, miR-15 and miR-16 are down-regulated in chronic lymphocytic leukaemia (CLL).

Interestingly B cell lymphoma 2 protein (Bcl2) is targeted by both of them, and is up-regulated in the B cells that cause CLL. Moreover the reduction of Bcl2 induces apoptosis and therefore can protect against CLL (Cimmino et al. 2005). These findings suggest that miR-15 and miR-16 may act as tumour-suppressor miRNAs. Curiously, the oncomiR miR-21 performs exactly the opposite function in similar haematological malignancies. Overexpression of miR-21 blocks apoptosis through the regulation of genes such us *PTEN*, *TPM1* and *PDCD4*. Also, miR-21 is found overexpressed in several blood malignancies (Garzon et al. 2009).

Besides differences in the expression levels of miRNAs between carcinogenic and healthy tissues, genetic variability in miRNA regions has also been analysed in the context of Cancer as well as for other previously mentioned diseases (See section 3.1 MiRNAs in disease). Interestingly, changes in miRNA sequence are capable of altering its expression and thus, may have effects similar to miRNA fluctuations. On the one hand, SNPs in promoter regions of miRNAs can alter its transcription and lead to differences in the amount of precursor miRNA molecules, and thus mature miRNAs. This is the case of miR-143 that is affected by variant rs4705342 in its promoter region. This variant has been genetically associated with prostate Cancer and it is known to alter miR-143 promoter by reducing its affinity for nuclear factors. Moreover, the addition of synthetic miRNAs can supress Prostate Cancers progression

through the regulation of *KLK2* (Chu et al. 2014a). On the other hand, changes in the miRNA precursor sequence can affect the stability of the hairpin structure and alter its maturation process causing a decrease or increase in the number of mature miRNA molecules. Two examples of these phenomena are the cases of hsa-miR-146a and hsa-miR-196a-2. Both of them are affected by genetic variants (rs2910164 and rs11614913, respectively) that alter their expression levels and that have been found genetically associated with Cancer (Jazdzewski et al. 2008; Hoffman et al. 2009; Jazdzewski et al. 2009). Moreover, changes in the expression of these miRNAs have also been associated with different types of Cancer (Jazdzewski et al. 2008; Jazdzewski et al. 2009; Xu et al. 2010a; Wei et al. 2013; Fu et al. 2014; Wan et al. 2014; Fan et al. 2015). More specific consequences of these variants are addressed in chapter one of this thesis.

c) MicroRNAs for diagnostic and therapeutics

A wide variety of body fluids, including plasma, serum or amniotic fluid, are enriched for extracellular circulating miRNAs that can be used as targets for diagnosis. Whether these molecules are performing a biological function or are just a product of cell machinery remains unclear. However their utility has been proven in different studies where expression of circulating miRNAs has been found to correlate with

cardiovascular disorders and Cancer. For instance, Boeri and co-workers were, not only able to distinguish between healthy and carcinogenic lung tissue using miRNA expression patterns, but also to predict clinical aspects such us growth rate, histology or outcome (Boeri et al. 2011). Similarly, miRNAs have been proposed as biomarkers for other type of tumors (Ellinger 2015; Sohn et al. 2015), autoimmune diseases (Zeng et al. 2014), CNS diseases (Jin et al. 2013), cardiovascular diseases and others (Oerlemans et al. 2012; Wang et al. 2013).

miRNAs have been proposed for the treatment and diagnosis of several diseases. In fact, many types of Cancers already use miRNAs as biomarkers for diagnosis (Mitchell et al. 2008; Chen et al. 2008b; Ng et al. 2009; Ellinger 2015). Considering the potential effect in health of miRNA expression deregulation, it is likely that the inhibition or supply of the disease-associated miRNA could restore the normal biological state of the affected tissue. Moreover, both the inhibition of oncogenic miRNAs and the supply of miRNAs with tumour-suppressor activity are already being investigated as treatment for different types of tumours. For example, the intravenous injection of synthetic antimirs of miR-10b in mice with metastatic Breast Cancer cells, resulted in the inhibition of the metastasis (Ma et al. 2010). Also, the reintroduction of synthetic miR-26a, which is down-regulated in hepatocellular carcinoma, inhibited Cancer proliferation and induced apoptosis in mice with this pathology (Kota et al. 2009). Although miRNA function can be blocked through the delivery

of antimirs (see **BOX3**), many aspects of treatment, either with (LNA)-modified oligonucleotides to supress miRNA function or with synthetic miRNA to restore miRNA effect, remain unsolved (Mendell and Olson 2012). However, some procedures have been able to overcome these limitations and now are promising solutions for disease handling. An example of treatment through the use of antimirs is the case of miR-122, which is essential for the replication of the hepatitis C virus (HCV). In fact, the use of (LNA)-modified oligonucleotides complementary to miR-122 supresses HCV in chimp livers without side effects or viral resistance (Lanford et al. 2010). Other treatments with (LNA)-modified oligonucleotides have been proven to treat pathologies with miRNA altered expression during the last decade (Stenvang et al. 2008), however the use of these molecules to treat pathologies caused by polymorphisms affecting miRNA sequences has not been addressed yet.

Clearly miRNAs perform a defined function in Cancer biology, behaving either like tumour suppressor or oncogenes, and their expression not only is deregulated in many tumours but also can correlate with different aspects of the tumour biology, helping both to diagnose and treat Cancer patients. However, little is known about the putative effect that genetic variation in miRNAs may have in Cancer. As previously mentioned, SNPs in miRNA regions are able to change the expression levels of its mature form or even their target repertory. Moreover,

changes of expression levels have been repetitively associated with Cancer and other diseases. The immediate conclusion is that genetic changes in miRNA genes are able to influence Cancer biology by altering their expression levels or changing the affinity for a given target gene. Despite the fact that several examples of miRNA variants involved in disease have already been described, the great majority of them remain unstudied, and their potential harm to human health is still unknown.

2. OBJECTIVES

General objectives:

The general objective of this thesis was to study human genetic variation in miRNA regions and its relationship with disease. On one hand, our intention was to study the main characteristics of variants located in human miRNA genes and, through their analysis, to better understand how natural selection has contributed to shape the distinct miRNA regions. On the other hand, we wanted to analyse the biological consequences of some of the variants most frequently associated with Cancer and try to find functional explanations to those associations through miRNA over-expression experiments.

Particular objectives:

- To construct a catalogue of genetic variants located in miRNA regions analysing their distribution among different miRNA regions and their involvement in disease.
- To analyse allele frequency differences of miRNA variants in different miRNA regions among human populations.
- To evaluate the effect of miRNA allele variants genetically associated with Cancer on miRNA expression.

- To study the effect of miRNA allele variants genetically associated with Cancer in the regulation of gene networks and pathways.
- To identify and validate potential miRNA-target gene interactions susceptible to explain the genetic associations of miRNA allele variants with cancer.

3. RESULTS

3.1 Chapter1: microRNA genetic variation, from population analysis to functional implications of three allele variants associated with Cancer

(Manuscript accepted for publication in Human Mutation)

Torruella-Loran I, Laayouni H, Dobon B, Gallego A, Balcells I, Garcia-Ramallo E, et al. MicroRNA Genetic Variation: From Population Analysis to Functional Implications of Three Allele Variants Associated with Cancer. Hum Mutat. 2016 Oct;37(10):1060–73. DOI: 10.1002/humu.23045

3.2 Chapter2: rs12416605:C/T in MIR938 associates with gastric cancer in Europeans though affecting miR-938 expression and the regulation of CXCL12 in an allele specific fashion

(Manuscript in preparation)

rs12416605:C/T in MIR938 associates with Gastric Cancer in Europeans though affecting miR-938 expression and the regulation of CXCL12 in an allele specific fashion

Ignasi Torruella-Loran[1], Xavier Muñoz[2,3], Catalina Bonet[4], Carlos A Gonzalez[4], Núria Sala[2,4] Yolanda Espinosa-Parrilla[1,5] on behalf of the EPIC gastric cancer working group.

Authors' Affiliations:

1 IBE, Institute of Evolutionary Biology (Universitat Pompeu Fabra-CSIC), Department of Experimental and Health Sciences, Barcelona, Catalonia, Spain

2 Molecular Epidemiology Group, Translational Research Laboratory, Catalan Institute of Oncology-IDIBELL, Barcelona, Spain

3 Human Molecular Genetics Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain

4 Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program, ICO-IDIBELL, Barcelona, Spain.

5 School of Medicine, Universidad de Magallanes, Punta Arenas, Chile

ABSTRACT

MicroRNAs are post-transcriptional gene regulators that participate in diverse biological pathways and with important roles in carcinogenesis. Nucleotide variants in microRNAs may contribute to cancer development causing changes in either microRNA expression or function. To look for a possible contribution of microRNA variants to gastric cancer susceptibility we performed association studies using 21 variants potentially functional because of their location in the seed, mature or precursor sequence of 22 microRNAs and gastric cancer anatomical and histological subtypes. Logistic regression analysis in 365 cases and 1,284 matched controls (European Prospective Investigation into Cancer and Nutrition cohort) showed that rs12416605:C/T, in the seed of miR-938, associated with the diffuse phenotype (p value= 0.0281, OR (95% CI)= 0.70 (0.51-0.97)) rs2114358:T/C, in the precursor miR-1206, associated with the non-cardias localization of the adenocarcinoma (p value= 0.0093, OR (95% CI)= 0.73 (0.58-0.93)) and rs3746444:A>G, in the mature miR-499b-5p and seed of miR-499a-3p, associated with the cardias CG subtype (p value= 0.0308, 0.64 (0.42-0.98)). We further investigated if differential gene regulation could be underlying the until now unreported association of gastric adenocarcinoma with rs12416605:C/T in MIR938. MicroRNA over-expression experiments in HeLa cells showed differences in the expression between rs12416605:C/T alleles and allele specific regulation of a subset of genes related to cancer and gastrointestinal disease. Moreover, among these genes differently

regulated in the transcriptome experiments by each microRNA variant we found the chemokine CXCL12 gene, which was predicted as target only by the miR-938 C allele. Luciferase-based assays validated this prediction and indicated that miR-938 could be regulating CXCL12 in an allele-dependent fashion.

INTRODUCTION

Gastric Cancer (GC) is the fifth most common cancer in the world in terms of incidence and the third leading cause of cancer death in both sexes worldwide (Ferlay et al. 2015). There are two main locations of gastric adenocarcinoma: proximal (cardia) and distal (noncardia); histologically there are two variants of the disease, the intestinal-type, which presents clearly defined glandular structures, and the undifferentiated diffuse-type, which consists of individually infiltrating neoplastic cells (Houghton and Wang 2005). GC is a multifactorial disease, whose major risk factors are environmental, such as Helicobacter pylori infection, diet or tobacco, as well as genetic (Rugge et al. 2015). A large part of susceptibility to GC (estimated to be 28% in Northern Europe) is inherited and several GC susceptibility genes have already been identified among which there are genes involved in immunoinflammatory response, DNA repair, cellular adherence, proliferative processes, mucosa protection, and Helicobacter pylori's cellular signalling pathways (Genta 1997; Loh et al. 2009; Lao-Sirieix et al. 2010; Sala et al. 2012; Saeki et al. 2013). GC is the result of a multifactorial complex process, for which a multistep model of carcinogenesis involving the effects of common low-penetrance and rare disease-causing variants is currently accepted (Correa 1992; Fletcher and Houlston 2010). Once the gastric adenocarcinoma has appeared the malignancy of GC is highly dependent on the progression and metastasis of the tumour, these are complex processes partially controlled by oncogenes and tumour suppressor genes, whose

expression is further regulated by microRNAs (miRNAs) (Esquela-Kerscher and Slack 2006).

MiRNAs are non-protein-coding small RNAs that negatively regulate gene expression post-transcriptionally participating in the control of most biological processes (Bartel 2004; Krol et al. 2010). The action of miRNAs as gene repressors is performed by means of partial complementarity to miRNA-binding sites at their target messenger RNAs (mRNAs) resulting in either degradation of the target mRNAs or inhibition of translation (Filipowicz et al. 2008). Each mRNA can be repressed by multiple miRNAs, as well as each miRNA is a potential regulator of hundreds of transcripts (Lewis et al. 2003; Friedman et al. 2009); accordingly, it has been estimated that more than half of the total human protein-coding genes are regulated by miRNAs through complex regulatory networks that control almost every cellular processes, including development, differentiation, proliferation and apoptosis, and having important roles in carcinogenesis (Kloosterman and Plasterk 2006; Friedman et al. 2009; Krol et al. 2010). Aberrant miRNA expression has been proven to correlate with the diagnosis and stage of many types of cancer and thus there are several studies running in order validate their prognostic and predictive value (Lu et al. 2005; Shen et al. 2012). In GC several miRNAs have been found differentially expressed between tumour and normal tissues and some of them are associated with progression and prognosis of GC (Lu et al. 2005; Katada et al. 2009; Kim et al. 2009; Ueda et al. 2010; Song and Meltzer 2012). For example, in tumor tissues the expression of the miR- 106b~25 cluster was correlated with tumor size and invasion,

as well as with tumor progression and metastases (Zhang et al. 2014).

Case-control association studies have pointed to several miRNAs as candidate genes to be involved in the susceptibility to GC, also genetic variants in miRNAs such as mirR-146a, miR-149, miR-196a-2, miR-499a and miR-27a have been associated with the GC susceptibility (Peng et al. 2010). Nevertheless, very little is known about the real contribution of these genetic variants to GC pathology. One exception is a recent report showing that the A allele of rs11671784 in miR-27a reduces the expression of the miRNA and it is associated with a reduction in the risk of GC (Yang et al. 2014). Acquiring a better understanding of the role of miRNA SNPs in the regulation of gastric carcinogenesis could be useful to assess individual susceptibility GC risk and to evaluate their functional importance and their possible use as a tool for diagnosis and therapy.

In this work we analysed the genetic association of 21 SNPs with GC in a cohort of 365 cases and 1284 healthy controls from the European Prospective Investigation into Cancer and Nutrition cohort. We found associations between three SNPs (rs2114358:T>C in the precursor sequence of miR-1206, rs12416605:C>T in the seed of the miR-938 and rs3746444:A>G in the seed of miR-499a-3p and mature 499-5p) and different GC subtypes. Moreover, to investigate the effect of rs2114358:T>C in the regulatory action of miR-938, we performed functional analyses based in miRNA over-expression experiments. We could detect differences between the

expression levels of the two miR-938 allele variants as well as several transcripts differentially regulated by the two rs2114358:T>C alleles. Furthermore, among these transcripts there were several ones predicted to be exclusively regulated by one of the two allele variants and found that the chemokine CXCL12, a gene largely involved in the aetiology of gastric cancer, is regulated by miR-938 in an allele an allele-dependent fashion.

MATERIALS AND METHODS

Subjects

Study subjects were selected from the EPIC cohort according to a nested case–control design aimed at assessing the genetic and environmental risk factors for GC (EurGast study). The study population has been extensively presented (Riboli et al. 2002). Briefly, the EPIC cohort includes 521,457 individuals recruited between 1992 and 2000 in 23 centers in ten European countries, from North to South. Cases were subjects having blood collected and diagnosed during the follow-up with adenocarcinoma of the stomach (code C16 of the International Classification of Diseases, 10th Revision). An independent panel of pathologists confirmed and validated the diagnosis, tumor subsite and morphology (Carneiro and Moutinho 2007). Prevalent tumors, cancer located in gastric stump as well as tumors other than adenocarcinoma were excluded. For each case, up to four control subjects were randomly selected among cohort members who were alive and free of cancer at the time of diagnosis and matched by center, gender, age (62.5 years) and date of blood collection (645 days). We selected 373 GC cases and 1,332 365 GC cases for genotyping from which eight cases and 48 controls whose DNA was not amplified or had a genotyping call rate lower than 80% were excluded from the association study. According to this, final analysis was performed

on 365 GC cases and 1,284 matched controls (Table 1). Regarding the localization of tumors, 29.3% were in the cardia or gastroesophageal junction, 49.6% were in the noncardia, 1.6% presented a mixed localization and 19.5% were unspecified; as for the histological type, 34.5% were intestinal, 35.1% were diffuse, 2.2% were mixed and 28.2% were unspecified (Table 1). Signed informed consent was obtained from all individuals. The study was approved by ethical committees at the International Agency for Research on Cancer and in each of the EPIC recruitment centers.

SNP genotyping, quality control and data filtering.

Genotyping of the 29 selected SNPs was included in a panel of 1,536 SNPs described in details elsewhere (Sala et al. 2012). Briefly, genomic DNA was extracted from buffy coat and genotyped at the Spanish National Genotyping Centre (CEGEN) by use of the Illumina BeadStation Platform and GoldenGate Technology (Illumina, San Diego, CA), according to the manufacturer's protocols. In addition to the internal genotyping controls, 5% of the samples (n=100) were genotyped in duplicate with overall agreement of 99.2% (Sala et al. 2012). Among the miRNA SNPs selected for our study, eight were excluded from analysis because of technical problems and none of the remaining SNPs deviated from Hardy–Weinberg equilibrium (HWE) among controls (Fisher exact test $p<10\text{-}4$); the total number of SNPs

successfully genotyped and available for analysis was therefore 21 (Supplementary Table 1).

Statistical analyses

Odds ratios (ORs) and 95% confidence intervals (95% CIs) in global GC as well as subgroup analysis by tumor localization (cardia/non-cardia) and by histological type (intestinal/diffuse) were estimated using unconditional logistic regression adjusted by the matching variables gender, age and country as previously described (Espinosa-Parrilla et al. 2014). Because the distribution of other risk factors for GC such as Helicobacter pylori infection, smoking and some other lifestyle and dietary factors was independent of the genetic variants, they were not included in models to retain maximal statistical power. The possibility of population stratification was considered and excluded as previously detailed (Sala et al. 2012). Association analyses and permutation tests were performed with SNPassoc R package (González et al. 2007).

miRNA cloning and transfection experiments in HeLa cells

A genomic region of 521 bp including the 83 bp of the precursor molecule of hsa-miR-938 and at least 100 bp at the 5' and 100 bp at the 3' flanking regions were amplified from human DNA samples

using the forward primer: 5' cacacacaAGATCTCCAAATCATTCTGGCAGTGA3' and reverse primer: 5'-cacacacaCTGCAGTTCATTGCTTGTTGGGATCA-3'. PCR fragments were cloned into the pmR-ZsGreen1 vector (Clontech) through PsTI and BglII restriction sites. Expression experiments by RT-qPCR were performed in HeLa cells as previously described (Lopez-Valenzuela et al. 2012). Briefly, after 24h of growth cells were double-transfected using Lipofectamine 2000 (Invitrogen) with 0.75 ng of DNA of the plasmid constructions carrying either one or the other allele variant of the studied miRNAs plus a control reference miRNA. Transcriptome experiments were performed in HeLa cells grown at 2x105 cells/well in six-well plates and transfected after 24h with the above-described constructs or a related negative miRNA control.

In all approaches, three independent experiments were performed including three technical replicates in each one. Transfection was stopped at 24 h for RT-qPCR expression and 48h for transcriptome analysis, by adding 1 mL QIAzol Lysis Reagent (QIAGEN) and total RNA was extracted using miRNeasy Mini Kit (QIAGEN).

Expression analysis by real time quantitative reverse transcription polymerase chain reaction (RT-qPCR)

MiRNA expression analyses were performed by RT-qPCR starting from 300 ng of total RNA from transfected cells by using the specific miR-938 forward primer: 5'-CAGTGCCCTTAAAGGTGA-3' and miR-938 reverse primer 5'-

CAGTTTTTTTTTTTTTTTACTGGGT-3' Primers design and RT-qPCR protocol were performed as previously described (Balcells et al. 2011; Busk 2014) using LightCycler® 480 SYBR Green I Master (Roche Diagnostics) following manufacturer's protocol. Standard curves were calculated by pooling samples from transfection experiments. Expression level comparisons between the two variants of miR-938 were calculated from RT templates 30x diluted using a modification of reported method (Pfaffl 2001) by quantifying the relative expression ratio of the studied miRNA based on the efficiency and the Cp deviation (ΔCp) between the two variants, and comparing vs. a reference miRNA gene as previously described (Gallego et al. 2016).

Folding analysis of the miRNA variants

Secondary hairpin structures and minimum free energy from precursor sequence carrying the studied allele variants were calculated using Mfold algorithm (http://mfold.rna.albany.edu/?q=mfold).

Whole-Genome Expression Analysis in HeLa Cells usig Beadchip Microarrays

RNA samples from three independent transfection experiments in HeLa cells were used for microarray expression experiments in Agilent SurePrint G3 Human Gene Expression microarrays (8x60k) starting from 300ng of total RNA. Data were analyzed using the Array File Maker (AFM) 4.0 software package (Array File Maker, Stanford, California). We considered genes deregulated when comparing each miRNA allele variant vs. the empty construct (nominal p-value < 0.05 with at least 1.2 fold change (FC)). We considered differential regulation between alleles comparing each miRNA allele variant vs. the other (nominal p-value < 0.05 with at least 1.2 FC). Analyses on the biological functions associated with deregulated genes were performed using the Ingenuity Pathway Analysis software (IPA, www.ingenuity.com) and DAVIDGO (http://david.abcc.ncifcrf.gov/home.jsp).

Whole-Genome Expression Analyses in HeLa Cells using Beadchip Microarrays

RNA samples from three independent transfection experiments in HeLa cells were used for microarray expression experiments in Agilent SurePrint G3 Human Gene Expression microarrays (8x60k) starting from 300ng of total RNA. Data were analyzed using the Array File Maker (AFM) 4.0 software package (Array File Maker,

Stanford, California). We considered genes deregulated when comparing each miRNA allele variant vs. the empty construct (nominal p-value < 0.05 with at least 1.2 fold change (FC)). We considered differential regulation between alleles comparing each miRNA allele variant vs. the other (nominal p-value < 0.05 with at least 1.2 FC). Analyses on the biological functions associated with deregulated genes were performed using the Ingenuity Pathway Analysis software (IPA, www.ingenuity.com) and DAVIDGO (http://david.abcc.ncifcrf.gov/home.jsp).

3'UTR cloning and luciferase activity assay in HeLa cells

A fragment of 1106 bp of the 3' UTRs of CXCL12 was amplified by PCR using Forward primer 5'-acacacgctagcTGTGTTACCTGAAAACACTGTGC-3' and Reverse primer 5'-acacactctagaAAGGGACAATTTTTGTTGATGG-3' s and cloned into PGL4.13 Promega vector after digestion with XbaI enzyme. Cotransfection experiments have been previously described (Lopez-Valenzuela et al. 2012). Briefly, HeLa cells were seeded at 1.3 x104 cells per well in 96-well plates and co-transfected 24 h later with the Firefly reporter constructs described above or the empty pGL4.13 vector (24 ng), the Renilla reporter plasmid pGL4.75 (3 ng), and 10 nM miRNA mimic for the mature miRNAs containing the studied allele variants (miR-938T and miR-938C) and negative control C2 (miRIDIAN; Dharmacon,

Lafayette, CO) using Lipofectamine 2000 (Invitrogen). Activity of Firefly and Renilla luciferases was measured 24h after transfection using Dual-Glo Luciferase assay system (Promega).

Relative reporter firefly luciferase activity was obtained by normalization to the Renilla luciferase activity. In order to correct for vector-dependent unspecific effects, each relative reporter activity was normalized to the empty vector co-transfected with the corresponding miRNA. Results were then compared with the mean of the C2 negative control. Each experiment was done in triplicate, and at least three independent experiments were performed for each miRNA. Statistical significance was determined using Student's t-test ($P < 0.05$).

RESULTS

Association of miRNA SNPs with gastric cancer and its histological and anatomical subtypes

To analyse the possible association between gastric adenocarcinoma and putative functional SNPs located in miRNAs we first we selected a panel of 29 non-monomorphic SNPs located in different regions of 30 human miRNAs (Suppplementary Table 1), eight of the SNPs were discarded from the analysis because of genotyping technical problems and, therefore, the final analysis was performed with a final set of 21 SNPs. Fifteen of the SNPs were located in the precursor out of the mature miRNA region, two in the mature out of the seed miRNA region, three in the seed miRNA region and one SNP, rs3746444:A>G, was located in the seed of miR-499a-3p and in the mature miR-499b-5p since both miRNAs are located in opposite strands of the same genomic region (Table 2).

Table 1. Main characteristics of the gastric cancer cases and controls analyzed.

		Cases		Controls	
		N	(%)	N	(%)
Sex	Male	214	(58.6%)	759	(59.1%)
	Female	151	(41.4%)	525	(40.9%)
Anatomical subtype of GC	Cardia	107	(29.3%)	-	-
	Noncardia	181	(49.6%)	-	-
	Mixed	6	(1.6%)	-	-
	Unknown	71	(19.5%)	-	-
				-	-
Histological subtype of GC	Intestinal	126	(34.5%)	-	-
	Diffuse	128	(35.1%)	-	-
	Mixed	8	(2.2%)	-	-
	Unknown	103	(28.2%)	-	-
		Mean	(Sd)	Mean	(Sd)
Age at recruitment (years)		58.4	(7.9)	58.4	(7.69)

Table 2. Main characteristics of the miRNA SNPs analyzed

dbSNP	miRNA gene	Chr.	MAF Controls	MAF cases	Location	Alleles	Minor allel	Global Fst
rs41291179	*MIR216A*	2	0.053	0.047	precursor	A/T	T	0.1068
rs13186787	*MIR1294*	5	0.002	0.003	precursor	A/G	G	0.0188
rs2910164	*MIR146A*	5	0.24	0.229	seed	C/G	G	0.1012
rs41274239	*MIR96*	7	0.004	0.003	precursor	A/G	G	0.001
rs2114358	*MIR1206*	8	0.395	0.358	precursor	T/C	C	0.0302
rs10505168	*MIR2053*	8	0.28	0.274	precursor	A/G	G	0.0768
rs11259096	*MIR1265*	10	0.048	0.054	precursor	T/C	C	0.2416
rs17091403	*MIR2110*	10	0.102	0.084	precursor	C/T	T	0.031
rs4919510	*MIR608*	10	0.191	0.178	mature	C/G	G	0.1406
rs12416605	*MIR938*	10	0.258	0.248	seed	C/T	T	0.2635
rs11020790	*MIR548L*	11	0.013	0.007	precursor	C/T	T	0.1503
rs7311975	*MIR1178*	12	0.028	0.026	seed	T/C	C	0.3701
rs11614913	*MIR196A2*	12	0.394	0.41	mature	C/T	T	0.2026
rs2289030	*MIR492*	12	0.058	0.047	precursor	C/G	G	0.1107
rs11844707	*MIR1185-2*	14	0.001	0	precursor	G/A	A	0.2655
rs6505162	*MIR423*	17	0.461	0.452	precursor	A/C	C	0.375
rs17759989	*MIR633*	17	0.029	0.022	precursor	A/G	G	0.041
rs895819	*MIR27A*	19	0.32	0.31	precursor	T/C	C	0.0412
rs3746444	*MIR499A*	20	0.192	0.185	seed	A/G	G	0.000
rs3746444	*MIR499B*	20	0.192	0.185	mature	A/G	G	0.000
rs4822739	*MIR548J*	22	0.051	0.049	precursor	C/G	G	0.0613
rs5965660	*MIR888*	X	0.187	0.218	precursor	T/G	G	NA

Next we investigated if miRNA SNPs were significantly associated with different GC subtypes in 1284 controls matched to 365 incident GC cases with different histological and tumor location phenotypes (European Prospective Investigation into Cancer and Nutrition (EPIC) cohort (Table 1). Results of the association analyses under the log-additive model are shown in Table 3. The

strongest association was found for rs2114358:T>C in the precursor miR-1206, which appeared inversely associated with the risk of the non-cardias localization of the adenocarcinoma (per allele OR=0.73, 95%CI=0.58-0.93, p-value=0.009) but not for the cardias localization. Conversely, rs3746444:A>G in the mature miR-499a-3p and seed of miR-499b-5p, appeared inversely associated only with the risk of the cardias localization of the adenocarcinoma (per allele OR=0.64, 95%CI=0.42-0.98, p-value=0.0308). As for the histological subtype, rs12416605:C>T in the seed of miR-938, appeared inversely associated with the risk for the diffuse subtype of GC (per allele OR=0.7, 95%CI=0.52-0.97, p-value=0.028) but not for the intestinal one (Table 3). These results are suggestive of a protective effect for GC susceptibility for the minor alleles of all three SNPs (C rs2114358, G rs3746444 and T rs2114358).

Analyses by other inheritance models (codominant, dominant and recessive) for the three SNPs that associated with GC risk under the log-additive model indicated that the unique association that was stronger with other models was this of rs3746444:A>G with the cardias phenotype, which was better explained by the dominant model (OR=0.58, 95%CI=0.36-0.94, p-value=0,021).

Table 3. Significant associations of miRNA SNPs with gastric cancer and its histological and anatomical subtypes (log-additive model)

		GASTRIC ADENOCARCINOMA		INTESTINAL		DIFFUSE		CARDIAS		NO CARDIAS	
dbSNP	**miRNA gene**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
rs41291179	*MIR216A*	0.87 (0.59-1.28)	0.4657	1.03 (0.58-1.84)	0.9165	0.96 (0.53-1.72)	0.8786	0.69 (0.33-1.45)	0.3036	0.98 (0.6-1.61)	0.9343
rs13186787	*MIR1294*	1.76 (0.32-9.81)	0.53087	0 (0-NA)	0.4122	5.38 (0.93-31.27)	0.0937	0 (0-NA)	0.3872	4.16 (0.73-23.89)	0.1452
rs2910164	*MIR146A*	0.95 (0.78-1.15)	0.593	1.11 (0.82-1.49)	0.4988	0.8 (0.58-1.11)	0.1723	1.18 (0.85-1.64)	0.3135	0.93 (0.71-1.21)	0.5839
rs41274239	*MIR96*	0.66 (0.14-3.01)	0.572	0 (0-NA)	0.1467	0.83 (0.1-6.67)	0.8590	0 (0-NA)	0.2888	0.55 (0.07-4.35)	0.5363
rs2114358	***MIR1206***	0.85 (0.71-1.01)	0.0600	0.8 (0.61-1.06)	0.1145	0.81 (0.62-1.06)	0.1267	1.04 (0.77-1.4)	0.8092	**0.73 (0.58-0.93)**	**0.0093**
rs10505168	*MIR2053*	0.97 (0.81-1.17)	0.7663	0.93 (0.7-1.24)	0.6348	0.88 (0.66-1.18)	0.3960	0.83 (0.6-1.14)	0.2486	1.07 (0.84-1.37)	0.5666
rs11259096	*MIR1265*	1.11 (0.77-1.61)	0.5721	1.36 (0.8-2.33)	0.2728	0.87 (0.45-1.69)	0.6764	1.48 (0.83-2.62)	0.1975	0.99 (0.58-1.69)	0.9816
rs17091403	*MIR2110*	0.81 (0.6-1.09)	0.1505	0.79 (0.49-1.27)	0.3074	0.87 (0.55-1.37)	0.5374	0.67 (0.39-1.16)	0.1345	0.85 (0.57-1.26)	0.4017
rs4919510	*MIR608*	0.92 (0.74-1.15)	0.4724	1.15 (0.83-1.58)	0.4140	0.81 (0.56-1.16)	0.2426	1.18 (0.83-1.69)	0.3594	0.87 (0.64-1.17)	0.3449
rs12416605	***MIR938***	0.95 (0.78-1.15)	0.5733	0.96 (0.71-1.31)	0.8152	**0.7 (0.51-0.97)**	**0.0281**	0.88 (0.63-1.24)	0.4604	1.07 (0.83-1.38)	0.5896
rs11020790	*MIR548L*	0.55 (0.21-1.42)	0.1846	0.59 (0.14-2.53)	0.4467	0.28 (0.04-2.09)	0.1300	0 (0-NA)	0.0493	0.6 (0.18-2)	0.3731
rs7311975	*MIR1178*	0.95 (0.56-1.6)	0.8504	1.33 (0.65-2.75)	0.4514	0.86 (0.36-2.03)	0.7197	1.2 (0.53-2.72)	0.6743	0.68 (0.3-1.5)	0.3144
rs11614913	*MIR196A2*	1.07 (0.9-1.27)	0.4543	1.12 (0.85-1.46)	0.4189	1.21 (0.92-1.59)	0.1649	0.9 (0.67-1.21)	0.4830	1.18 (0.93-1.48)	0.1702
rs2289030	*MIR492*	0.79 (0.54-1.16)	0.2166	0.6 (0.3-1.19)	0.1192	0.98 (0.56-1.71)	0.9371	0.62 (0.3-1.29)	0.1739	1.11 (0.7-1.76)	0.6630
rs11844707	*MIR1185-2*	0 (0-NA)	0.3216	0 (0-NA)	0.4983	0 (0-NA)	0.6115	0 (0-NA)	0.4749	0 (0-NA)	0.5279
rs6505162	*MIR423*	0.96 (0.8-1.15)	0.6746	0.81 (0.6-1.09)	0.1627	1.06 (0.8-1.42)	0.6662	1.04 (0.76-1.42)	0.8249	0.88 (0.69-1.13)	0.3195
rs17759989	*MIR633*	0.71 (0.41-1.25)	0.2212	0.28 (0.07-1.16)	0.0305*	0.77 (0.33-1.8)	0.5299	1.07 (0.45-2.54)	0.8757	0.46 (0.18-1.15)	0.0632
rs895819	*MIR27A*	0.96 (0.8-1.16)	0.6801	1.08 (0.81-1.45)	0.5919	0.92 (0.69-1.25)	0.6048	1.02 (0.74-1.41)	0.8912	0.91 (0.7-1.17)	0.4577
rs3746444	***MIR499A/B***	0.95 (0.77-1.18)	0.6560	0.95 (0.68-1.32)	0.7490	1.07 (0.77-1.48)	0.6977	**0.64 (0.42-0.98)**	**0.0308**	1.1 (0.83-1.46)	0.5004
rs4822739	*MIR548J*	0.96 (0.65-1.41)	0.8297	1.57 (0.93-2.65)	0.1063	0.81 (0.43-1.53)	0.5027	1.3 (0.7-2.4)	0.4172	0.77 (0.43-1.35)	0.3401
rs5965660	*MIR888*	1.12 (0.96-1.32)	0.1625	1.13 (0.88-1.45)	0.3360	1.02 (0.77-1.35)	0.8729	1.16 (0.89-1.51)	0.2716	1.24 (1-1.54)	0.0524

* In this case, the confidence interval does not support the genetic association.

Significant associations are shown in bold.

rs12416605:C>T affects the expression of miR-938 and several gene regulatory pathways

rs2114358 in miR-1206, and rs3746444:A>G in miR-499a-3p and miR-499b-5p, have already been shown to affect the dosage of the corresponding miRNAs by altering their biogenesis or stability (Kim et al. 2012b,Torruella-Loran et al., 2016). Conversely, no studies have been conducted to analyse the functional effect of rs12416605:C>T. We therefore investigated if there were differences in miRNA expression or in gene regulation between both miR-938 alleles of rs12416605:C>T that could be underlying the genetic association with GC.

To test the effect of rs12416605:C>T in miR-938 expression, we cloned both miRNA allelic variants, over-expressed them in HeLa cells (after confirming that these cells had undetectable levels of endogenous miR-938) and measured their expression by RT-qPCR. As shown in Figure 1a, the rs12416605 C allele of miR-938 was 1.49 times more expressed than the T allele (Student t-test, p-value < 0.05) after correction by control miRNA and an empty construct (Student t-test, p-value < 0.05) indicating that this variant in miR-938 may be affecting the final expression of the miRNA.

We further analysed the secondary structure of the hsa-miR-938 molecule by predicting the folding structure for both rs12416605 allele variants using the RNA fold algorithm. However, no differences were observed, nor in the hairpin structure neither in the minimum free energy between both allele variants.

In addition to differences in the expression levels between the rs12416605 C and T allele variants, this nucleotide substitution could also change the spectrum of miRNA target genes, thus we investigated which genes could be regulated by each miRNA allele variant through miRNA over-expression experiments and transcriptome analyses. After transfection of both miRNA allele variants, we selected those trascripts statistically significant up or down regulated with respect to the control miRNA (p-value < 0.05, over a 1.2 FC). The miR-938 C allele variant deregulated 2045 trascripts, whereas the T allele variant deregulated 1825 trascripts (Figure 1b). Among those, 829 where deregulated by both variants (40.5% of deregulated trascripts by the C allele and 45.4% by the T allele) indicating that about 50% of the transcripts were exclusively regulated by each of the rs12416605:C>T allele variants. Next we used Ingenuity Pathway Analysis software to analyse the functions associated with the genes deregulated after over-expression of each miR-938 allele variants. Both variants presented Cancer as their top disease associated with their deregulated genes, and Endocrine System Disorder was also one of the diseases in common. Interestingly, variant C but not variant T presented Gastrointestinal Disease among the top five diseases to which their deregulated genes are associated (data not shown).

According to cancer as the top associated disease, there were several genes that are implicated in the aetiology of cancer among the top 10 up-regulated and down-regulated genes such as the Fibroblast Growth Factor Receptor 1 (FGFR1) or the NK2 Homeobox 8 gene (NKX2-8) (Table 4). Remarkably, most of the

top ten down regulated transcripts but none of the top ten up regulated transcripts were shared between the two rs12416605:C>T allele variants (Table 4).

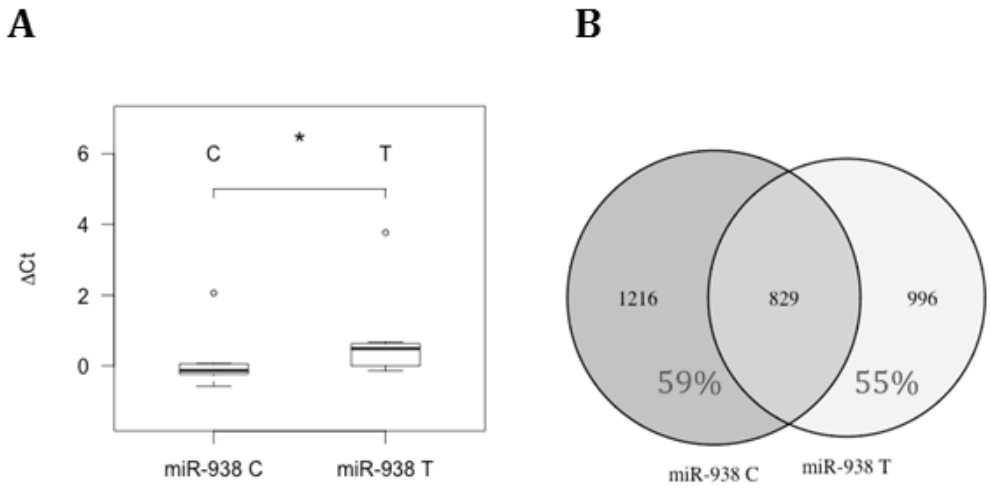

Figure 1: Results of over-expression experiments of rs12416605:C>T miR-938 in HeLa cells. a) Differences in the expression levels ((Cp) measured by qPCR in transfected HeLa cells for the C and T miR-938 allele variants related to a control reference miRNA. Asterisks indicate $p < 0.05$ in the t-test comparisons; b) Venn diagram of deregulated transcripts in the transcriptome experiment for the C and T miR-938 allele variants related to a control reference miRNA Absolut numbers of deregulated transcripts for each category are shown, as well as percentage of exclusively deregulated transcripts for each allele variants.

Table 4: Top ten deregulated genes by miR-938 rs12416605:C>T alleles with the highest fold change compared with the control. In grey deregulated transcripts in common between both alleles.

	miR-938 C	**miR-938 T**
Top 10 up-regulated genes	*TSPAN19*	*HPGDS*
	C18orf34	*TRPA1*
	SLC4A5	*SORBS1*
	CHD6	*PEX5L-AS2*
	BC015129	*OR1D5*
	IL34	*SLC23A2*
	OR5B2	*FLJ21408*
	GRIP1	*PRR15L*
	SRRM4	*RBMS3*
	ABCA8	*RNF157*
Top 10 down-regulated genes	*NKX2-8*	*NKX2-8*
	ZNF730	*ZNF730*
	THPO	*THPO*
	MFSD4	*DNAJC3*
	PTPN11	*HNF1A*
	DPCR1	*MFSD4*
	HNF1A	*FGF1*
	MDM1	*PTPN11*
	C1orf210	*DPCR1*
	CCDC103	C1orf210

The chemokine CXCL12 gene shows differential allele regulation by miR-938

To investigate which of the transcripts deregulated in the microarray experiment could be directly regulated by any rs12416605:C>T allele variant of miR-938, we first used PITA algorithm to predict targets among deregulated genes for each miR-938 allele variant. When looking at the list of candidate genes to be exclusively regulated by one of the miR-938 allele variants (Table 5), the chemokine CXCL12 presented the higher difference in the FC betwenn alleles (4.27). In addition, the C allele variant showed six predicted target sites and a best pita score of -9.08 whereas the T allele did not show any good prediction (Figure 2a). Therefore we focused our attention in CXCL12 as it is known to be involved in GC. To investigate the interaction between CXCL12 with the rs12416605 miR-938 C and the T alleles, functional validation was performed using a dual-luciferase assay in HeLa cells. A luciferease reporter pGL4.13 construct, either empty or carrying the 3'UTR of the CXCL12, was co-transfected with the corresponding miRNA mimics: miR-938T, miR-938C or a control miRNA. As shown in Figure 2b, a statistically significant reduction of the luciferase activity was observed between the rs12416605 C allele variant of miR-938 and both the control miRNA and the T allele variant when co-transfected with 3' UTR of the CXCL12 gene after correcting with the empty construct ($p<0.05$, Student's t test). The observed reduction of the luciferase activity would be compatible with a 25%

repression of CXCL12 exclusively by the rs12416605 C allele of miR-938 suggesting that differences in the regulation of CXCL12 between both rs12416605 alleles might be a factor underlying the genetic association of miR-938 with GC.

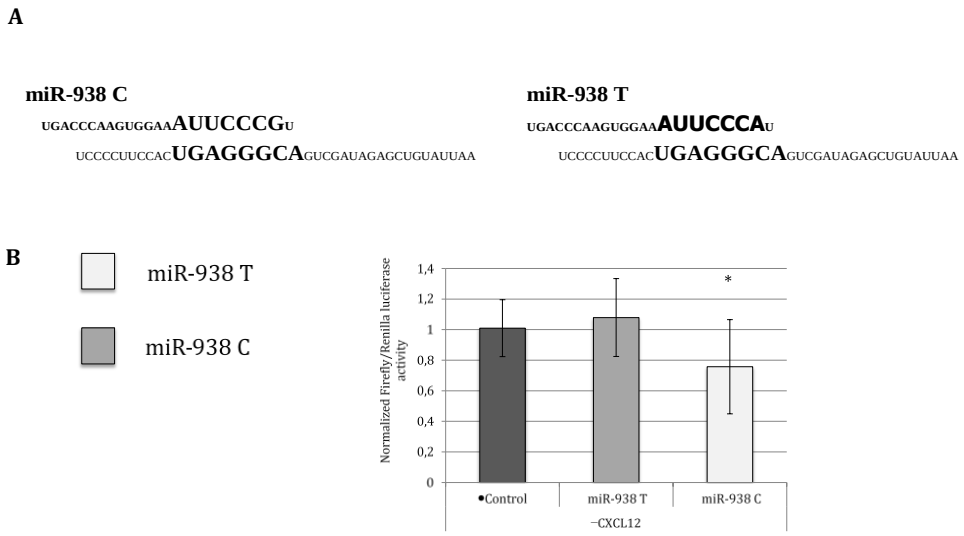

Figure 2: **A:** *CXCL12* mRNA with PITA predicted target sites marked in Green (miR-938 T allelic variant) and Yelow (miR-938 C allelic variant), primer sequences marquet in Blue and rs1801157 marqued in Red **B:** Results of the luciferase-reporter assay, which was used to test the interaction between miR-938-3p alleles and the 3'UTR of the CXCL12 gene in HeLa cells. Ratios of the Firefly and Renilla luciferase luminescence are presented after normalization to the empty plasmid pGL4.13 and to a miRNA control. Each experiment was done in triplicate and eight independent experiments were performed for each miRNA tested, at a concentration of 10 nM. Data reported here are the means+/-SEM of all experiments performed. Significant reduction of the luciferase activity ($P<0.01$, Student's t-test) is indicated with an asterisk.

Table 5: Gastric cancer related genes found to be differentially regulated by miR-938 rs12416605:C>T alleles variants accorindg to the transcriptome experiments and PITA target site predictions. The best Pita scores for a transcript of each gene, the total number of Pita target sites and the fold change, both *versus* a control miRNA and between alleles (C *versus* T), are shown in the table.

Gene Symbol	Gene name	Allele	Best Pita Score	Total Pita Sites	Control FC	C *versus* T FC
CXCL12	Chemokine C-X-C motif ligand 12	C	-9.08	6	-1.54	4.27
		T	-0.93	2	-1.58	
CCR5	C-C chemokine receptor 5	C	-5.79	2	1.63	1.27
		T	-12.81	3	2.05	
CFLAR	CASP8 And FADD-Like Apoptosis Regulator	C	-15.51	12	-	-1.07
		T	-11.4	15	1.39	
LEPR	Leptin receptor	C	-4.77	6	-	-1.26
		T	-11.81	6	-1.21	

DISCUSSION

Even though changes in the expression of miRNAs have been largely associated with GC and impaired miRNA expression may be related to genetic variation in their sequences, still there is lack of evidences explaining the real contribution of miRNA genetic variants to GC pathology. In the present study we interrogated several putative functional allele variants in miRNAs as candidate factors for GC by mean of association studies and demonstrated a possible causative effect by mean of a functional approach for at least one of these variants.

We identified three common genetic variants in miRNA genes that associated with GC, one of them, rs3746444:A>G, has been reported to be associated with cancer in several reports that further suggest a role for this SNP in the aetiology GC (Ahn et al. 2013; Chen et al. 2014). We identied the G allele as a protective factor for GC under the dominant model, which is partially in agreement with a recent meta-analysis study in which this SNP was reported to be associated with GC in Assian but not European populations (Chen et al. 2014). Although the study concludes that no association of rs3746444:A>G with GC in Europeans, other studies support our finding of association of this SNP with other cancer types in populations of European origin (Nikolić et al. 2015). Furthermore, we have recently shown that the two alleles of rs3746444:A>G, both in miR-499a-3p and miR-499b-5p, may affect the processing and expression of these miRNAs, which ultimately may be affecting

the regulation of their target genes in a dosage and allele-dependent manner (Torruella-Loran et al., 2016). More importantly, we showed that miR-499a-3p could be regulating the cadherin CDH1 and the cell adhesion molecule CLH1 in an allele-dependent fashion with the A allele repressing both genes. The cadherin down-regulation caused by the miR-499a-3p A allele could be associated with a malignant behaviour (Behrens et al. 1989), and could underlie the protective effect of the G allele.

Another of the SNPs that we found associated with GC was rs2114358:T>C in the precursor molecule of miR-1206. This is the first report of a genetic association of this miRNA with a disease, besides a previous report that suggested a role for this miRNA in cancer and further demonstates an effect of the SNP in the biogenesis of the mature miRNA forms (Kim et al. 2012b). Thus, and because the SNP is located in the miRNA precursor regeion outside the mature miRNA molecule, we estimate that the effect of this variant might be throught the variation of the miRNA expression rather than the alterion of the regulation of specific targeted genes. Interestingly several genes involved in gastric cancer such us FGF2 and TP53INP1 appear as validated targets of miR-1206 in mirtarbase through next generation sequencing. Future studies should analyse these interactions and identify new targets of miR-1206 that could lead to discern a posible role for this miRNA in carcinogenesis.

Finnally, the third SNP identified to be associated with GC was rs12416605:C>T in the seed of miR-938, this is the first report that

associates this SNP with cancer. Interestingly, a previous study has reported association of GC with another SNP (rs2505901) in the miR-938 primary precursor molecule as well as another variant (rs2275913) in IL17A, a miR-938 targeted gene (Arisawa et al. 2012). To investigate the possible contribution of was rs12416605:C>T to GC, we performed over-expression experiments in HeLa cells and analyzed the effect of both allele variants in the expression levels of miR-938. Remarkably, statistically significant differences were observed between the expression levels of each miRNA variant with the C variant of miR-938 being 1.45 times more expressed than the T variant. However these results could not be explained by the stability of the secondary structure; other facts such as the interaction between the miR-938 variants miRNA and miRNA sponges like long non-coding or circular RNAs could be affecting the detection of the miRNAs (Bak and Mikkelsen 2014). Besides the alteration on the expression levels of miR-938, the studied allele variant was also shown to have an effect in the regulatory networks in which this miRNA is involved. Interestingly among deregulated genes in the microarray there were several genes that might be strongly regulated by miR-938 according to PITA predictions (Table 5). For instance the CASP8 and FADD-Like Apoptosis Regulator (CFLAR) gene has strong Pita scores and high number of predicted sites for both allelic variants of miR-938. CEFLAR is a gene that encodes the FLICE-like inhibitory protein (c-FLIP) that acts as a inhibitor of the Fas-mediated apoptosis (Olsson and Zhivotovsky 2011) and helps tumour cells to escape from TRAIL mediated apoptosis, hence promoting metastasis and

tumour progression (Zhang et al. 2004; Zhou et al. 2004). Interestingly, several SNPs in the CFLAR gene have been associated with GC in Chinese populations (Hyland et al. 2014). Moreover CFLAR mRNA and its protein were found highly expressed in gastric adenocarcinomas compared with normal gastric mucosa tissue in patients of Chinese origin (Zhou et al. 2004). In contrast to CFLAR, which might be regulated by both allele variants of miR-938, we found several genes that were predicted as possible exclusive targets of one of the miR-938 variants and that are known to play important roles in GC (Table 5). For instance the C-C chemokine receptor 5 (CCR5) and the Leptin receptor (LEPR), are related to GC and are predicted to be regulated exclusively by the T allele variant. LEPR was found deregulated in the transcriptome only after over-expression of the T allele variant (Chang et al. 2014; Shi et al. 2014). Moreover, Pita predictions also suggest a strong regulation of LEPR by the miR-938 T allele with a top score of -11.81 and 6 predicted target sites. Leptin and its receptor are known to play a role in H. pillory infection (Azuma et al. 2001), and a polymorphism in LEPR has been associated with GC risk in Korean populations (Kim et al. 2012a). Similarly, CCR5 was predicted as a target exclusively for the T miR-938 allele with a top score of -12.81 and three predicted target sites. In this context, the expression of CCR5 has been found to favor metastasis in GC patients (Cao et al. 2011; Mencarelli et al. 2013). Finnally, the chemokine CXCL12 gene was predicted to be regulated only by the C miR-938 allele variant with six predicted target sites and a best pita score of -9.08. In fact CXCL12 is well known to be involved in

GC and the expression of this gene has been related to metastasis. Although it seems clear that CXCL12 has a critical role in GC, several authors have reported that a lack of CXCL12 expression might lead to GC metastasis (Zhi et al. 2012; Rubie et al. 2015) whereas some others suggest that CXCL12 over-expression promotes cell invasion (Izumi et al. 2016). The here reported SNP might be causing this lack of agreement and thus further investigation is required. However, this case is of special interest, because despite of its moderate target prediction, CXCL12 regulation by miR-938, showed strong differences between alleles both in the microarray fold change and the Pita predictions. For this reason we further analyzed the interaction between both allelic variants of miR-938 and the 3'UTR of CXCL12 through a luciferase reporter assay and showed a possible repression of the gene by the C but not the T allele, being the last a protective factor according to our results. All together would suggest that a reduction of the expression of miR-938 in carriers of the T rs12416605 allele might lead to a lack of repression of the CXCL12 that, by means of an still unrevealed mechanism, could be protective for GC.

In conclusion we have replicated a previously described association of miR-499 with GC in European populations and identified two new unreported associations of miR-1206 and miR-938 with GC. In all three cases a reduction in the expression of the miRNA was related to one of the alleles and, furthermore, we have shown that miR-938 could be regulating the chemokine CXCL12 in an allele-dependent fashion.

REFERENCES:

Ahn DH, Rah H, Choi Y-K, Jeon YJ, Min KT, Kwack K, Hong SP, Hwang SG, Kim NK. 2013. Association of the miR-146aC>G, miR-149T>C, miR-196a2T>C, and miR-499A>G polymorphisms with gastric cancer risk and survival in the Korean population. Mol. Carcinog. 52 Suppl 1: E39–51.

Arisawa T, Tahara T, Shiroeda H, Matsue Y, Minato T, Nomura T, Yamada H, Hayashi R, Saito T, Matsunaga K, Fukuyama T, Hayashi N, et al. 2012. Genetic polymorphisms of IL17A and pri-microRNA-938, targeting IL17A 3'-UTR, influence susceptibility to gastric cancer. Hum. Immunol. 73: 747–752.

Azuma T, Suto H, Ito Y, Ohtani M, Dojo M, Kuriyama M, Kato T. 2001. Gastric leptin and Helicobacter pylori infection. Gut 49: 324–329.

Bak RO, Mikkelsen JG iehm. 2014. miRNA sponges: soaking up miRNAs for regulation of gene expression. Wiley interdisciplinary reviews RNA 5, 3:317-333.

Balcells I, Cirera S, Busk PK. 2011. Specific and sensitive quantitative RT-PCR of miRNAs with DNA primers. BMC Biotechnol. 11: 70.

Bartel DP. 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell. 116 2:281-297.

Behrens J, Mareel MM, Roy FM Van, Birchmeier W. 1989. Dissecting tumor cell invasion: Epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion. J. Cell Biol. 108: 2435–2447.

Busk PK. 2014. A tool for design of primers for microRNA-specific quantitative RT-qPCR. BMC Bioinformatics 15: 29.

Cao Z, Xu X, Luo X, Li L, Huang B, Li X, Tao D, Hu J, Gong J. 2011. Role of RANTES and its receptor in gastric cancer

metastasis. J. Huazhong Univ. Sci. Technol. - Med. Sci. 31: 342–347.

Carneiro F, Moutinho C. 2007. Pathology findings and validation of gastric and esophageal cancer cases in a European cohort (EPIC/EUR-GAST). 42 5:618-627.

Chang W-J, Du Y, Zhao X, Ma L-Y, Cao G-W. 2014. Inflammation-related factors predicting prognosis of gastric cancer. World J. Gastroenterol. 20: 4586–4596.

Chen C, Yang S, Chaugai S, Wang Y, Wang DW. 2014. Meta-analysis of Hsa-mir-499 polymorphism (rs3746444) for cancer risk: evidence from 31 case-control studies. BMC Med. Genet. 15: 1–11.

Correa P. 1992. Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. Cancer Res. 52: 6735–40.

Espinosa-Parrilla Y, Muñoz X, Bonet C, Garcia N, Venceslá A, Yiannakouris N, Naccarati A, Sieri S, Panico S, Huerta JM, Barricarte A, Menéndez V, et al. 2014. Genetic association of gastric cancer with miRNA clusters including the cancer-related genes MIR29, MIR25, MIR93 and MIR106: Results from the EPIC-EURGAST study. Int. J. Cancer 135: 2065–2076.

Esquela-Kerscher A, Slack FJ. 2006. Oncomirs - microRNAs with a role in cancer. Nat. Rev. Cancer 6: 259–269.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2015. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer 136: E359–E386.

Filipowicz W, Bhattacharyya SN, Sonenberg N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9: 102–14.

Fletcher O, Houlston RS. 2010. Architecture of inherited susceptibility to common cancer. Nat. Rev. Cancer 10: 353–361.

Friedman RC, Farh KKH, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19: 92–105.

Gallego A, Melé M, Balcells I, García-Ramallo E, Torruella-Loran I, Fernández-Bellon H, Abelló T, Kondova I, Bontrop R, Hvilsom C, Navarro A, Marquès-Bonet T, et al. 2016. Functional Implications of Human-Specific Changes in Great Ape microRNAs. PLoS One 11: e0154194.

Genta RM. 1997. The immunobiology of Helicobacter pylori gastritis. Semin. Gastrointest. Dis. 8: 2–11.

González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, Moreno V. 2007. SNPassoc: An R package to perform whole genome association studies. Bioinformatics 23: 644–645.

Houghton J, Wang TC. 2005. Helicobacter pylori and gastric cancer: A new paradigm for inflammation-associated epithelial cancers. Gastroenterology. 128 6:1567-1578.

Hyland PL, Lin S-W, Hu N, Zhang H, Wang L, Su H, Wang C, Ding T, Tang Z-Z, Fan J-H, Qiao Y-L, Xiong X, et al. 2014. Genetic variants in fas signaling pathway genes and risk of gastric cancer. Int. J. Cancer 134: 822–31.

Izumi D, Ishimoto T, Miyake K, Sugihara H, Eto K, Sawayama H, Yasuda T, Kiyozumi Y, Kaida T, Kurashige J, Imamura Y, Hiyoshi Y, et al. 2016. CXCL12/CXCR4 activation by cancer-associated fibroblasts promotes integrin β1 clustering and invasiveness in gastric cancer. Int. J. Cancer 138: 1207–1219.

Katada T, Ishiguro H, Kuwabara Y, Kimura M, Mitul A, Mori Y, Ogawa R, Harata K, Fujii Y. 2009. MicroRNA expression profile in undifferentiated gastric cancer. Int. J. Oncol. 34: 537–542.

Kim EY, Chin HM, Park SM, Jeon HM, Chung WC, Paik CN, Jun KH. 2012a. Susceptibility of gastric cancer according to leptin and leptin receptor gene polymorphisms in Korea. J. Korean Surg. Soc. 83: 7–13.

Kim HK, Prokunina-Olsson L, Chanock SJ. 2012b. Common Genetic Variants in miR-1206 (8q24.2) and miR-612 (11q13.3) Affect Biogenesis of Mature miRNA Forms. PLoS One 7.:

Kim YK, Yu J, Han TS, Park SY, Namkoong B, Kim DH, Hur K, Yoo MW, Lee HJ, Yang HK, Kim VN. 2009. Functional links between clustered microRNAs: Suppression of cell-cycle inhibitors by microRNA clusters in gastric cancer. Nucleic Acids Res. 37: 1672–1681.

Kloosterman WP, Plasterk RH. 2006. The diverse functions of microRNAs in animal development and disease. Dev Cell 11: 441–450.

Krol J, Loedige I, Filipowicz W. 2010. The widespread regulation of microRNA biogenesis, function and decay. Nat. Rev. Genet. 11: 597–610.

Lao-Sirieix P, Caldas C, Fitzgerald RC. 2010. Genetic predisposition to gastro-oesophageal cancer. Current Opinion in Genetics and Development. 20: 210-217.

Lewis BP, Shih I, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian microRNA targets. Cell 115: 787–98.

Loh M, Koh KX, Yeo BH, Song CM, Chia KS, Zhu F, Yeoh KG, Hill J, Iacopetta B, Soong R. 2009. Meta-analysis of genetic polymorphisms and gastric cancer risk: Variability in associations according to race. Eur. J. Cancer 45: 2562–2568.

Lopez-Valenzuela M, Ramírez O, Rosas A, García-Vargas S, La Rasilla M De, Lalueza-Fox C, Espinosa-Parrilla Y. 2012. An ancestral miR-1304 allele present in neanderthals regulates genes involved in enamel formation and could explain dental differences with modern humans. Mol. Biol. Evol. 29: 1797–1806.

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, et al. 2005. MicroRNA expression profiles classify human cancers. Nature 435: 834–838.

Mencarelli A, Graziosi L, Renga B, Cipriani S, D'Amore C, Francisci D, Bruno A, Baldelli F, Donini A, Fiorucci S. 2013. CCR5 Antagonism by Maraviroc Reduces the Potential for Gastric Cancer Cell Dissemination. Transl. Oncol. 6: 784–93.

Nikolić Z, Savić Pavićević D, Vučić N, Cidilko S, Filipović N, Cerović S, Vukotić V, Romac S, Brajušković G. 2015. Assessment of association between genetic variants in microRNA genes hsa-miR-499, hsa-miR-196a2 and hsa-miR-27a and prostate cancer risk in Serbian population. Exp. Mol. Pathol. 99: 145–150.

Olsson M, Zhivotovsky B. 2011. Caspases and cancer. Cell Death Differ. 18: 1441–1449.

Peng S, Kuang Z, Sheng C, Zhang Y, Xu H, Cheng Q. 2010. Association of MicroRNA-196a-2 Gene Polymorphism with Gastric Cancer Risk in a Chinese Population. Dig. Dis. Sci. 55: 2288–2293.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29: e45.

Riboli E, Hunt K, Slimani N, Ferrari P, Norat T, Fahey M, Charrondière U, Hémon B, Casagrande C, Vignat J, Overvad K, Tjønneland a, et al. 2002. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 5: 1113.

Rubie C, Kauffels A, Kï¿½lsch K, Glanemann M, Justinger C. 2015. CXCL12/CXCR4 display an inverse mRNA expression profile in gastric carcinoma that correlates with tumor progression. Oncol. Lett. 360–364.

Rugge M, Fassan M, Graham DY. 2015. Epidemiology of gastric cancer. Gastric Cancer Princ. Pract. 12: 23–34.

Saeki N, Ono H, Sakamoto H, Yoshida T. 2013. Genetic factors related to gastric cancer susceptibility identified using a genome-wide association study. Cancer Science 1:1-8

Sala N, Muñoz X, Travier N, Agudo A, Duell EJ, Moreno V, Overvad K, Tjonneland A, Boutron-Ruault MC, Clavel-Chapelon F,

Canzian F, Kaaks R, et al. 2012. Prostate stem-cell antigen gene is associated with diffuse and intestinal gastric cancer in Caucasians: Results from the EPIC-EURGAST study. Int. J. Cancer 130: 2417–2427.

Shen J, Stass S a, Jiang F. 2012. MicroRNAs as Potential Biomarkers in Human Solid Tumors. Cancer Lett.

Shi H, Shu H, Huang C, Gong J, Yang Y, Liu R, Yang Y, Liu P. 2014. Association of LEPR K109R polymorphisms with cancer risk: A systematic review and pooled analysis. J. B.U.ON. 19: 847–854.

Song JH, Meltzer SJ. 2012. MicroRNAs in pathogenesis, diagnosis, and treatment of gastroesophageal cancers. Gastroenterology 143.:

Ueda T, Volinia S, Okumura H, Shimizu M, Taccioli C, Rossi S, Alder H, Liu CG, Oue N, Yasui W, Yoshida K, Sasaki H, et al. 2010. Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. Lancet Oncol 11: 136–146.

Yang Q, Jie Z, Ye S, Li Z, Han Z, Wu J, Yang C, Jiang Y. 2014. Genetic variations in miR-27a gene decrease mature miR-27a level and reduce gastric cancer susceptibility. Oncogene 33: 193–202.

Zhang R, Wang W, Li F, Zhang H, Liu J. 2014. MicroRNA-106b~25 expressions in tumor tissues and plasma of patients with gastric cancers. Med Oncol 31: 243.

Zhang X, Jin T, Yang H, Dewolf WC, Khosravi-far R, Olumi AF. 2004. Persistent c-FLIP (L) Expression Is Necessary and Sufficient to Maintain Resistance to Tumor Necrosis Factor-Related Apoptosis-Inducing Ligand − Mediated Apoptosis in Prostate Cancer Persistent c-FLIP (L) Expression Is Necessary and Sufficient to Main. Cancer rRsearch 64: 7086–7091.

Zhi Y, Chen J, Zhang S, Chang X, Ma J, Dai D. 2012. Down-regulation of CXCL12 by DNA hypermethylation and its

involvement in gastric cancer metastatic progression. Dig. Dis. Sci. 57: 650–659.

Zhou X-D, Yu J-P, Liu J, Luo H-S, Chen H-X, Yu H-G. 2004. Overexpression of cellular FLICE-inhibitory protein (FLIP) in gastric adenocarcinoma. Clin. Sci. (Lond). 106: 397–405.

4. DISCUSSION

Despite existing evidence that genetic variation in human miRNA regions is capable of altering miRNA function and eventually lead to disease (Hoffman et al. 2009; Mencía et al. 2009), there is a lack of studies that deeply analyse the specific biological consequences of disease associated allelic variants in miRNAs and how they may ultimately lead to disease. The work performed in this thesis, among others (Saunders et al. 2007; Quach et al. 2009), suggests that natural selection may be shaping variation in these regulatory elements, and thus proposes that variation in miRNAs may have important functional consequences. However, still many questions arise: How harmful genetic variation in miRNAs may be? How likely is that a variant affecting a miRNA alters its expression or its affinity for a target gene? How often these variants are altering biological pathways and causing disease? What are the molecular mechanisms that involve these variants in disease?

Interestingly, it seems clear that variants in miRNAs may generate different outcomes, depending on the specific region of the miRNA gene where they are located. The first chapter of the thesis, according with previous analysis (Saunders et al. 2007; Quach et al. 2009), revealed that regions coding for the mature miRNA molecule, not only have less SNP density than the regions flanking the miRNA gene, but also have less variation than the region encoding the miRNA precursor molecule. The most intuitive explanation for these facts is that the region of the miRNA that do not code for the mature miRNA molecule never interacts with the

targeted mRNA and, as a consequence, changes in these regions can only affect the stability of the hairpin molecule and thus, the final expression level of the mature miRNA molecule. Conversely, changes in the mature molecule are able to either increase or decrease the affinity of the miRNA for its target sites in the genes it regulates. This does not necessarily mean that changing the affinity for a targeted gene is more harmful than altering the expression pattern of a miRNA, however, mutations in the mature molecule might affect both the expression patterns and the target spectrum of the miRNA where they are located. Nonetheless, the seed region has long been considered crucial for target recognition and thus, we expected it to accumulate less variation than the rest of the mature miRNA. However, we did not find statistically significant differences between the seed and the mature region in terms of SNP density. This may probably respond to a lack of power of our approach due to the reduced length of the seed region. Other explanations might be that there are other regions within the mature miRNA of nearly equal importance, a hypothesis that has also been suggested by other authors.

The relevance of the seed region was reflected in the population differentiation analysis that we performed. Our approach revealed that variants affecting the seed region tend to have higher F_{ST} values than variants affecting the precursor and mature region of the miRNA, the miRNA flanking regions or the rest of the non-coding genome. Altogether suggesting that, although the mature miRNA molecule and its seed show similar conservation patterns and SNP densities, those variants showing the highest allele frequency

differences among human populations tend to accumulate more in the seed, which is considered the most important region for a proper function of the miRNA. Thanks to those analyses, we found several miRNAs, which are known to perform important roles in disease, affected by genetic variants with high differences in allele frequency across human populations. The study of these genetic variants and their effect in the function of the miRNAs where they are located is of major relevance for population genetics and disease. For instance, rs2168518:G>A in *MIR4513*, which has been proposed as a biomarker for prognosis of coronary artery disease, is among the miRNA variants with higher F_{ST} values in multiple population comparisons (Ghanbari et al. 2014; Li et al. 2015). In this regard, it is worth noticing that coronary artery disease has been reported to have differences in prevalence between populations of African and European descent (Bertrand et al. 1995). Similarly, rs12355840:T>C in *MIR202* was found to be potentially implicated in differences in Cancer outcome between African and non-African populations (Rawlings-Goss et al. 2014). These findings encourage functional approaches similar to the ones performed in this thesis in order to better understand the mechanism underlying phenotype differences among human populations that may be related to miRNA regulation.

Besides total SNP density and pairwise F_{ST}, in the first chapter of this thesis, we also analyzed how variation accumulates in the distinct miRNA regions across frequency. We observed differences in the accumulative allele frequency of the variants of the different miRNA regions. Interestingly all of the regions analyzed (flanking,

precursor, mature and seed) accumulate similar numbers of low frequency allele variants but different amounts of the high frequency ones. The variants located in the flanking region of miRNAs are the ones less affected by allelic frequency, followed by the variants affecting the precursor miRNA region. The seed and mature regions seem to accumulate less high frequency variants than the other regions but no differences were observed between them. Again, the reduced length of the seed region as well as an unknown role of the rest of the mature region could be influencing this result.

In chapters one and two of this thesis, we analyzed in depth several variants genetically associated with Cancer, which affect the mature miRNA molecules, through miRNA over-expression experiments. We compared the expression levels of each miRNA allelic variant focusing only on the affected miRNA branch. Surprisingly, all of the analyzed variants were capable of altering the expression levels of the miRNA where they are located, with the exception of rs35770269 in hsa-miR-449c, our control variant and the only SNV studied that has not been associated with disease. From these results, together with other previously published work (Jazdzewski et al. 2008; Hoffman et al. 2009; Jazdzewski et al. 2009), we could speculate that variation in the secondary structure of the miRNA hairpin is very likely to affect its expression levels. However, we must consider that the studied miRNAs are highly expressed miRNAs involved in disease, and other miRNAs might not present similar expression changes when affected by genetic variation.

These changes in miRNA expression between allelic variants correlated with the differences in minimum free energy of the predicted miRNA folding structures. In all the studied cases, the allele predicted to have a less stable structure, was the same that showed less expression in the RT-qPCR approach. Yet, in the case of the miR-938 described in chapter two, miRNA expression did not correlate with those predictions as both variants were predicted to fold in structures with the same minimum free energy. Nonetheless, we must consider that RNA folding algorithms are far from perfection, since different algorithms often suggest different folding structures; that in biochemistry there is not an absolute folding structure rather than a more probable one: Finally we should also consider that our over-expression approach is based in an artificial system that may magnify what is occurring in physiological al conditions. Also, it would be interesting to confirm the obtained results through the measurement of miRNA levels after transcription inhibition. Through these experiments the expression shifts could be entirely attributed, or not, to stability changes. Unfortunately there is still much to understand about the processes that control miRNA stability before these issues can be resolved.

More studies are necessary in this field in order to understand how these variants, that are able to alter miRNA expression, are influencing disease. In fact, another method that could confirm that a given miRNA variant is altering its miRNA expression and favoring disease would be to compare the expression of each miRNA variant in healthy and pathogenic tissue as well as their regulated/predicted target genes. These experiments would be of

great interest and should be performed in order to establish causality between the expression shift and the studied pathology. In any case, the stability of the precursor miRNA molecule seems to be of enormous importance to miRNA expression levels. This stability is very fragile and seems likely that changes in the mature region of miRNAs can alter it and thus changes the expression levels of the mature miRNA molecule.

In order to study the effect of the miRNA variants, not only in their expression levels, but also in their regulatory function, we analyzed the changes in the pathways associated with the genes deregulated in the array between allelic variants of the same miRNA. We could observe that all the studied miRNA allelic variants deregulated genes that had Cancer among the Top Diseases in which they were associated. Contrary, the test of the networks and pathways in which deregulated genes were found to be involved were a lot more different between miRNA variants.

Besides analyzing the altered pathways in which miRNAs are involved, we also looked at the changes in the regulation of several transcripts of the microarray caused by the miRNAs SNVs. Although these results do not necessarily indicate a direct miRNA regulation, indirect regulation of a transcript can also indicate a potential role of a miRNA in disease and should be studied. Some of the most interesting deregulated transcripts are mentioned in both chapters; however, we focused our attention in those transcripts that, not only were deregulated in the array of a specific miRNA, but also were predicted to be regulated by the same miRNA allelic

variant using PITA algorithm. We played special attention to those transcripts that showed differences between alleles and that were also involved in Cancer. The luciferase reporter analysis performed with those transcripts, showed regulation of *CXCL12* by miR-938 and *CDH1* and *CHL1* by miR-499a-3p in an allele dependent fashion. All of these genes are involved in carcinogenic diseases such us the ones to which those miRNAs variants have been associated. Of special interest is the case of *CXCL12*, which not only it is regulated by miR-938, but also it is differentially regulated by each of its rs12416605 allele variants. So far, very few such cases have been reported.

CDH1 is an important element in cell adhesion and has been related to tumor differentiation, metastasis, and infiltrative tumor growth among other biological aspects of Cancer in different cancer types (Liu and Chu 2014; Yu et al. 2015). In fact, mutations in this gene, which has been proposed as marker of prognosis, and down-regulation of it, favors tumor progression in different types of Cancers (Berx et al. 1998; Wang et al. 2004; Masciari et al. 2007; van der Post et al. 2015). Moreover, *CDH1* is also involved in apoptosis, which represents an important defense against tumor formation. Thus, finding that a SNV associated with Cancer is able of altering *CDH1* regulation, is a remarkable finding and should be further studied.

Similarly, *CHL1* is another gene involved in cell adhesion involved in many Cancer types (Senchenko et al. 2011; He et al. 2013). Further, the regulation of *CHL1* by other miRNAs has previously been found to promote several types of Cancer (Zhu et al. 2014;

Chu et al. 2014b). Therefore, the here described regulation of *CHL1* altered by a Cancer associated genetic variant, is a notable result and should be analyzed in depth.

Also, the equilibrium between *CXCL12* and its receptor *CXCR4* is very important for tumor growth, metastasis, tumor vasculogenesis and cellular microenvironment maintenance in several types of Cancer (Domanska et al. 2013). In fact, *CXCR4* antagonists are used for Cancer therapy (Domanska et al. 2013). However, although the involvement of CXCL12 in Cancer is clear, the detailed role that it plays still generates controversy among researchers (Zhi et al. 2012; Rubie et al. 2015; Izumi et al. 2016). In this regard, we have shown that Cancer related *CXCL12* is a direct target of miR-938. Moreover, not only we have found that a genetic variant located in miR-938 is genetically associated with Gastric Cancer, but also that each of the allele variants from rs12416605 shows significant differences in the capacity of regulating *CXCL12*. This represents a relevant finding.

Nevertheless, all these results require further analysis such as mutagenesis experiments It is worth to remark that all the above-described results cannot fully demonstrate a physiological miRNA-target gene regulation and would require further analyses such as the study of endogenous protein and mutagenesis experiments. The last could lead to confirm the specific target site where the miRNA binds to repress the transcript translation. It is also important to consider the relevance of the cell type used in these experiments. HeLA cells are considered a very useful model, especially when studying Cancer. However, each of our variants has been associated

with a particular type of Cancer and other cellular models would have been more suitable for the study of their effects. For instance, rs12416605 in miR-938, is associated with Gastric Cancer, as described in chapter two, and cells derived from gastric adenocarcinoma, such as AGS, would have been more appropriated.

However, what is more interesting, besides the specific effect that each variant might have in the expression of one of the miRNA mature molecules or in the regulation of specific target genes, is the final effect that these variants might have in the cell where they are expressed. In this regard, it is interesting to consider that although the differences in the regulation of *CDH1* and *CHL1* between the allele variants of miR-499a-3p did not reach statistical significance when being compared between them, in *in vivo* conditions the differences in regulation between alleles might be greater, as expression levels were artificially equal between allele variants in the luciferase reporter assay, while expression approach suggested significant differences in the expression of both alleles. The same happens with both allelic variants of the miR-938, although in this case the differences in *CXCL12* regulation between allele variants were statistically significant. Because the constructs used in the microarray approach contained the complete miRNA hairpin, the expression of each mature miRNA was not artificially forced to be the same. Thus, any difference in the expression of an array transcript could be the result of either a variation in the expression of one of the allelic variants of the miRNA, or a major affinity between a miRNA variant for a transcript. Also, when analyzing the

results of the microarray, it is important to consider that both of the potentially functional mature miRNA molecules that can result from a hairpin structure might be participating of the effects observed in this approach. Despite resembling more the natural scenario with this approach than with the use of specific synthetic mature miRNAs, the expression of both potentially mature miRNA molecules, together with the differences in expression between allelic variants, could be causing a lack of power to detect deregulated transcripts in this approach and should be considered when analyzing these results. Furthermore, another way that we could have been used in order to analyze allele differences in gene regulation could have been the transfection of antimirs in cells expressing both the targeted gene and the specific miRNA variant. This could have helped us to confirm the hypothesis of miRNA allele specific regulation as an important genetic factor contributing to Cancer. It is also worth noticing that in HeLa cells transfected with miR-499a constructs we observed an increase in cellular death compared with cells transfected with control constructs. Functional experiments that distinguish between different biological processes such us cellular survival, necrosis or apoptosis would be crucial to further understand the effect of these variants in the biological function of these miRNAs

When analyzing variation in miRNAs it is important to remember that ncRNAs follow different rules to coding regions. For instance, although a change in a DNA coding region can truncate the function of a protein, it is very unlikely that changing a nucleotide of this

sequence will make this protein to perform a new and noxious function, whereas a change in the seed region of a miRNA not only may result in a reduction of its expression or affinity for a given mRNA, but also may affect the expression of other mRNAs that are not directly targeted by the miRNA. Moreover, although a single miRNA can regulate up to several hundreds of genes, each mRNA can also be regulated by many miRNAs at the same time, thus attenuating the potentially detrimental effect of genetic variants in miRNAs. As other ncRNA regulators, miRNAs tolerate more variants than coding regions. However, a great part of this variation seems to be rare and affect mainly the precursor region of the miRNA altering only the expression levels of it.

Some authors have found that highly expressed miRNAs are more conserved than miRNAs with lower expression (Berezikov 2011b). It is difficult to know whether these miRNAs with reduced expression are functional miRNAs or are just a biological source of innovation able to tolerate variation in order to acquire a new function. Obviously, changes in the expression of less expressed miRNAs will be less harmful as their absolute expression level will rarely change much and their potential effect on a highly expressed target will be limited. Moreover, the organization of miRNAs in families that share seed sequence and even location, together with the fact that some miRNAs have several copies in the genome, may be a sign that miRNAs are organized in a way to tolerate variation.

Altogether supports miRNA variation as an important source of phenotypic variation that may cause pathology and thus should thus be considered when studding the molecular bases of human disease.

5. CONCLUSIONS

- Distinct miRNA regions show significant differences in genetic variation density, being the seed and mature the miRNA regions with less variation compared with the precursor and flanking regions.
- The genetic variation density in distinct miRNA regions is differentially affected by SNV allele frequency. High frequency variants are accumulated faster in the flanking and precursor miRNA regions than in the mature and seed regions.
- The miRNA seed region consistently shows higher mean F_{ST} index than any of the other miRNA regions. These differences are statistically significant when comparing the seed and the miRNA flanking regions between African and European populations suggesting the existence of local adaptation in the seed region.
- The variants genetically associated with Cancer rs2910164 in miR-146a, rs11614913 in miR196a, rs12416605 in miR-938 and rs3746444 both in miR-499a and miR-499b alter the expression of these miRNAs in an allele dependent manner.
- The variants genetically associated with Cancer rs2910164 in miR-146a, rs11614913 in miR196a, rs12416605 in miR-938 and rs3746444 both in miR-499a and miR-499b affect the direct or indirect regulation of Cancer related genes and pathways in an allele specific manner.
- MicroRNA miR-499a directly targets the cadherin *CHL1* and the cell adhesion molecule *CDH1* and

both genes, which are differentially regulated by the two alleles of rs3746444:A>G.

- MicroRNA miR-938 directly targets the chemokine *CXCL12* gene, which is differentially regulated by the two alleles of rs12416605: C>T.

References:

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56–65.

Ahn K, Gim JA, Ha HS, Han K, Kim HS. 2013. The novel MER transposon-derived miRNAs in human genome. Gene 512: 422–428.

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, et al. 2003. A uniform system for microRNA annotation. RNA 9: 277–9.

Bartel DP. 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell. 116 2:281-297.

Berezikov E. 2011a. Evolution of microRNA diversity and regulation in animals. Nat. Rev. Genet. 12: 846–860.

Berezikov E. 2011b. Evolution of microRNA diversity and regulation in animals. Nat. Rev. Genet. 12: 846–60.

Berezikov E, Cuppen E, Plasterk RH a. 2006. Approaches to microRNA discovery. Nat. Genet. 38 Suppl: S2–7.

Betrand E. 1995. Coronary heart disease in black Africans: an overview. East African Med. J. 72: 37-41.

Berx G, Becker KF, Höfler H, Roy F Van. 1998. Mutations of the human E-cadherin (CDH1) gene. Human Mutation. 12:226-237

Blow MJ, Grocock RJ, Dongen S van, Enright AJ, Dicks E, Futreal PA, Wooster R, Stratton MR. 2006. RNA editing of human microRNAs. Genome Biol. 7: R27.

Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, Calabrò E, Croce CM, Pastorino U, Sozzi G. 2011. MicroRNA

signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. Proc. Natl. Acad. Sci. U. S. A. 108: 3713–8.

Borchert GM, Holton NW, Williams JD, Hernan WL, Bishop IP, Dembosky JA, Elste JE, Gregoire NS, Kim J-A, Koehler WW, Lengerich JC, Medema AA, et al. 2011. Comprehensive analysis of microRNA genomic loci identifies pervasive repetitive-element origins. Mob. Genet. Elements 1: 8–17.

Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, Croce CM. 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc. Natl. Acad. Sci. U. S. A. 101: 2999–3004.

Carbonell J, Alloza E, Arce P, Borrego S, Santoyo J, Ruiz-Ferrer M, Medina I, Jimenez-Almazan J, Mendez-Vidal C, Gonzalez-del Pozo M, Vela A, Bhattacharya SS, et al. 2012.

Casals F, Bertranpetit J. 2012. Human Genetic Variation, Shared and Private. Science (80-.). 337: 39–40.

Chen J-F, Murchison EP, Tang R, Callis TE, Tatsuguchi M, Deng Z, Rojas M, Hammond SM, Schneider MD, Selzman CH, Meissner G, Patterson C, et al. 2008a. Targeted deletion of Dicer in the heart leads to dilated cardiomyopathy and heart failure. Proc. Natl. Acad. Sci. U. S. A. 105: 2111–6.

Chen K, Rajewsky N. 2006. Natural selection on human microRNA binding sites inferred from SNP data. Nat. Genet. 38: 1452–1456.

Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, Guo X, Li Q, Li X, et al. 2008b. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. 18: 997–1006.

Chu H, Zhong D, Tang J, Li J, Xue Y, Tong N, Qin C, Yin C, Zhang Z, Wang M. 2014a. A functional variant in miR-143 promoter contributes to prostate cancer risk. Arch. Toxicol. 403–414.

Chu Y, Ouyang Y, Wang F, Zheng A, Bai L, Han L, Chen Y, Wang H. 2014b. MicroRNA-590 promotes cervical cancer cell growth and invasion by targeting CHL1. J. Cell. Biochem. 115: 847–853.

Cimmino A, Calin GA, Fabbri M, Iorio M V, Ferracin M, Shimizu M, Wojcik SE, Aqeilan RI, Zupo S, Dono M, Rassenti L, Alder H, et al. 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A 102: 13944–13949.

Clark MB, Choudhary A, Smith MA, Taft RJ, Mattick JS. 2013. The dark matter rises: the expanding world of regulatory RNAs. Essays Biochem. 54: 1–16.

Clark MB, Mattick JS. 2011. The dark matter rises: the expanding world of regulatory RNAs.Essays in Biochemistry. 54:1-16.

Consortium T 1000 GP. 2015. A global reference for human genetic variation. Nature 526: 68–74.

Consortium TEP. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science (80-.). 306: 636–640.

Domanska UM, Kruizinga RC, Nagengast WB, Timmer-Bosscha H, Huls G, Vries EGE de, Walenkamp AME. 2013. A review on CXCR4/CXCL12 axis in oncology: no place to hide. Eur. J. Cancer 49: 219–30.

Dweep H, Sticht C, Gretz N. 2013. In-Silico Algorithms for the Screening of Possible microRNA Binding Sites and Their Interactions. Curr. Genomics 14: 127–36.

Ellinger J. 2015. Circulating serum miRNA (miR-367-3p, miR-371a-3p, miR-372-3p and miR-373-3p) as biomarkers in patients with testicular germ cell cancer. J. Urol. 193: 331–337.

Esquela-Kerscher A, Slack FJ. 2006. Oncomirs - microRNAs with a role in cancer. Nat. Rev. Cancer 6: 259–69.

Fan Y, Fan J, Huang L, Ye M, Huang Z, Wang Y, Li Q, Huang J. 2015. Increased expression of microRNA-196a predicts poor prognosis in human ovarian carcinoma. 8: 4132–4137.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2015. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer 136: E359–E386.

Filipowicz W, Bhattacharyya SN, Sonenberg N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9: 102–14.

Friedman RC, Farh KKH, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19: 92–105.

Fu B, Song P, Lu M, Wang B, Zhao Q. 2014. The association between miR-146a gene rs2910164 polymorphism and gastric cancer risk: A meta-analysis. Biomed. Pharmacother. 68: 923–928.

Garzon R, Calin GA, Croce CM. 2009. MicroRNAs in Cancer. Annu. Rev. Med. 60: 167–79.

Ghanbari M, Vries PS de, Looper H de, Peters MJ, Schurmann C, Yaghootkar H, Dörr M, Frayling TM, Uitterlinden AG, Hofman A, Meurs JBJ van, Erkeland SJ, et al. 2014. A Genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. Hum. Mutat. 35: 1524–1531.

Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, Lander ES, Schaffner SF, et al. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327: 883–886.

Hansen T, Olsen L, Lindow M, Jakobsen KD, Ullum H, Jonsson E, Andreassen OA, Djurovic S, Melle I, Agartz I, Hall H, Timm S, et al. 2007. Brain expressed microRNAs implicated in schizophrenia etiology. PLoS One 2.:

Hayashizaki Y. 2003. The Riken mouse genome encyclopedia project. Comptes Rendus - Biol. 326: 923–929.

HD collaborative, research group. 1993. A novel gene containing a

trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group [see comments]. Cell 72: 971–983.

He LH, Ma Q, Shi YH, Ge J, Zhao HM, Li SF, Tong ZS. 2013. CHL1 is involved in human breast tumorigenesis and progression. Biochem. Biophys. Res. Commun. 438: 433–438.

Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF. 2006. The expansion of the metazoan microRNA repertoire. BMC Genomics 7: 25.

Hertel J, Stadler P. 2015. The Expansion of Animal MicroRNA Families Revisited. Life 5: 905–920.

Hoffman AE, Zheng T, Yi C, Leaderer D, Weidhaas J, Slack F, Zhang Y, Paranjape T, Zhu Y. 2009. microRNA miR-196a-2 and breast cancer: A genetic and epigenetic association study and functional analysis. Cancer Res. 69: 5970–5977.

Hudson RS, Yi M, Esposito D, Glynn SA, Starks AM, Yang Y, Schetter AJ, Watkins SK, Hurwitz AA, Dorsey TH, Stephens RM, Croce CM, et al. 2013. MicroRNA-106b-25 cluster expression is associated with early disease recurrence and targets caspase-7 and focal adhesion in human prostate cancer. Oncogene 32: 4139–47.

Hutvágner G, Zamore PD. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. Science 297: 2056–60.

Iwama H, Kato K, Imachi H, Murao K, Masaki T. 2013. Human microRNAs originated from two periods at accelerated rates in mammalian evolution. Mol. Biol. Evol. 30: 613–626.

Izumi D, Ishimoto T, Miyake K, Sugihara H, Eto K, Sawayama H, Yasuda T, Kiyozumi Y, Kaida T, Kurashige J, Imamura Y, Hiyoshi Y, et al. 2016. CXCL12/CXCR4 activation by cancer-associated fibroblasts promotes integrin β1 clustering and invasiveness in gastric cancer. Int. J. Cancer 138: 1207–1219.

Jansson MD, Lund AH. 2012. MicroRNA and cancer. Mol. Oncol. 6: 590–610.

Jazdzewski K, Liyanarachchi S, Swierniak M, Pachucki J, Ringel MD, Jarzab B, la Chapelle A de. 2009. Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. Proc. Natl. Acad. Sci. U. S. A. 106: 1502–5.

Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, la Chapelle A de. 2008. Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. Proc. Natl. Acad. Sci. U. S. A. 105: 7269–7274.

Jin XF, Wu N, Wang L, Li J. 2013. Circulating microRNAs: A novel class of potential biomarkers for diagnosing and prognosing central nervous system diseases. Cell. Mol. Neurobiol. 33: 601–613.

Junn E, Mouradian MM. 2012. MicroRNAs in neurodegenerative diseases and their therapeutic potential. Pharmacology and Therapeutics. 13:143-150.

Kamaraj B, Purohit R. 2014. Mutational analysis of oculocutaneous albinism: A compact review. Biomed. Research International.

Kawahara Y, Megraw M, Kreider E, Iizasa H, Valente L, Hatzigeorgiou AG, Nishikura K. 2008. Frequency and fate of microRNA editing in human brain. Nucleic Acids Res. 36: 5270–5280.

Kloosterman WP, Plasterk RHA. 2006. The diverse functions of microRNAs in animal development and disease. Dev. Cell 11: 441–50.

Kota J, Chivukula RR, O'Donnell KA, Wentzel EA, Montgomery CL, Hwang HW, Chang TC, Vivekanandan P, Torbenson M, Clark KR, Mendell JR, Mendell JT. 2009. Therapeutic microRNA Delivery Suppresses Tumorigenesis in a Murine Liver Cancer Model. Cell 137: 1005–1017.

Kozomara A, Griffiths-Jones S. 2014. MiRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42.:

Krol J, Loedige I, Filipowicz W. 2010. The widespread regulation of microRNA biogenesis, function and decay. Nat. Rev. Genet. 11:

597–610.

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, et al. 2007. A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. Cell 129: 1401–1414.

Lanford RE, Hildebrandt-Eriksen ES, Petri A, Persson R, Lindow M, Munk ME, Kauppinen S, Ørum H. 2010. Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. Science 327: 198–201.

Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 75:843-854.

Lewis BP, Shih I, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian microRNA targets. Cell 115: 787–98.

Lewis M a, Quint E, Glazier AM, Fuchs H, Angelis MH De, Langford C, Dongen S van, Abreu-Goodger C, Piipari M, Redshaw N, Dalmay T, Moreno-Pelayo MA, et al. 2009. An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. Nat. Genet. 41: 614–618.

Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74: 175–195.

Li J, Liu Y, Xin X, Kim TS, Cabeza EA, Ren J, Nielsen R, Wrana JL, Zhang Z. 2012. Evidence for positive selection on a number of MicroRNA regulatory interactions during recent human evolution. PloS. Genet. 8:1002578.

Li Q, Chen L, Chen D, Wu X, Chen M. 2015. Influence of microRNA-related polymorphisms on clinical outcomes in coronary artery disease. Am. J. Transl. Res. 7: 393–400.

Liu X, Chu KM. 2014. E-cadherin and gastric cancer: Cause, consequence, and applications. Biomed Research International.

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D,

Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, et al. 2005. MicroRNA expression profiles classify human cancers. Nature 435: 834–838.

Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q. 2008. An analysis of human microRNA and disease associations. PLoS One 3: 1–5.

Ma L, Reinhardt F, Pan E, Soutschek J, Bhat B, Marcusson EG, Teruya-Feldstein J, Bell GW, Weinberg RA. 2010. Therapeutic silencing of miR-10b inhibits metastasis in a mouse mammary tumor model. Nat. Biotechnol. 28: 341–7.

Maas S, Patt S, Schrey M, Rich A. 2001. Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. Proc. Natl. Acad. Sci. U. S. A. 98: 14687–92.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, et al. 2009. Finding the missing heritability of complex diseases. Nature 461: 747–753.

Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozlowski P. 2011. Copy number variation of microRNA genes in the human genome. BMC Genomics 12: 183.

Marco A, Ninova M, Ronshaugen M, Griffiths-Jones S. 2013. Clusters of microRNAs emerge by new hairpins in existing transcripts. Nucleic Acids Res. 41: 7745–52.

Masciari S, Larsson N, Senz J, Boyd N, Kaurah P, Kandel MJ, Harris LN, Pinheiro HC, Troussard a, Miron P, Tung N, Oliveira C, et al. 2007. Germline E-cadherin mutations in familial lobular breast cancer. J. Med. Genet. 44: 726–731.

Mattick JS. 2011. The double life of RNA. Biochimie 93: viii–ix.

Mencía A, Modamio-Høybjør S, Redshaw N, Morín M, Mayo-Merino F, Olavarrieta L, Aguirre L a, Castillo I del, Steel KP, Dalmay T, Moreno F, Moreno-Pelayo MA. 2009. Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat. Genet. 41: 609–613.

Mendell JT, Olson EN. 2012. MicroRNAs in stress signaling and human disease. Cell 148: 1172–1187.

Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, et al. 2008. Circulating microRNAs as stable blood-based markers for cancer detection. Proc. Natl. Acad. Sci. U. S. A. 105: 10513–8.

Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA. 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 18: 610–621.

Morris K V, Mattick JS. 2014. The rise of regulatory RNA. Nat. Rev. Genet. 15: 423–37.

Mu XJ, Lu ZJ, Kong Y, Lam HYK, Gerstein MB. 2011. Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project. Nucleic Acids Res. 39: 7058–7076.

Muinos-Gimeno M, Montfort M, Bayes M, Estivill X, Espinosa-Parrilla Y. 2010. Design and evaluation of a panel of single-nucleotide polymorphisms in microRNA genomic regions for association studies in human disease. Eur J Hum Genet 18: 218–226.

Negi V, Paul D, Das S, Bajpai P, Singh S, Mukhopadhyay A, Agrawal A, Ghosh B. 2015. Altered expression and editing of miRNA-100 regulates iTreg differentiation. Nucleic Acids Res. 43: 8057–8065.

Neilsen CT, Goodall GJ, Bracken CP. 2012. IsomiRs - The overlooked repertoire in the dynamic microRNAome. Trends in Genetics. 28:544-549.

Ng EK, Chong WW, Jin H, Lam EK, Shin VY, Yu J, Poon TCW, Ng SS, Sung JJY. 2009. Differential expression of microRNAs in plasma of colorectal cancer patients: A potential marker for colorectal cancer screening. Gut.

Oerlemans MIFJ, Mosterd A, Dekker MS, Vrey EA de, Mil A van, Pasterkamp G, Doevendans PA, Hoes AW, Sluijter JPG. 2012. Early assessment of acute coronary syndromes in the emergency department: The potential diagnostic value of circulating microRNAs. EMBO Mol. Med. 4: 1176–1185.

Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R. 2010. Long noncoding RNAs with enhancer-like function in human cells. Cell 143: 46–58.

Peterson SM, Thompson J a., Ufkin ML, Sathyanarayana P, Liaw L, Congdon CB. 2014. Common features of microRNA target prediction tools. Front. Genet. 5: 1–10.

Piriyapongsa J, Mariño-Ramírez L, Jordan IK. 2007. Origin and evolution of human microRNAs from transposable elements. Genetics 176: 1323–1337.

Post RS van der, Vogelaar IP, Carneiro F, Guilford P, Huntsman D, Hoogerbrugge N, Caldas C, Schreiber KEC, Hardwick RH, Ausems MGEM, Bardram L, Benusiglio PR, et al. 2015. Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. J. Med. Genet. 52: 361–374.

Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, Macdonald PE, Pfeffer S, Tuschl T, Rajewsky N, Rorsman P, Stoffel M. 2004. A pancreatic islet-specific microRNA regulates insulin secretion. Nature 432: 226–30.

Qu H, Fang X. 2013. A Brief Review on the Human Encyclopedia of DNA Elements (ENCODE) Project. Genomics, Proteomics Bioinforma. 11: 135–141.

Quach H, Barreiro LB, Laval G, Zidane N, Patin E, Kidd KK, Kidd JR, Bouchier C, Veuille M, Antoniewski C, Quintana-Murci L. 2009. Signatures of purifying and local positive selection in human miRNAs. Am. J. Hum. Genet. 84: 316–27.

Rawlings-Goss RA, Campbell MC, Tishkoff SA. 2014. Global population-specific variation in miRNA associated with cancer risk and clinical biomarkers. BMC Med. Genomics 7: 53.

ReyesHerrera PH, Ficarra E. 2012. One Decade of Development and Evolution of MicroRNA Target Prediction Algorithms. Genomics, Proteomics Bioinforma. 10: 254–263.

Riken T, Exploration G, Ii P, Consortium F, Gene M, Project E, Mouse T, Encyclopaedia G, Strategies D, Fig SI, Decoder R, Encyclo- MG, et al. 2001. Functional annotation of a full-length mouse cDNA collection. Nature 409: 685–90.

Rooij E Van. 2011. The art of MicroRNA research. Circ. Res. 108: 219–234.

Rubie C, Kauffels A, Kï¿½lsch K, Glanemann M, Justinger C. 2015. CXCL12/CXCR4 display an inverse mRNA expression profile in gastric carcinoma that correlates with tumor progression. Oncol. Lett. 360–364.

Saini HK, Griffiths-Jones S, Enright AJ. 2007. Genomic analysis of human microRNA transcripts. Proc. Natl. Acad. Sci. U. S. A. 104: 17719–24.

Saunders M a, Liang H, Li W-H. 2007. Human polymorphism at microRNAs and microRNA target sites. Proc. Natl. Acad. Sci. U. S. A. 104: 3300–5.

Senchenko VN, Krasnov GS, Dmitriev AA, Kudryavtseva A V., Anedchenko EA, Braga EA, Pronina I V., Kondratieva TT, Ivanov S V., Zabarovsky ER, Lerman MI. 2011. Differential expression of CHL1 Gene during development of major human cancers. PLoS One 6.:

Seong IS, Ivanova E, Lee JM, Choo YS, Fossale E, Anderson MA, Gusella JF, Laramie JM, Myers RH, Lesort M, MacDonald ME. 2005. HD CAG repeat implicates a dominant property of huntingtin in mitochondrial energy metabolism. Hum. Mol. Genet. 14: 2871–2880.

Shingara J, Keiger K, Shelton J, Laosinchai-Wolf W, Powers P, Conrad R, Brown D, Labourier E. 2005. An optimized isolation and labeling platform for accurate microRNA expression profiling. RNA 11: 1461–70.

Small EM, Olson EN. 2011. Pervasive roles of microRNAs in cardiovascular biology. Nature 469: 336–342.

Sohn W, Kim J, Kang SH, Yang SR, Cho J-Y, Cho HC, Shim SG, Paik Y-H. 2015. Serum exosomal microRNAs as novel biomarkers for hepatocellular carcinoma. Exp. Mol. Med. 47: e184.

Stenvang J, Silahtaroglu AN, Lindow M, Elmen J, Kauppinen S. 2008. The utility of LNA in microRNA-based cancer diagnostics and therapeutics. Semin. Cancer Biol. 18: 89–102.

Su Z, Yang Z, Xu Y, Chen Y, Yu Q. 2015. MicroRNAs in apoptosis, autophagy and necroptosis. Oncotarget 6: 8474–90.

Tao J, Wu H, Lin Q, Wei W, Lu X-H, Cantle JP, Ao Y, Olsen RW, Yang XW, Mody I, Sofroniew M V, Sun YE. 2011. Deletion of astroglial Dicer causes non-cell-autonomous neuronal dysfunction and degeneration. J. Neurosci. 31: 8306–8319.

Vlachos IS, Hatzigeorgiou AG. 2013. Online resources for miRNA analysis. Clin. Biochem. 46: 879–900.

Volinia S, Calin GA, Liu C-G, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, et al. 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc. Natl. Acad. Sci. U. S. A. 103: 2257–61.

Wan D, Gu W, Xu G, Shen C, Ding D, Shen S, Wang S, Gong X, He S, Zhi Q. 2014. Effects of common polymorphisms rs2910164 in miR-146a and rs11614913 in miR-196a2 on susceptibility to colorectal cancer : a systematic review meta-analysis. 792–800.

Wang F, Long G, Zhao C, Li H, Chaugai S, Wang Y, Chen C, Wang DW. 2013. Plasma microRNA-133a is a new marker for both acute myocardial infarction and underlying coronary artery stenosis. J. Transl. Med. 11: 222.

Wang HD, Ren J, Zhang L. 2004. CDH1 germline mutation in hereditary gastric carcinoma. World J. of Gastroenterol. 10:3088-3093.

Wei WJ, Wang YL, Li DS, Wang Y, Wang XF, Zhu YX, Yang YJ, Wang ZY, Ma YY, Wu Y, Jin L, Ji QH, et al. 2013. Association between the rs2910164 Polymorphism in Pre-Mir-146a Sequence and Thyroid Carcinogenesis. PLoS One 8.:

Weir BS, Hill WG. 2002. Estimating F-statistics. Annu. Rev. Genet. 36: 721–750.

Xu B, Feng N-H, Li P-C, Tao J, Wu D, Zhang Z-D, Tong N, Wang J-F, Song N-H, Zhang W, Hua L-X, Wu H-F. 2010a. A functional polymorphism in Pre-miR-146a gene is associated with prostate cancer risk and mature miR-146a expression in vivo. Prostate 70: 467–472.

Xu B, Hsu P-K, Stark KL, Karayiorgou M, Gogos JA. 2013. Derepression of a Neuronal Inhibitor due to miRNA Dysregulation in a Schizophrenia-Related Microdeletion. Cell 152: 262–275.

Xu B, Karayiorgou M, Gogos JA. 2010b. MicroRNAs in psychiatric and neurodevelopmental disorders. Brain Research. 1338:78-88.

Xu P, Vernooy SY, Guo M, Hay BA. 2003. The Drosophila microRNA mir-14 suppresses cell death and is required for normal fat metabolism. Curr. Biol. 13: 790–795.

Yin JQ, Zhao RC, Morris K V. 2008. Profiling microRNA expression with microarrays. Trends in Biotechnology. 26:70-76.

Yu Q, Guo Q, Chen L, Liu S. 2015. Clinicopathological significance and potential drug targeting of CDH1 in lung cancer: A meta-analysis and literature review. Drug Design, Development and Therapy. 9:2171-2178.

Zeng L, Cui J, Wu H, Lu Q. 2014. The emerging role of circulating microRNAs as biomarkers in autoimmune diseases. Autoimmunity 6934: 1–11.

Zhi Y, Chen J, Zhang S, Chang X, Ma J, Dai D. 2012. Down-regulation of CXCL12 by DNA hypermethylation and its involvement in gastric cancer metastatic progression. Dig. Dis. Sci. 57: 650–659.

Zhou X, Chen J, Tang W. 2014. The molecular mechanism of HOTAIR in tumorigenesis, metastasis, and drug resistance. Acta Biochimica et Biophysica Sinica. 46:1011-1015.

Zhu H, Fang J, Zhang J, Zhao Z, Liu L, Wang J, Xi Q, Gu M. 2014. MiR-182 targets CHL1 and controls tumor growth and invasion in papillary thyroid carcinoma. Biochem. Biophys. Res. Commun. 450: 857–862.