Addressing functional and evolutionary implications of microRNA variation at the DNA and RNA levels in primates

Alicia Gallego Jiménez

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 mis padres, Enrique y Sol

We must always remember with gratitude and admiration the first sailors who steered their vessels through storms and mists, and increased our knowledge of the lands of ice in the South

Roald Amundsen (1911)

"Men wanted for hazardous journey. Small wages, bitter cold, long months of complete darkness, constant danger, safe return doubtful. Honour and recognition in case of success."

Newspaper ad to recruit men for the Imperial Trans-Antarctic Expedition. Attributed to Ernest Shackleton (1914)

ACKNOWLEDGEMENTS

El doctorado pasa, quizás, por largos y repetidos periodos de angustia, soledad y pérdida del sentido *¿Hacia dónde quiero llegar con lo que estoy haciendo? Esta gente ya lo dijo hace años. ¿Puede una sola persona hacer frente a tantas preguntas? Esto no va a funcionar. En la historia del pensamiento faltan muchos nombres de mujer. ¿Cuántas vidas necesitaría yo para poder leer todo esto?* Durante estos años a veces perdí un poco el norte. Es por eso quizás que a veces también me quejé demasiado. Pero algunos ya se dieron cuenta que, bajo esta fachada de escepticismo y crítica incesantes, hay una persona que aprecia la vida por encima de todo. A pesar de todos los momentos de flaqueza, mi sentimiento es de enorme alegría y gratitud por estos cuatro años. Lo más importante que aprendí en esta tesis fue que ningún proyecto puede llevarse a cabo en solitario. El trabajo que aquí se recoge ha sido posible gracias a tantas personas que solo he podido citar a algunas de ellas. Muchas gracias:

Al PRBB, al IBE y a BioEvo por ser un espacio de intercambio y reflexión permanentes. Ha sido un privilegio enorme poder rodearme de tantas personas, escuchar tantos seminarios, proyectos, debates y hacerlo desde tantos y tan variados puntos de vista. Gracias por ser una fuente de inspiración y crítica continuas donde todas las voces tienen cabida. Ojalá sepamos trabajar por transmitir la importancia de mantener y multiplicar lugares de encuentro y trabajo como éste.

Al microgrupo, que siempre quedará en mi recuerdo bebiendo café y discutiendo sobre cuadernos y apuntes llenos de fotos de geles y garabatos.

A Yolanda, por haber confiado en mí, haber puesto todos los recursos disponibles y haberme ayudado a desarrollar mi propia visión sobre la investigación y la ciencia. Gracias por haberme apoyado en momentos de máxima dificultad, en el plano laboral pero también, y sobre todo, en el personal. Gracias por todo el cariño de tu familia y en especial, el de Ivo. A Eva, por ser mi maestra y compañera. Gracias por tu paciencia, y por acompañarme y enseñarme siempre, más allá de tus responsabilidades. A mis camaradas y hermanos, María e Ignasi.

Porque el trabajo que aquí se recoge es en gran parte vuestro. Gracias por guiarme desde el primer día. Por tantas horas juntos. Por los debates espirituales y teóricos. Y sobre todo, gracias por tantas risas y por hacer tan alegre esta travesía.

A Marta e Ingrid, por todo su esfuerzo. Por siempre tener tiempo para explicarme cualquier duda, grande o pequeña. Por trabajar en equipo y mantener la cabeza alta y optimista ante cualquier contratiempo. A Diego y a Marina por vuestras geniales ideas, por las horas de escucha y por tanta motivación y entusiasmo. Ha sido una suerte inmensa poder aprender tanto de los cuatro y de vuestra manera cuidadosa, crítica y honesta de hacer ciencia.

Thanks a lot Kawahara-sensei and all the colleagues from the Osaka University for your kindness and for dedicating me all your time and efforts during my stay. The second part of this thesis would have not been possible without your knowledge and suggestions. It has been a pleasure to learn from one of the masters, and to feel part of your group. I will never forget my days in Osaka and the smell of the bamboo.

A Marcelo González y a todas las personas del Instituto Antártico Chileno y de la expedición 2014/2015. Infinitas gracias por permitirme conocer y estudiar uno de los rincones más increíbles del mundo y por enseñarme el valor del trabajo en equipo y el de no rendirse nunca ante la adversidad. En especial, gracias a Claudia y Andrea por demostrarme que los problemas logísticos siempre tienen una solución. Solo hacen falta ingenio y disposición.

Al Taper. Gracias con todo mi corazón por ser el *"by far"* de mi doctorado. Gracias Adriano, Àlex, Arturo, Diego, Elena, Fede, Joha, Juan, Lara, Lisa, Marc, Marco, María, Marina, Nino, Pierre, Vero, Vicky y a todos los amigos visitantes, por ser una gran familia y enseñarme que sólo las cosas tienen sentido cuando se comparten, y que todo (lo profundo y también lo aparente) pueden y deben debatirse hasta la saciedad, una y mil veces. Espero que nos juntemos por muchos años a discutir sobre cualquier tema entorno a grandes

mesas y copiosos platos. Realmente, lo más importante de mi doctorado lo he aprendido junto a vosotros.

A todos los bioevos que me dieron cuerda para polemizar y fantasear en el laboratorio, en los pasillos y hasta en la radio: a Jess, Ludovica, Guillem, Javi, Max, Salva, Carlos, Raúl, Mònica, Belén, Gabriel, Mayukh, Íñigo, Nuria, André, Carla. A Neus por estar a mi lado cada día con esa sonrisa y por demostrarme que hacer deporte no está reñido con ser una maravillosa persona. A Judit por entender mis serios problemas con la burocracia y por ayudarme desde los rincones más perdidos del planeta. Gracias en general a todo bioevo por crear entre todos este lugar de trabajo tan gratificante y amable que, estoy segura, voy a extrañar mucho.

A mis amigas de Barcelona, Madrid, Valencia, Copenhagen, Punta Arenas y Valdivia. Gracias por hacer que las ciudades merezcan tanto la pena. A Diego, por los jugos, las quesadillas, la hamaca y por obligarme a ir sin zapatillas. Por enseñarme que la optimización del tiempo es una utopía y por hacer que los meses en Poble Sec hayan sido tan divertidos e inolvidables. A María por los grupos de culto, por las sesiones de peluquería, por el Raval, la radio, por el yoga, las chelas y los tacos, los viajes en coche y por esperar siempre a que llegue. Llegaré. A Silvia por las noches de pizza, la política, el teatro, por obligarme a darle una visión más abierta a la cadenita de la vida y por ser un pilar fundamental que sustenta Barcelona. A Mattea, por ayudarme tanto con esta tesis y permitir que esta etapa se cierre con la alegría de su llegada. A Zaida, Ana, Manuela, Lalos, Dani, Blanca y Carlota porque fuisteis indispensables en la distancia. A Edu Circonite, Club de Canciones, Miguel Atienza y Hi Jauh USB, por hacer de Barcelona un experimento colectivo ilusionante al que siempre voy a querer volver.

A mi madre Sol y a mi hermana Julia, por confiar siempre en mí y animarme a superar cualquier miedo, por absurdo que parezca. Por estar cerca estando lejos, por vuestras siempre sosegadas palabras, pero sobre todo, por vuestro amor infinito. A Paula e Irene por tanta dulzura, alegría incontenible y por mantener siempre despiertas la

fantasía y la sorpresa. ¡Y por tantos dibujos que viajaron a Barcelona! A mi hermano Luismi, por ser eso mismo. Al calor de la mancha albaceteña y a todos mis tíos y primos por haber construido un espacio de amor y apoyo incondicional que va más allá de filogenias y consanguinidades. ¡Todos somos contingentes, pero vosotros sois necesarios! A mis abuelos que me enseñaron el valor de la memoria y la humildad. Gracias por una vida de esfuerzo y amor. Por ser tantas veces más abiertos, flexibles y respetuosos que yo misma y mis coetáneos. En muchos días de agotamiento vuestro recuerdo fue el motor que me permitió continuar.

A Marco. Por ser la otra cara de esta tesis. Ésa que no se ve pero hace posible ésta otra que aquí aparece. Gracias por acompañarme siempre desde el amor y la camaradería, sin importar las distancias ni los viajes. Por tantas horas de escucha, debates infinitos, risas y paciencia. Por Barcelona y Madrid, y la genial mezcla de las dos. Por comprender lo incomprensible y por los proyectos imposibles. ¡Gracias Marco por las fresas y la filosofía!

Esta tesis tiene, sin embargo, una persona a la que quisiera agradecer por encima de todas. Su recuerdo me ha acompañado cada día durante estos años, animándome a luchar incluso cuando todas las batallas parecían perdidas. Gracias a mi padre, Enrique, por enseñarme a no rendirme nunca y a amar a través del espacio y el tiempo. Me aplico la colleja cariñosa que sé que me darías si pudieras estar leyendo estas páginas. Como tú solías decirme, *te quiero inmensamente*.

Alicia Gallego Jiménez, Barcelona, Septiembre 2016

This research was supported by the PhD fellowship "Formación de Profesorado Universitario" (FPU-AP2010-2469) funded by the "Secretaría de Estado de Educación, Formación Profesional y Universidades" of the "Ministerio de Educación, Cultura y Deporte de España".

ABSTRACT

microRNAs are small non-coding RNAs with crucial roles in gene regulation and whose contribution to animal evolution has been largely demonstrated. In this thesis we explored some functional consequences of microRNA sequence variation and their possible effects in primate evolution. We first evaluated microRNA nucleotide variation at the genomic level in great apes. Taking advantage of recently published whole-genome sequencing data from 82 individuals including oragutans, gorillas, bonobos, chimpanzees and humans, we analyzed microRNA sequence conservation patterns, both among and within populations. We observed that the entire microRNA mature region was significantly conserved, suggesting its central role for the microRNA regulatory function. We additionally observed that more conserved microRNAs tend to be older, duplicated, clustered, highly associated with disease and show higher expression levels. Further functional analyses of mir-299, mir-503, mir-508 and mir-541 revealed that lineage-specific changes in the microRNA mature sequences and/or in the length of the precursor molecules altered their expression levels and redirected the spectrum of target genes and regulatory networks, some of them linked to neuronal functions. We secondly investigated microRNA sequence variation generated by RNA editing. Focusing on mir-376a1 we studied the RNA editing patterns among different primate individuals including human placenta and macaque, gorilla, chimpanzee and human brain cortex samples. Although mir-376a1 editing showed high inter-individual variation, it was more frequently detected in brain than in placenta and in one particular site. This highly edited site altered the spectrum of target genes and conferred the highest stability to the hairpin molecule, revealing the important contribution of RNA editing to microRNA processing and function. In summary, we provide evidence on how DNA and RNA nucleotide changes may drive microRNA diversification and redefine new regulatory functions, which could have importantly contributed to primate phenotypic diversification processes and to the recent evolution of our species.

RESUMEN

Los microARNs son ARNs de pequeño tamaño no codificantes para proteína que juegan un papel crucial en la regulación génica y la evolución. En esta tesis exploramos algunas consecuencias funcionales asociadas a la variación en las secuencias de los microARNs y sus posibles implicaciones evolutivas en primates. En primer lugar evaluamos la variación nucleotídica de los microARNs a nivel del genoma en los grandes simios. Aprovechando datos de secuenciación de genomas completos de 82 orangutanes, gorilas, bonobos, chimpancés y humanos, analizamos los patrones de conservación de las secuencias de microARNs, entre y dentro de las poblaciones. La región madura del microARN apareció significativamente conservada, sugiriendo su papel central en la función reguladora del microARN. Los microARNs más conservados mostraron ser más antiguos, estar duplicados, agrupados, asociados a enfermedad y más expresados. Análisis funcionales en mir-299, mir-503, mir-508 y mir-541 revelaron que cambios específicos de linaje en las secuencias maduras y/o en la longitud de las moléculas precursoras alteraron sus niveles de expresión y redefinieron el conjunto de genes dianas y redes génicas reguladas, implicadas algunas en funciones neuronales. En segundo lugar, investigamos la variación de los microARNs generada por la edición del ARN. Estudiamos la edición de mir-376a1 en diferentes individuos de primates incluyendo muestras de placenta humana y córtex cerebral de macaco, gorila, chimpancé y humano; la cual mostró gran variación entre individuos y se detectó en mayor frecuencia en cerebro que en placenta. Una posición concreta del microARN apareció altamente editada alterando el conjunto de genes diana y confiriendo mayor estabilidad a la molécula precursora, indicando la importancia de la edición del ARN en la función de los microARNs. En resumen, nuestro trabajo demuestra cómo los cambios nucleotídicos en el ADN o ARN podrían conducir a la diversificación de los microARNs y definir nuevas funciones reguladoras, las cuales podrían haber contribuido en los procesos de diversificación fenotípica en primates y en la evolución reciente de nuestra especie.

PREFACE

Natural variation and how to start

Humans have always been a curious type of animal, interested in - even obsessed by - understanding how the world around them works. To understand how things work you need to start by observing. At a glance, what you can probably see is a world full of things with different sizes, appearances and behaviours. Thus, it could be reasonable to consider "variation" as the main attribute of the surrounding world you perceive. However, in a second step, in the process of understanding this world of variation, you will probably have to classify and simplify all its complexity. To properly classify many different elements in, for instance, two groups, you have to identify the main features that distinguish both groups, as well as what unifies everything within each group. Imagine these groups as boxes. Once you form two boxes of things, you realize that both could be further classified, as more and more common and singular features start appearing. If you are very resolute, you may end up with numerous boxes containing boxes with sub-boxes. The number of boxes may be huge, but you should not have any problems keeping them organized as long as you can describe the features you employed for your classification.

Unbelievably, one day you go back to your boxes and find that all your elements have completely changed and sadly, you cannot recognize their initial features anymore. It is very frustrating but, since you are perseverant, you create a new classification using new boxes. However, this time you assign a number to each element. Imagine that you number 1000 different elements. After a while you carefully observe every group and, as expected, all the elements exhibit new features but surprisingly, all of them show combinations of two numbers such as 0926-0032. Some elements also seem to have exchanged their boxes. After several viewings over time you finally reach three important conclusions. First, your elements are constantly moving and exchanging their boxes. Second, your elements are somehow interconnected as they seem to be combinatorial results of previous elements. And third, these combinations produce new features since new characteristics emerge as time passes. Thus, it

seems your undertaking to make groups was useless because whenever you organize variation, it always reappears differently.

You then run far away without looking back, leaving all your stuff and boxes in a mess. Once you overcome the trauma and you come back, you observe all your fidgety elements again from afar. Then, the "variation" you easily recognized at the very beginning suddenly appears connected with this constant "change" from old elements to new elements. Exhausted, you realize that if you want to understand the world you are observing, you will probably have to understand what is happening behind its elements. This means understanding how the elements interact, how they move, how long they live, how fast they combine, why some features emerge, why others do not. You will probably have to observe all these elements again, one by one and collectively, as well as manipulating and measuring them, focusing on their details and broadening your view to larger scales. You will probably have to talk with others who have also tried to understand the same world, but perhaps from different perspectives. And finally, the only thing you can accept is that this process of understanding the world around you will take a long time. Probably more time than your modest life's length, but the process will be amazing because it will be worked through by many people, from one generation to another. People who will be learning together while that world is continuously changing.

Therefore, please take these chapters as a small, surely minimal, contribution to this huge and universal task of understanding which mechanisms govern this variation we observe everyday around us. With this purpose in mind, the following pages attempt to delve into the molecular scales that have been ceaselessly working behind living beings for a very long time, experiencing change in themselves and contributing to the processes of change - perhaps only a few or perhaps many - on larger scales. Thus, from here on, let us discuss small (micro) things but with big things (individuals, populations and species) in mind.

TABLE OF CONTENTS

I. INTRODUCTION

1. The masters of regulation: the non-coding genome

The identification of the double-stranded helical structure of the DNA molecule (Watson and Crick 1953), undoubtedly thanks to the contribution of previous studies based on X-ray diffraction images (Franklin and Gosling 1953; Wilkins et al. 1953), permitted the identification of the molecular bases implicated in the transmission of biological information from one individual to another. From that moment, deciphering the genetic code became one of the main motivations in biology. The establishment of the *central dogma* of molecular biology (Crick 1958; Crick 1970) allowed for the linking of the genetic information flow from DNA through RNA to proteins. Although, this conception did not explicitly reject other possible interactions among DNA, RNA and proteins; in fact, at that moment and for a long period of time, RNA was considered as a temporary template in the protein synthesis (**Figure 1**).

In the following years, the roles of different types of RNA molecules such as ribosomal (Palade 1955), transfer (Hoagland et al. 1958) and messenger (Brenner et al. 1961) RNAs were identified, demonstrating that genetic expression is strongly regulated and that not all RNAs are translated into polypeptide amino acid sequences. Although these newly identified RNAs had functions associated with the protein synthesis process and they did not essentially alter the *central dogma*, several hypotheses started to arise, suggesting the role of specific RNA molecules as direct regulators of transcription (Jacob and Monod 1961; Britten and Davidson 1969). Shortly after, the discovery of introns (Williamson 1977) indicated that a significant proportion of the transcribed genome was not translated in eukaryotes, but they probably had important functions, most of which are yet to be identified. Since then, increasing studies have demonstrated the existence of different types of RNAs, called non-protein-coding RNAs (npcRNAs) or non-coding RNAs (ncRNAs); showing different lengths, locations and expression patterns, and for which non-associated protein products could be described (Morris and Mattick 2014).

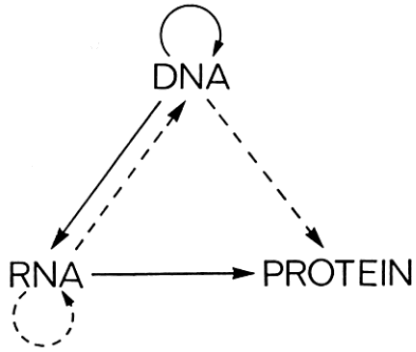


Figure 1. The central dogma of molecular biology. Flow of genetic information represented by Crick in the central dogma (first proposed in 1958 and reviewed in 1970). Solid arrows represent regular genetic information transfers, while dotted arrows represent unlikely transfers (adapted from Crick 1970).

In the past few years, the appearance of the *-omic* data, mostly generated thanks to widespread access to high-throughput sequencing technologies, has made it possible to draw a complex picture of the transcriptional architecture in eukaryotes. Recently, two main projects developed by international consortia, the Encyclopedia of DNA Elements (ENCODE) and the Functional Annotation of the Mammalian Genome (FANTOM), have revealed that three quarters of the genome are transcribed in humans and mice, mostly with non-coding capacity (Okazaki et al. 2002; Djebali et al. 2012). The human genome seems to be highly overlapping and dynamic with numerous coding and non-coding transcripts that are alternatively spliced. Most genomic regions are sense and antisense transcribed and, indeed, the majority of nucleotides are associated with, at least, one primary transcript (Djebali et al. 2012).

The latest estimates have shown that humans express around 20,000 protein-coding genes (Hattori 2005) and that this number does not significantly differ among vertebrates, and even around half of them are shared among evolutionary distal eukaryotes (Taft et al. 2007). In

fact, mammals encode less protein-coding genes than certain plants (Haas et al. 2005; International Rice Genome Sequencing Project 2005) and protists (Aury et al. 2006). Thus, it is currently accepted that the genetic information underlying phenotypic differences does not only lie in the protein-coding content of the genome, but non-protein-coding genomic regions could also have an important role, being involved, among others, in regulating the expression of these proteins (Kleinjan and Van Heyningen 2005; Djebali et al. 2012; Morris and Mattick 2014).

Moreover, in the total genomic content, the portion that is non-coding genome appears to substantially vary among taxa. Although *biological complexity* is a very diffuse complex trait (McShea 1996; Wideman 2016), generally it is naively defined as the number of different cell types present in an organism and their connectivity. Under this assumption, a general pattern has been observed in which "more complex organisms" present higher proportions of non-coding sequences in their genomes (Mattick 2007; Taft et al. 2007). In humans, for instance, it is estimated that only 1.1% of the genome is codifying for protein sequences, a proportion that is widely maintained across mammals (Taft et al. 2007; Djebali et al. 2012).

Among the non-coding genome many different potentially functional elements can be considered, such as ncRNA genes, untranslated, intronic and intergenic regions, as well as gene expression regulatory factors acting in *cis*- (DNA regions that mediate the transcription of nearby genes as promoters or enhancers) and in *trans*- (transcribed regions, which diffused products bind and regulate the expression of distal genetic sequences such as microRNAs or long ncRNAs). Also, a huge proportion of the genome (i.e. around 50% in the human genome) is constituted by repetitive sequences of different lengths (Treangen and Salzberg 2012; Palazzo and Gregory 2014).

The roles and origins of the ncRNAS are still under extensive debate. It has been shown that in some particular cases, they can perform catalytic activities and act as protein cofactors (Teixeira et al. 2004; Salehi-Ashtiani et al. 2006; Morris and Mattick 2014). Nevertheless,

the vast majority of ncRNAs seem to be involved in gene regulation, including their participation in the regulation of chromatin structure, transcription and translation, splicing, RNA modification, RNA stability and cellular signalling processes (Mattick 2007) (**Figure 2**). Additionally, ncRNAs show specific intra-cellular localizations (Liao et al. 2010; Djebali et al. 2012; Cabili et al. 2015) and restrictive patterns of expression across tissues (Melé et al. 2015) and development (Amaral and Mattick 2008; Sauvageau et al. 2013), which suggest the participation of ncRNAs in tightly regulated processes.

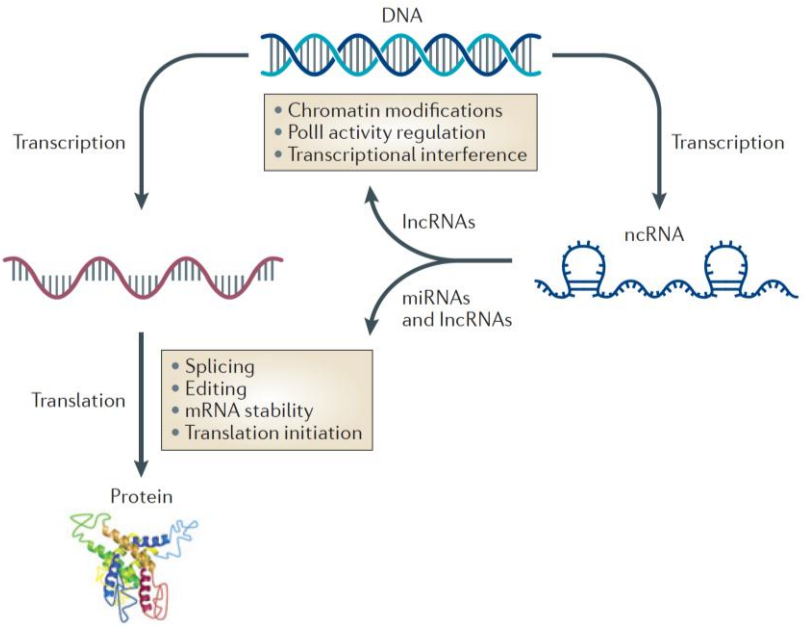


Figure 2. The central dogma in the context of regulatory non-coding RNAs. Genetic information flow including the notion of non-coding RNAs and their regulatory roles in gene expression at both, translational and transcriptional levels (adapted from Wahlestedt 2013).

Since the identification of ribosomal and transfer RNAs (Palade 1955; Hoagland et al. 1958), many different types of ncRNA molecules have been described. Classification of RNAs has been largely discussed as it must deal with different structural and evolutionary properties such as size, location, conservation and function. Assuming that none of these categorizations completely cover the complex ncRNA landscape, the most common one separates ncRNAs based on their length into >200 nucleotides (long ncRNAs) and <200 nucleotides (small or short ncRNAs) (Djebali et al. 2012; St.Laurent et al. 2015), which is the criterion that we have followed throughout this thesis (**Table 1**).

Table 1. Non-protein-coding RNAs. Main classes and known functions of long and small ncRNAs.

Non-coding RNA class	**Size**	**Function**
Long non-coding RNAs (lncRNAs)	≥300 to thousands nt	From structural mechanisms to gene regulation by epigenetic modifications
Small non-coding RNAs (small ncRNAs)	~20–300 nt	From RNA modification to genomic imprinting in eukaryotic cells
PIWI-interacting RNAs (piRNAs)	26–31 nt	Transposon silencing in the genome and regulation of DNA methylation affecting gene expression. Principally restricted to the germline
small interfering RNAs (siRNAs)	20–25 nt	Silencing of messenger RNAs; establishment of silenced chromatin (heterochromatin)
microRNAs (miRNAs)	~22 nt	Post-transcriptional regulation of hundreds to thousands of messenger RNAs by gene silencing

1.1. Long non-coding RNAs

Long ncRNAs (lncRNAs) constitute a heterogeneous group of RNAs and, although the first lncRNAs were discovered in the nineties (Brown et al. 1992; Lee et al. 1999), the majority that are currently known were found in the past decade. Indeed, in recent years, the ENCODE project has identified around 10,000 new lncRNAs in the human genome (Derrien et al. 2012). Extensive characterization analyses covered by ENCODE and subsequent projects demonstrated that most lncRNAs show tissue-specific expression patterns (Derrien et al. 2012; Melé et al. 2015), and many of them (40% of the GENCODE annotated lncRNAs) are located within protein-coding loci and co-expressed with their host messenger RNAs (mRNAs), suggesting regulatory interactions between them (Derrien et al. 2012). Also, long ncRNAs show significant conservation among species and, although the majority of them have non-coding capacity, they are usually constituted by an exonic-intronic structure (most of them formed by only two exons), being spliced or alternatively spliced in the same way as protein-coding genes (Ørom et al. 2010; Derrien et al. 2012).

Despite the fact that the specific functions of the vast majority of lncRNAs remain largely unknown, their great abundance, evolutionary constraints and presence in many and variable processes, all suggest that they play crucial roles in gene regulation. An increasing number of studies have reported multiple functions associated with lncRNAs. Most of them have shown the involvement of lncRNAs in the silencing of gene expression. This seems to be the case of *Xist*, the first lncRNA identified, and its antisense transcript, *Tsix*, both implicated in the X chromosome inactivation and dosage compensation in mammalian females (Brockdorff et al. 1992; Brown et al. 1992; Lee et al. 1999).

Also, lncRNAs may act as enhancers and activate the expression of neighbouring protein-coding genes, as shown for a set of ncRNAs in a study included in the ENCODE project (Ørom et al. 2010). In this

work, a fraction of lncRNAs were depleted using specific interference RNAs, which produced a decrease in the expression of closed protein-coding genes that had important functions in development and differentiation. One of these lncRNAs was *ncRNA-a3,* which activated the expression of *SCL/TAL1*, a gene involved in hematopoiesis regulation (Ørom et al. 2010). The precise mechanisms behind the enhancer-like functions of lncRNAs are still not well understood, and could be based on the direct binding of ncRNAs over genomic protein-coding regions, or could be due to wider *trans*-mediated interactions such as the recruitment of specific factors required for gene transcription (Ørom et al. 2010).

Regarding their regulatory capabilities, lncRNAs have been associated with complex diseases like cancer. Most cancer-expressed lncRNAs show specificity for one or few tumour types and/or aberrant expression levels as is the case of *BRAFP1*, which overexpression correlates with the up-regulation of its coding homologue *BRAF* and the promotion of diffuse large B cell lymphomas (Karreth et al. 2015; Schmitz et al. 2016); or *MALAT1*, a lncRNA that can induce the expression of several genes implicated in lung cancer metastasis (Gutschner et al. 2013).

Moreover, some lncRNAs are able to regulate chromatin organization through histone modifications and DNA methylation, to sequester and repress specific transcripts, and act as a source of small ncRNA molecules when processed (Schmitz et al. 2016). Interestingly, many of them may affect different regulatory pathways at the same time, increasing their overall impact on the whole gene regulatory system. One well-studied example is the antisense *HOX* gene transcript *HOTAIR*, a lncRNA involved in cancer proliferation and metastasis (Gupta et al. 2010), and in chromatin modification and gene silencing of the homeotic gene cluster, *HOXD* (Rinn et al. 2007; Schmitz et al. 2016).

Even though we have only mentioned a few examples, long ncRNAs represent an extended and heterogeneous universe of genetic and genomic regulators with visible roles in development, evolution and

disease, whose origins, properties and functions are only just beginning to be understood.

1.2. Small non-coding RNAs

The other group of ncRNAs encompasses an equally complex universe formed by the small RNA fraction. Again, many groups of small ncRNAs have been described based on their evolutionary history, location, structure and ways of action. Increasing studies have reported the existence of different types of small ncRNAs such as PIWI-interacting RNAs (piRNAs), small interfering RNAs (siRNAs), microRNAs (miRNAs), small nucleolar miRNAs (snoRNAs), transcription initiation RNAs (tiRNAs) and splice site RNAs (spliRNAs), among others (Morris and Mattick 2014) (**Table 1**). Even though ncRNA categorizations are not yet fully established, the existence of the first three groups has been widely accepted. These ncRNAs have been studied in more detail and are described below.

PIWI-interacting RNAs (piRNAs) are ~26-31 nucleotides (nt) RNAs present in metazoans that interact with a class of Argonaute (Ago) proteins, PIWI proteins, forming the piRISC complex that has been described as being able to repress circulating transposable elements (Morris and Mattick 2014). The silencing of transposons mediated by piRNAs has been shown to be required in embryonic development and during gametogenesis in the germ cell lines (Grimson et al. 2008; Ghildiyal and Zamore 2009; Costa 2010). Recent studies have also demonstrated the role of piRNAs in animal brain functions by regulating DNA methylation patterns in certain target genes such as *CREB2*, involved in the memory process (Ishizu et al. 2012).

On the other hand, small interfering RNAs (siRNAs) are ~20-25 nt long present in animals and plants that derive from processed products of endogenous or exogenous double-stranded RNAs (dsRNAs). By perfect sequence complementarity, siRNAs can recognize and degrade their target mRNAs through Ago-mediated cleavage (Amaral and Mattick 2008; Costa 2010; Morris and Mattick 2014). Regarding the fact that siRNAs fully repress their target genes

by degradation, it has been considered that siRNAs could act as protection mechanisms against the invasion of external genetic material, such as virus infections, along with being commonly used as knockdown tools in experimental and therapeutic biology (Ghildiyal and Zamore 2009; Ramachandran and Ignacimuthu 2013).

Finally, microRNAs (miRNAs) are ~22 nt RNAs present in metazoans, plants and viruses (although functional and structural variations exist among these groups) that act in very similar ways to siRNAs, although they are processed from endogenous dsRNAs and present partial sequence complementarity with their target genes. miRNAs can block the expression of hundreds to thousands of target genes by translational repression or molecular degradation. Thus, rather than avoiding the introgression of external material, miRNAs would post-transcriptionally regulate the expression of endogenous genes, being involved in almost all known biological processes, from development and cell differentiation, to evolution and disease (Filipowicz et al. 2008; Morris and Mattick 2014). miRNAs are the main object of study in this thesis, thus an extensive description and characterization of these molecules can be found from section number 2 onwards.

In general, small ncRNAs show high levels of sequence and secondary structure conservation, which suggest an early origin of these regulators and their potentially crucial roles at the very basal functions of living beings. Also, many of them are species-specific and show tissue-specific expression, which may indicate their essential contribution in specific and highly regulated processes. The most studied functions of small ncRNAs are mainly based on the destabilization of longer transcripts, recognized in a sequence-specific manner, by the coupled effect of protein complexes. This mechanism is usually known as RNA interference (RNAi), an extended resource of post-transcriptional repression present in eukaryotes and prokaryotes (Shabalina and Koonin 2008).

2. The microRNA world

The main focus of this thesis is the study of miRNA variation at different levels and the possible implications of this variation on their regulatory functions. As briefly outlined, although the non-coding transcriptome represents a large and miscellaneous universe formed by multiple elements, whose functions and properties are still poorly understood, miRNAs are one of the most well-defined groups among the ncRNA repertoire. Despite only having been studied for as little as two decades, their common features such as their structure, biogenesis and ways of action, have permitted the discovery of evidence for new miRNAs in a short period of time. The first described miRNA, *lin-4*, was discovered in 1993 (Lee 1993) and it was not until almost ten years later that the second, *let-7,* was found (Reinhart et al. 2000). Both of them were shown to regulate the timing of larval development in *Caenorhabditis elegans*, and to present very similar structures consisting of 21-22 nt RNA transcripts capable of binding and regulating specific mRNAs. These findings motivated the search for new miRNAs and soon after, several studies reported the existence of many more examples across different species, pointing to miRNAs as a new and extended class of ncRNAs with crucial roles in gene regulation (Lagos-Quintana et al. 2001; Lau et al. 2001). Currently, more than 35,500 miRNAs belonging to 223 species are compiled in miRBase, the largest public miRNA repository (Kozomara and Griffiths-Jones 2014). These small ncRNAs molecules are one of the masters of post-transcriptional regulation and their biogenesis, function and strong regulation have been intensively studied (Krol et al. 2010). Nevertheless, there are still many questions regarding their features. In particular, it is largely unknown how genetic and genomic variation could have affected miRNA function, as well as how miRNAs may have contributed to shaping phenotypic traits during evolution. Trying to shed light on these questions has been the principal motivation of this study.

2.1. microRNA biogenesis

In order to understand how miRNAs evolve and operate, it is essential to know how they are processed, what kind of secondary structures they form and which genetic and genomic properties determine their function. miRNAs may be located either in intergenic regions or within introns of other genes, thus they can be transcribed through their own promoters or through the transcription machinery of their host genes (Lau et al. 2001; Baskerville and Bartel 2005; Liang et al. 2007; Campo-Paysaa et al. 2011; Meunier et al. 2013). Among metazoans, canonical miRNAs are transcribed in the nucleus by RNA polymerases II or III, in a primary transcript called pri-miRNA. This pri-miRNA contains an imperfect palindromic sequence that creates a secondary hairpin stem-loop structure constituted by a ~33 nt stem with a terminal loop and flanked by single-stranded RNA (ssRNA) sequences (Bartel 2004; Lee et al. 2004). Adjacent or clustered miRNAs are usually transcribed together in a polycistronic pri-miRNA containing different hairpins, which can reach several kilobases (Olena and Patton 2010). Most pri-miRNAs are cleaved by a complex formed by the ribonuclease III (RNase III), Drosha, and its cofactor, the dsRNA binding protein DGCR8 (DiGeorge syndrome Critical Region 8, also known as Pasha in *Drosophila melanogaster* and *C. elegans*). The Drosha-DGCR8 complex cleaves the pri-miRNA at the flanking ssRNA sequences, around 11 nt from the stem-loop base, giving a ~70 nt precursor miRNA or pre-miRNA that contains a 5' phosphate group and a 3' 2 nt overhang (Rana 2007; Olena and Patton 2010). Some miRNAs, termed as mirtrons, are located within intronic sequences, between 5' and 3' splice sites, and are directly processed by the Spliceosome instead of Drosha. These kinds of miRNAs were first identified in *D. melanogaster* and *C. elegans*, and soon after in different mammal species (Berezikov et al. 2007; Okamura et al. 2007; Ruby et al. 2007). Spliced mirtrons form hairpins and show the same structure as pre-miRNAs, being processed in the same way (**Figure 3**).

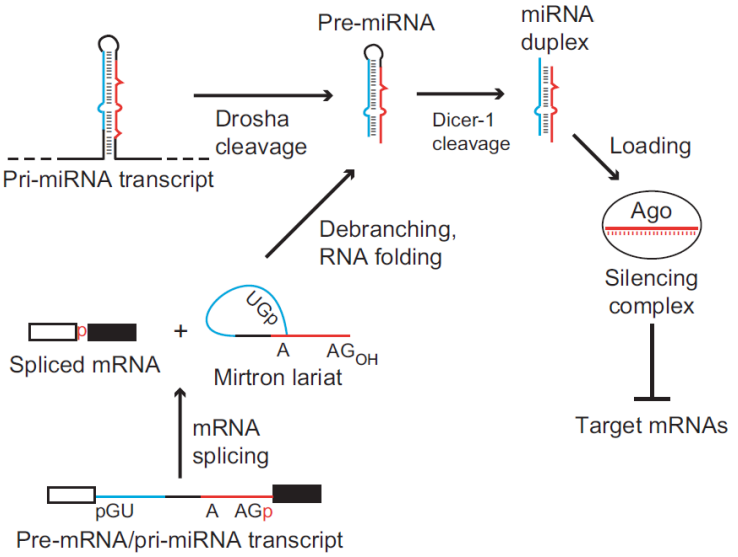


Figure 3. Canonical miRNA and mirtron biogenesis. Canonical miRNAs are transcribed as pri-miRNAs and processed by Drosha, while mirtrons originate from spliced introns. Both mechanisms generate pre-miRNA hairpins that are exported into the cytoplasm by Exportin-5, and cleaved by Dicer, giving a miRNA duplex. One of the mature strands is further incorporated into the RISC complex, which guides the miRNA to its target mRNAs, silencing their expression (adapted from Ruby et al. 2007).

The pre-miRNA is transported by Exportin-5 and the Ran-GTP cofactor into the cytoplasm (Yi et al. 2003), where it is cleaved close to the terminal loop by another RNase III, Dicer, giving a miRNA duplex constituted by the -5p and the -3p mature miRNA strands. One of these strands is further incorporated into the RISC (RNA induced silencing complex), which contains Ago, Dicer and several RNA-binding proteins such as TRBP and PACT in humans. The RISC complex binds to the mature miRNA and forms the miRISC complex that guides the miRNA mature strand to its target mRNAs, suppressing their expression by translational repression or mRNA degradation (Faller and Guo 2008; Krol et al. 2010). Even though the

exact mechanism of action is not fully understood, target gene repression is based on the partial sequence complementarity between the miRNA and the target mRNA, although the miRNA *seed region*, located between nt 2-8 of the 5' extreme, must show a perfect match with the target site on the mRNA, with only a few exceptions in which certain G:U wobbles are permitted (Lewis et al. 2003; Grimson et al. 2007a; Friedman et al. 2009) (**Figure 4**).

Plant miRNAs show a very similar structure, function and biogenesis to animal miRNAs, although some important differences exist between them. Plant miRNAs are entirely processed in the nucleus and are larger than animal miRNAs (20-24 nt long). Additionally, plant miRNAs show nearly perfect base-pairing complementarity with their target genes (most of which are cleaved and degraded); thus, they are expected to have a more limited number of target genes compared with animal miRNAs. Some authors suggest that resemblance between animal and plant miRNAs is the result of convergent processes with different origins, while other studies maintain that both types of miRNAs diverged from a common eukaryotic ancestor (Voinnet 2009; Axtell et al. 2011). Other groups such as certain viruses and protist

---

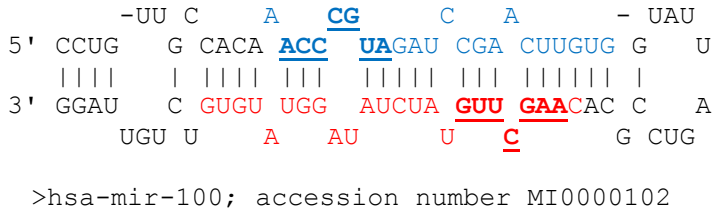


Figure 4. The miRNA secondary structure. This example of human mir-100 shows a pre-miRNA hairpin of 80 nt with an imperfect stem-loop constituted by the further-processed mature strands, miR-100-5p (blue) and miR-100-3p (red). Seed regions within each mature strand are shown in bold-underlined letters (miRBase, release 21). Notice that the lower case "r" (mir-) refers to the precursor molecule stage, while the upper case "R" (miR-) refers to the mature molecule.

also show miRNAs, although their evolutionary relationships among plant and animal miRNAs are yet to be determined (Ghosh et al. 2009; Cullen 2010; Tarver et al. 2012). Further comparative analyses are still needed in order to clarify the evolutionary relationships among miRNAs in different phyla.

2.2. microRNA ways of action

It is estimated that most human protein-coding genes show miRNA target sites and are susceptible to their regulation (Lewis et al. 2005; Friedman et al. 2009). Considering that a single mRNA transcript may contain many target sites for numerous miRNAs and that, at the same time, one particular miRNA can regulate hundreds and even thousands of target mRNAs (Selbach et al. 2008; Schnall-Levin et al. 2011), it is easy to imagine how miRNA-mediated regulation can generate complex networks of gene interactions.

It has been widely accepted that miRNAs regulate their target genes by binding the mRNA 3' untranslated region (UTR) (Grimson et al. 2007a). The 3' UTR is probably the most effective region in miRNA-mediated gene repression (Bartel 2009); however, it is currently known that miRNAs can target almost all sites along their target genes such as protein coding regions (Rigoutsos 2009; Schnall-Levin et al. 2011), 5' UTRs (Zhou and Rigoutsos 2014), intronic and intergenic sequences (Zisoulis et al. 2010), as well as ncRNAs such as pseudogenes, lncRNAs and circular RNAs (Tay et al. 2011; Hansen et al. 2013; Kosik 2013; Memczak et al. 2013) (**Figure 5**). Importantly, it has recently been described that not all miRNA targeted sites mediate repression of the mRNA despite binding a miRNA (Agarwal et al. 2015). Moreover, target gene repression does not only depend on the co-expression of both the miRNA and its target gene. Adding another layer of complexity, repression varies in intensity depending on several aspects. For instance, the number of target sites for a particular miRNA in a gene, as well as the cooperative/competitive effects among several miRNAs over one mRNA may differentially affect the intensity of the gene repression. Also, the level of complementarity

along the miRNA-mRNA duplex may determine the stability of the union, while the expression levels and the number of copies for both, the target gene and the miRNA can define the probability of said union. Finally, target gene repression may also be influenced by the presence of other binding competitors or *off-targets* such as protein-coding transcripts, circular RNAs or ncRNAs (Doench et al. 2003; Grimson et al. 2007a; Garcia et al. 2011; Hausser and Zavolan 2014) (**Figure 5**).

Post-transcriptional repression modulated by miRNAs is thus a dynamic and complex process in which protein synthesis is ultimately disrupted by different mechanisms, whose exact way of action is still unclear. It is currently known that miRNAs can lead to mRNA poly-A tail deadenylation and its subsequent degradation through the binding of the miRISC complex into the mRNA 3' UTR. In addition, miRISC

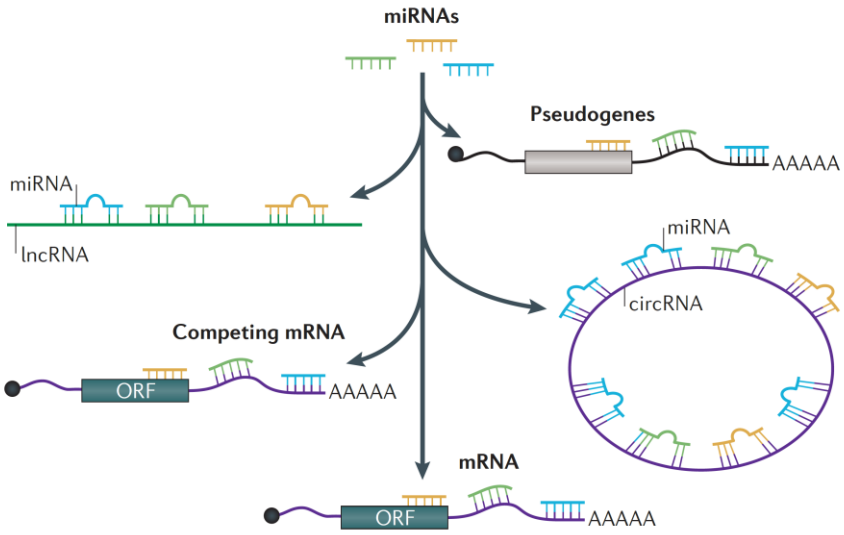


Figure 5. Potential miRNA target genes and target sites. Pseudogenes, long non-coding RNAs (lncRNAs), circular RNAs (circRNAs) and messenger (mRNAs) can compete for miRNA binding. Also, miRNAs can interact at different sites along their target genes (adapted from Hausser & Zavolan 2014).

can compete for the 5'-cap region of the mRNA with the translational machinery preventing the action of the latter. In both cases, the repressed mRNAs are moved to cellular processing bodies (p-bodies) where they are degraded or temporarily stored. Additionally, miRNAs may repress translational elongation by blocking the transition of the ribosomes along the mRNA, and can degrade the synthesized polypeptides by the recruitment of proteases nearby (Filipowicz et al. 2008; Krol et al. 2010).

On the other hand, the levels of base pairing affinity between the miRNA and the target mRNA seem to be the main determinant in the balance between mRNA translational repression and degradation (Hutvágner and Zamore 2002). It has been largely proven that the seed is the most important region for target gene recognition since it shows perfect sequence complementarity with its target sites. Therefore, it is expected that mutations in the seed dramatically change the miRNA function and, in fact, few changes seem to be permitted in this region that shows high levels of sequence conservation (Lewis et al. 2005; Saunders et al. 2007; Selbach et al. 2008). Even though they also seem to play significant roles, the relevance of the remaining positions along the mature region is still not well understood. Certain sites of the mature region, out of the seed, seem to be crucial for target gene interaction, as evidenced for nt 13-16 (Grimson et al. 2007a), which also show significant levels of conservation across different metazoans (Wheeler et al. 2009). In a similar way, purifying selection seems to have globally constrained the first 14 nt of the mature miRNA molecule in humans (Quach et al. 2009). Hence, apart from the seed, other positions along the mature miRNA might be decisive for the miRNA function by shaping the type of post-transcriptional repression (translational suppression or mRNA degradation), by compensating seed mismatches, or by other mechanisms that are yet to identified (Grimson et al. 2007a; Hausser and Zavolan 2014).

2.3. microRNA regulatory functions

The contribution of miRNAs in almost all biological functions and, in particular, in tightly regulated processes such us development, differentiation, proliferation and apoptosis has been widely evidenced (Drakaki and Iliopoulos 2009). However, addressing the regulatory functions associated with one particular miRNA is not an easy task considering that these regulators can directly and indirectly affect the expression of many different genes, with a wide range of intensities. Several approaches including *in silico* predictions, and *in vivo* and *in vitro* experiments are regularly used with the aim of determining specific miRNA functions and, usually, the combination of different techniques has permitted the development of more comprehensive interpretations of miRNA function.

2.3.1. Approaches to the analysis of microRNA functions

One of the most logical steps to start the process of understanding miRNA functions is to identify where and when one miRNA (or a set of miRNAs) is expressed. Several methods may permit the study of the miRNA expression patterns such as microarrays, which are used to assess the expression levels of large numbers of miRNAs in a single experiment. This high-throughput method is based on the hybridization by sequence complementarity between the small RNAs present in our sample (previously retrotranscripted and labelled with fluorescent dyes), and the collection of nucleotide sequences or *probes* corresponding to known annotated miRNAs (generally attached in spots to a solid surface). Sequence interactions are quantified by fluorescence emission, thus we can measure the presence and the relative abundance of specific miRNAs in our sample, as well as comparing expression levels among different conditions. These assays are powerful tools for miRNA detection, although they require prior knowledge of the miRNAs present in our studied species (Calin and Croce 2006; Krützfeldt et al. 2006; Yin et al. 2008). In humans, for whom more than 2,500 mature miRNAs have been described, microarrays represent a useful method for miRNA detection and

quantification. Nevertheless, they are less informative in other species, for which less data have been generated (for instance, only 587 mature miRNAs have been described in chimpanzee, one of the most studied species after humans) (Kozomara and Griffiths-Jones 2014). Additionally, microarrays are employed for transcriptome analysis to measure the expression levels of particular genes. In this case, the attached probes represent known gene fragments that hybridize with fluorescent-labelled material from total RNA samples. This approach allows for the quantification of the expression levels of many genes and, once again, to compare the different conditions. Through transcriptome microarray analyses we can, for example, evaluate which genes are expressed and at which levels in the presence or albescence of certain miRNAs. This procedure has enormously contributed to the understanding of miRNA regulatory functions (Lim et al. 2005; Zhao et al. 2014).

RNA sequencing (RNA-seq) based on next generation sequencing (NGS) technologies applied to the small RNA fraction is also commonly used for the detection and quantification of multiple miRNA sequences in parallel. Since these methods do not require a pre-design of complement probes nor sequence hybridization steps, they can directly determine the absolute amounts of any sequence present in a sample regardless of their expression levels and without requiring previous knowledge concerning the region we want to analyze. RNA-seq is particularly informative in the analysis of non-model organisms whose genomic sequences are not well determined; lowly-expressed transcripts that could be underrepresented in microarray platforms; and sequences with genetic variations that would remain masked in hybridization-based approaches (Landgraf et al. 2007; Wang et al. 2009; Zhao et al. 2014).

Analysis of miRNA expression patterns can also be addressed with more low-throughput procedures such as reverse transciption real-time quantitative Polymerase Chain Reaction (RT-qPCR) (Chen et al. 2005; Balcells et al. 2011), Sanger sequencing of cloned fragments (Kawahara 2012), northern blot experiments (Válóczi et al. 2004) and

in situ miRNA detection (Kloosterman et al. 2006). In fact, although these methods can require more time and higher costs compared with microarrays or RNA-seq, they are highly reliable and sensitive, and have been commonly applied for validating specific miRNAs and their expression profiles.

Apart from examining the timing and location of miRNA expression, knowing which genes are directly regulated by miRNAs could give important clues about the processes in which they are involved. *In silico* Target gene predictions have become a crucial step in the analysis of miRNA functions and a prior explorative tool for further experimental target gene validations. In this regard, several software packages such as *PITA* (Kertesz et al. 2007), *Miranda* (Betel et al. 2008) or *TargetScan* (Friedman et al. 2009), which are among the most popular, have been developed over the past few years in order to identify the most probable target genes for a particular miRNA. These programs are essentially based on algorithms that take into account several parameters affecting the miRNA-mRNA interaction such as the level of complementarity between the miRNA seed and the mRNA target site, the number of target sites in a gene for one particular miRNA, the miRNA-mRNA duplex stability, the conservation levels of the target and the miRNA (it is expected that more phylogenetically conserved sequences will play functional roles), as well as the secondary structure of the mRNA and the accessibility of the miRNA to its target site (Grimson et al. 2007a; Garcia et al. 2011; Peterson et al. 2014). All these programs are continuously modified as more information becomes available on the genetic and biochemical bases acting in miRNA-mediated gene suppression.

In order to establish reliable miRNA-mRNA interactions, predicted target genes must be eventually validated. One of the most common methods, the luciferase gene reporter assay, is based on the cellular transfection of a reporter vector construct containing the predicted target site sequence downstream from the luciferase coding region. The pertinent miRNA should be present in the system, thus when it is not naturally expressed in the cell, it can be co-transfected within a

different vector construct or as a synthetic RNA molecule that mimics the studied miRNA sequence (miRNA mimics). miRNA-mediated repression can then be established if the expression of the reporter luciferase gene is reduced in comparison with a control scenario (Krützfeldt et al. 2006).

The most reliable approach to identifying miRNA roles is probably looking at the phenotypic consequences of a mutated, disrupted or altered miRNA expression through *in vivo* or *in vitro* analyses (Bartel 2004). An extensive collection of miRNA knockout and over-expression experiments, together with their associated phenotypes has been generated in the past decade; however, many of them have been related to cancer, immune and cardiovascular diseases, possibly because of a primary biased biomedical interest (Park et al. 2010). Several studies have shown the relevance of miRNAs for normal development and organism survival. One of the most popular examples is the role of lin-4 and let-7 in *C. elegans* timing development. Ectopic expressions of these miRNAs induced larval growth alterations and lethal phenotypes through the repression of several transcription factors and RNA-binding proteins involved in the progression from larval to adult phases (Lee 1993; Reinhart et al. 2000). On the other hand, deletion of miR-1 in flies and mice caused larval and embryonic death with important defects in muscle differentiation and cardiac function, suggesting its role in muscle and cardiac development (Kwon et al. 2005; Zhao et al. 2007). Similarly, genetic ablation of the miR-17/miR-92 cluster produced lung, heart and B cell developmental defects associated with perinatal death in mice; while over-expression of this cluster resulted in lymphocyte proliferative deregulation, autoimmunity and, once again, premature death, revealing the implication of miR-17/miR-92 in mice's normal development and survival (Ventura et al. 2008; Xiao et al. 2008). Phenotypic characterizations have supplied a lot of information about miRNA functions; although, they also have important limitations, such as the use of genetically modified organisms or viable cell lines. Furthermore, even though there are significant examples, it is not easy

to identify a well-defined phenotype when one miRNA is missed, over-expressed or modified. This is mainly because miRNAs participate in complex gene interactions and their functions are generally redundant (Bartel 2004; Williams 2008; Wu et al. 2009; Park et al. 2010).

In summary, addressing which are the main regulatory functions associated with specific miRNAs is a complex process that may be facilitated by the combination of different approaches including target gene predictions, transcriptome analysis and loss-/gain-of-function studies.

2.3.2. microRNAs as buffers and fine-tuners of gene expression

Most miRNAs are evolutionary conserved and show unique tissue, developmental, stage and disease-specific expression, suggesting that they are implicated in establishing the particular features and phenotypes behind these conditions (Liang et al. 2007; Niwa and Slack 2007; Rana 2007). Redundancy of miRNA function is generally extended, thus alterations in one miRNA can normally be supplied by cooperative effects carried out by other miRNAs. Nevertheless, when the miRNA processing machinery is disrupted, dramatic changes arise. Individual knockout experiments of Dicer, DGCR8, Drosha and Ago led to lethality in embryonic mice, indicating that, although alterations in the normal expression of one or several miRNAs probably do not have severe functional consequences for the animal (with the exception of some important miRNAs), the absence of the whole miRNA repertoire is not compatible with normal development and survival (Bartel 2004; Williams 2008; Park et al. 2010).

Different roles have been proposed for miRNA functions depending on their patterns of expression and the type of genes they regulate. Since miRNAs can supposedly suppress the expression of numerous target genes, have redundant functions and simultaneously participate in positive and negative feedbacks, it has been suggested that they act as *fine-tuners* of gene expression (Klein et al. 2010), while also providing *robustness* to gene networks by buffering, canalizing or stabilizing

variation imposed by environmental perturbations (X. Li et al. 2009; Peterson et al. 2009; Herranz and Cohen 2010). The miRNA-mRNA interaction map (miRNAome) produces complex and intricate networks in which there is usually more than one way to get the same result or phenotype (Murray et al. 2010) (**Figure 6**). It is expected that alteration in crucial miRNAs regulating many genes within the network (central nodes) would alter the interactome map and the robustness of the entire system more dramatically than miRNAs acting in the periphery. An extensive study performed on *Drosophila* sensory organs demonstrated the interlocking role of miR-7 in feedback and feedforward genetic loops through the participation of multiple genes.

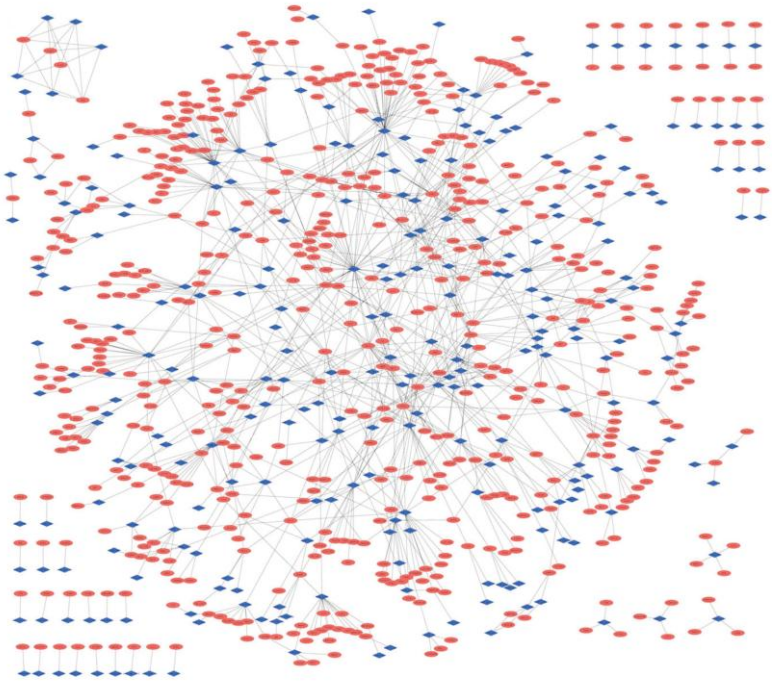


Figure 6. The miRNA-target gene interactome (miRNAome). Interactome of target genes regulated by miRNAs described in the literature. Lines represent interactions among genes (red circles) and miRNAs (blue diamonds). This representation of miRNAome has 270 miRNAs and 581 genes connected by 1165 interactions (adapted from Murray et al. 2010).

External perturbations, such as the depletion of regulated genes or the introduction of temperature fluctuations, revealed the essential contribution of miR-7 in conferring robustness and gene expression stability throughout developmental programs (X. Li et al. 2009). Some authors suggest a distinction between the *buffering* and *tuning* effects of miRNAs, and propose that miRNA *tuners* adjust the target gene expression levels, while miRNA *buffers* reduce gene expression variation (Wu et al. 2009; Zhu et al. 2012).

On the contrary, some other miRNAs seem to suppress several relevant targets and/or to be involved in feedback loops that unidirectionally reinforce the expression of certain genes, acting in these cases as *switchers* of gene networks (Herranz and Cohen 2010; Park et al. 2010). For instance, loss- and gain-of-function studies in mice demonstrated that miR-150 directly regulates *c-Myb* expression, and that this specific interaction is essential for B cell differentiation (Park et al. 2010). In addition, miRNAs could have diverse roles, acting as *buffers* and subsequently as *switchers,* reducing the genetic variance imposed by external perturbations until the expression level thresholds of particular genes are exceeded, and then causing irreversible changes. This seems to be the case of miR-14, involved in development and hormonal responses in *Drosophila*. This miRNA can repress EcR, a positively autoregulated hormonal ecdysone receptor, which at the same time can repress miR-14 expression. Mutual repression of miR-14 and EcR keeps the levels of both factors in balance until the presence of an ecdysone hormone induces a transcription burst of EcR. The increase of EcR expression strongly blocks the production of miR-14, but the pre-existing miRNA can still continue regulating EcR levels for a while. Only significant and constant hormonal induction would produce a permanent decrease in miR-14 and the activation of EcR, which involves profound changes in the gene expression program of the organism (Herranz and Cohen 2010) (**Figure 7**).

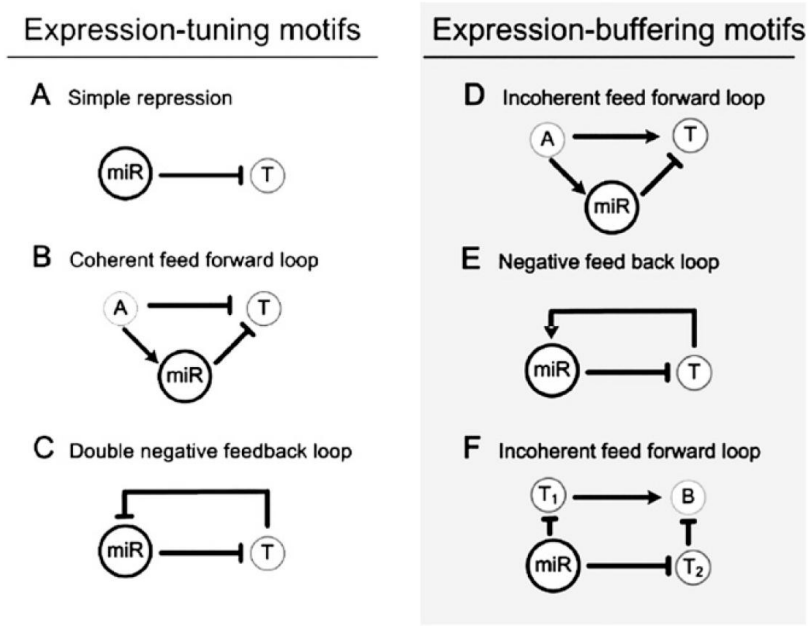


Figure 7. Simple networks involving miRNA function as tuners or buffers of gene expression. T represents miRNA target genes; A and B represent other non-target genes. The tuning networks include: **(A)** simple repression, in which the miRNA reduces the expression of the target; **(B)** coherent feedforward loop, in which the miRNA reinforces the repression of the target; **(C)** double-negative feedback loop, in which the miRNA and the target are mutually repressed. The buffering networks include: **(D)** incoherent feed-forward loop, in which the miRNA reduces the expression variation of the target against the fluctuation imposed by A; **(E)** negative feedback loop, in which the miRNA and the target mutually buffer their expression; **(F)** incoherent feed-forward loop, in which targets T1 and T2 buffer gene B against fluctuations in the miRNA (adapted from Wu et al. 2009).

In this regard, theories based on evidence from the fossil record have hypothesized that miRNAs may have contributed to the canalization of animal development during evolution based on three main observations. First, animal phenotypic variation has decreased over time. This is based on the fact that, although more species and genera have emerged, higher taxonomical groups have become extinct since the Cambrian explosion (Gould 1989). Second, as previously mentioned, miRNA function is able to buffer and reduce external

genetic noise (X. Li et al. 2009; Peterson et al. 2009; Herranz and Cohen 2010). And third, miRNAs have been continuously added to metazoan genomes (see section 4.1. for a more detailed explanation). According to this view, animal phenotypic variation would have been reduced throughout evolutionary time owning to an increase in the genetic expression precision carried out, at least partially, by miRNA-mediated regulation (Peterson et al. 2009).

Other authors have found that genes targeted by miRNAs tend to show more expression variation among human populations than non-target genes and consequently suggest dual roles for miRNAs. For a subset of genes, miRNAs could canalize and reduce their expression variation, or on the contrary, in some other cases miRNAs could promote this variation (Lu and Clark 2012).

miRNAs also play crucial roles in rapid, adaptive responses imposed by external environmental signals. In general, changes at the transcriptional level, such as those triggered by transcription factors, increase reaction times since the whole sophisticated transcription machinery must be initiated. In contrast, post-transcriptional regulation, carried out by miRNAs, can rapidly vary the expression of one particular target gene merely by binding an existing transcript. Moreover, the miRNA rate of biogenesis is faster than that of proteins, and their repression activity can be briefly reverted if the suppressed transcripts are liberated from the miRNA "trappers" and moved to the translational machinery (Hobert 2008; X. Li et al. 2009). Some examples suggest the role of miRNAs as timers of rapid stress responses such as the expression of miR-9, miR-155, and miR-146 in macrophages induced by the nuclear factor NF-kB during inflammatory processes (Leung and Sharp 2010). Interestingly, miRNAs can also mediate during fast neuronal synapses, as shown for miR-134. In the absence of synaptic activity, this miRNA represses the target gene *Limk1*, involved in dendritic spine growth. Under a synaptic stimulation, the presence of neurotrophic factors and the activation of tyrosine kinase signalling pathways would inactivate

miR-134 and promote dendritic spine development by *Limk1* mRNA translation (Schratt et al. 2006).

2.3.3. Specific biological processes involving microRNA regulation

Even though miRNAs have been shown to be expressed in a wide range of tissues and developmental stages, there are specific contexts in which their presence seems to be particularly relevant. miRNAs are significantly abundant in the brain, and many of them show specific expression in this organ, which is also enriched in genes with particularly long 3' UTRs (Meunier et al. 2013). These observations stress the importance of miRNA-mediated regulation in plastic and tightly-regulated systems such as the central nervous system. For instance, a study analyzing approximately 250 small RNA libraries from 26 different organ systems and cell types belonging to humans and rodents showed that miR-9, miR-124, miR-128a and miR-128b were the most highly and specifically expressed miRNAs in all brain regions (38.6% relative cloning frequency), and that, interestingly, the miRNA brain profile changed substantially between the embryonic and adult stages, as well as between tumour diseases with glial or neuronal cellular origins (Landgraf et al. 2007). Neuronal Dicer ablation assays performed on different animal brain regions and cell types were associated with severely compromised neural phenotypes evidencing the essential contribution of the miRNA repertoire to normal brain development and function (Saba and Schratt 2010).

It has also been shown that the contribution of miRNA regulation to animal development is crucial. Species-specific differences in the temporal expression and regulatory functions of miRNAs throughout development act as key factors in determining tissue identities within organisms. Some examples come from previously mentioned heterochronic miRNAs, lin-4 and let-7, which are central regulators in the larval phases of *C. elegans* (Lee 1993; Reinhart et al. 2000) or miR-17/miR-92, an essential miRNA cluster in mammalian development (Ventura et al. 2008; de Pontual et al. 2011). Also, a substantial

number of the early expressed miRNAs are maternally supplied and play central roles in cell fate decisions. Knockdown experiments in combination with supplied miRNAs showed that maternally expressed miR-71 is needed for developmental transitions before blastulation in sea urchins (Song et al. 2012). In this regard, miRNA-mediated regulation seems to be indispensable for stem cell differentiation processes. High-throughput sequencing analyses showed that almost 200 miRNAs were differentially expressed when comparing embryonic stem cells before and after their differentiation into embryoid bodies (Morin et al. 2008).

Overall, the role of miRNAs in gene regulation seems to be complex and diverse, affected by external perturbations and environmental determinants, the relative position of each miRNA within a genetic network, and its timing and local patterns of expression. An existing bias in the study of miRNA function (usually focused on specific and unidirectional miRNA-target interactions observed under constant and unchangeable environmental conditions) has probably hidden the important contribution of miRNAs to establishing *rhizomatic* gene connections and their supposed capacity to generate diverse regulatory effects depending on the biological context. Significant evidence demonstrated that miRNAs are indispensable for development, cell differentiation and cell-cycle progression, as well as for tightly and dynamic processes such as neural function or stress responses (Ambros 2004; Landgraf et al. 2007; Morin et al. 2008; Murray et al. 2010; Meunier et al. 2013).

2.4. The role of microRNAs in disease

Given the contribution of miRNAs to many gene regulatory processes, it is not surprising that alterations in their normal expression have been related to a great spectrum of diseases, such as cardiovascular, rheumatologic, infectious, inflammatory, autoimmune and metabolic diseases, as well as complex diseases like cancer and mental disorders. In some cases, changes in the expression levels of miRNAs can be directly linked with particular diseases, as has been the

case of miR-375, for which increased expression seems to reduce the glucose-induced insulin secretion, related to diabetes mellitus recurrence (Poy et al. 2004); or miR-122, whose inhibition has been associated with the fatty acid metabolism by reducing the plasma cholesterol levels and promoting the accumulation of lipids in liver cells (Esau et al. 2006).

In some other cases, changes in the nucleotide sequences of miRNA genes may alter the expression levels of the mature miRNA molecules or the repertoire of miRNA targeted genes, which could ultimately lead to altered miRNA regulatory functions and disease (Sun et al. 2009). A single nucleotide polymorphism (SNP) in the precursor miR-146a has been reported to decrease almost 2-fold the expression levels of the mature molecule (Jazdzewski et al. 2008; Torruella-Loran et al. 2016), reducing their target gene interactions and contributing to papillary thyroid carcinoma predisposition (Jazdzewski et al. 2008). On the other hand, certain cancers have been associated with the deregulation of several miRNAs. Breast cancer, for instance, has been related to changes in the normal levels of miR-125b, miR-145, miR-21 and miR-155 (Iorio et al. 2005). The role of miRNAs in neurologic disorders and diseases is also understood. Among others, changes in the expression patterns of miRNAs regulating brain and neuron development have been associated with Alzheimer's disease (Hébert et al. 2008) and schizophrenia (Hansen et al. 2007), and nucleotide changes in miRNA regions have been associated with anxiety and panic disorder susceptibility (Muiños-Gimeno et al. 2011). The significant association of specific deregulated miRNAs with several diseases has made it possible to demonstrate the potential of miRNAs as precise and valuable biomarkers in the diagnosis and treatment of many different pathologies (Lu et al. 2005; Esquela-Kerscher and Slack 2006; Mitchell et al. 2008).

Apart from changes in the miRNA sequence and/or deregulation of miRNA expression, changes in the target sites of their regulated genes have been also associated with different disorders. It has been shown that SNP variants rs1982073 inside *TGFB1* (the transforming growth

factor beta 1) and rs1799782 inside *XRCC1* (the X-ray repair complementing defective repair protein) significantly alter the target sites recognized by, respectively, miR-187 and miR-138, and subsequently the expression of these proteins, contributing to the likelihood of breast cancer susceptibility (Nicoloso et al. 2010). Moreover, it has been shown that modifications to the proteins in charge of miRNA processing cause different diseases. Deletion of the human chromosome 22q11.2 has been associated with DiGeorge syndrome, which involves learning disabilities, psychiatric illness and heart and thymus defects, among other symptoms. This region contains the DiGeorge syndrome critical region 8 (*DGCR8*), required for the maturation of primary miRNAs (Landthaler et al. 2004). Also, 27% of tumours with various tissue origins present hemizygous deletion of Dicer (Di Leva et al. 2014), indicating the tumour suppressor activity of this protein. This evidence stresses the importance of miRNA maturation proteins, such as DGCR8 and Dicer, in brain function and other complex biological processes, as well as their involvement in several disorders, probably due to significant alterations in the maturation of the whole miRNA collection.

3. The evolution of microRNAs

As with (almost) everything in biology, in order to understand how the biological processes work on the organism-level, it is essential to understand those that work on the molecular one. Therefore, to comprehend how miRNAs could have contributed in animal evolution requires knowing how miRNAs have evolved themselves.

3.1. Emergence of new microRNAs and their genomic organization

Although still under controversial debate, different mechanisms of miRNA emergence and innovation have been proposed. Information about miRNA genomic locations and their molecular ages has shed light on the evolutionary history behind these regulators. From the serendipitous appearance of new functional RNA hairpins to genome/genomic duplication processes, as well as the long-term effects of strong/relaxed selective pressures, a wide range of events may explain the current miRNA repertoires we can see today in the metazoan genomes.

A substantial number of miRNAs are grouped within clusters, many of which conserve their organization across species. As shown in humans, the majority of clusters contain from two to eight miRNAs, but there are some exceptions like the miR-379 and miR-512-1 clusters, respectively located on human chromosomes 14 and 19, and respectively involving 42 and 46 miRNA members (Guo et al. 2014) (**Figure 8**). Considering that clusters are highly conserved and that, in many cases, clustered miRNAs are involved in similar regulatory functions, it has been hypothesized that this kind of genomic organizations confer important selective advantages such as the transcriptional co-regulation and the simultaneous expression of grouped miRNAs required in a particular response (Berezikov 2011; Campo-Paysaa et al. 2011). Although not all clustered miRNAs belong to the same family (they do not necessarily share homologous sequences), many of them do, suggesting that they originated after

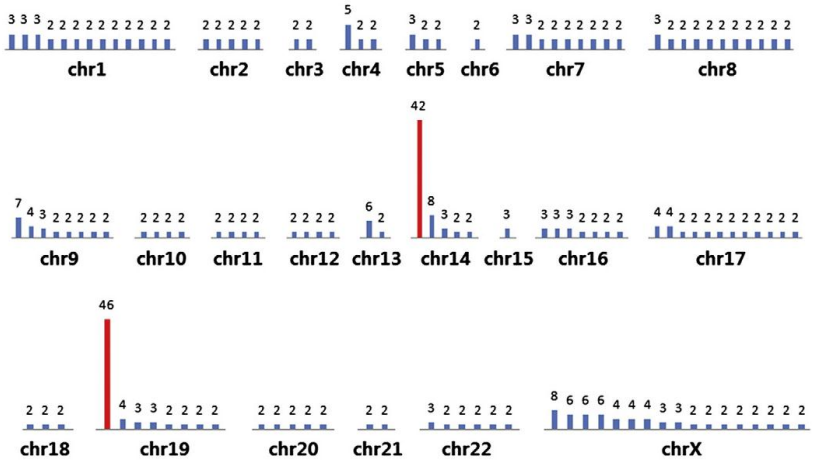


Figure 8. Human miRNA gene clusters and their chromosomic distribution. The numbers in each bar correspond to the number of miRNAs in each cluster (adapted from Guo et al. 2014).

---

duplication events of different magnitude, ranging from local to genome-wide processes (Hertel et al. 2006).It is estimated that tandem duplications have been the main source in the expansion of paralogous clustered miRNAs, as seems to be the case of the miR-15, miR-141/miR-200 and miR-302 clusters, arisen from old tandem duplications at the base of tetrapoda, chordates and amniotes, respectively (Hertel et al. 2006). Paralogous miRNAs present at distal genomic locations could have more intricate histories. A few of them have been associated with the metazoan genome-wide duplication events that occurred before the vertebrate and teleost expansions (Hertel et al. 2006). Many miRNA clusters, however, are the result of different events of reorganization, duplication and loss. For instance, the already mentioned miR-17/miR-92 cluster, closely linked to the early evolution of vertebrates, probably experimented different duplication episodes affecting single miRNA genes, the whole cluster and the complete genome, interspersed by independent loss of particular miRNAs within each lineage (Tanzer and Stadler 2004).

Duplicated miRNAs have reported interesting examples of rapid evolution and new miRNA emerging, probably due to a relaxation of the selective pressures over paralogous sequences. The miR-506 cluster, present in mammals at chromosome X and preferentially expressed in testis, has registered multiple lineage-specific changes across primates, including sequence substitutions and miRNA copy number changes (Zhang et al. 2007). The increased rate of change in primates has been proposed as a source of miRNA innovation, probably linked with the acquisition of new regulatory functions in primates related to testis development, sexual maturation and male reproduction (Zhang et al. 2007; Guo et al. 2014).

In addition, a vast proportion of miRNAs reside in introns of protein and non-protein coding genes (e.g. at least 63% of all human miRNAs are intronic), and are encoded in the same orientation as their host genes, mainly transcribed using their promoters. Most intronic miRNAs are younger than those located along intergenic regions, which could be explained, at least partially, if the host genes were degenerated and lost, but in this case their contained miRNAs would have remained. Some authors have proposed introns as one of the most plausible regions for giving rise to novel miRNAs since they are naturally transcribed without needing their own promoters and only the formation of a hairpin secondary structure is enough to integrate into the canonical miRNA maturation pathway (Wang et al. 2010; Campo-Paysaa et al. 2011; Meunier et al. 2013). Nevertheless, some intronic miRNAs seem to have their own promoter regions and to be autonomously expressed (Berezikov 2011) (**Figure 9**).

Apart from introns and duplications, other non-coding regions such as lncRNAs, snoRNAs, tRNAs, pseudogenes, transposable elements, antisense miRNA transcripts and, essentially, any RNA product that could adopt a miRNA-like hairpin structure with RNAi properties, have all shown signs of being possible sources of miRNA *de novo* formation (Berezikov 2011; Meunier et al. 2013).

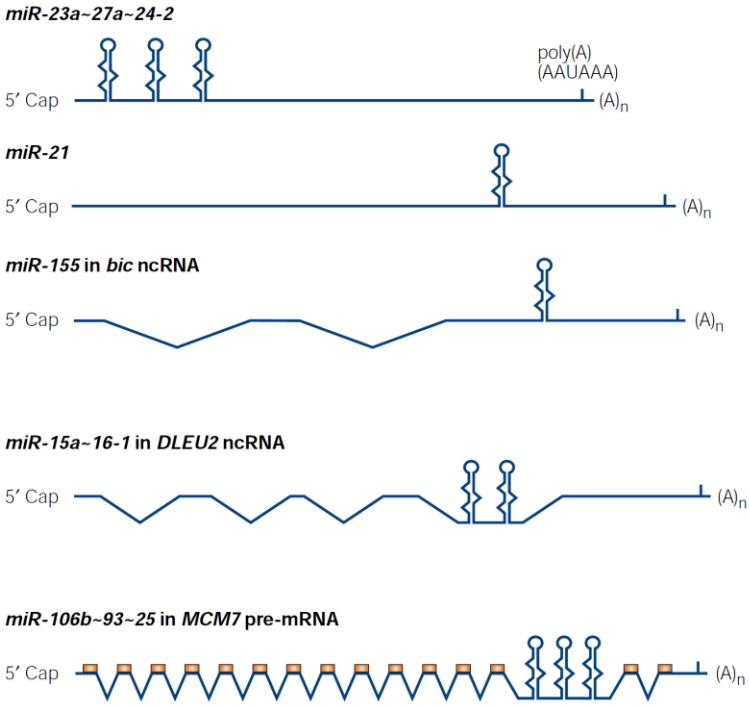


Figure 9. Relative genomic location of five pri-miRNAs. Exonic miRNAs in non-coding transcripts such as the human examples of the miR-23a/27a/24-2 cluster, miR-21 and miR-155 (miR-155 was found in a previously defined non-coding gene, *BIC*). Intronic miRNAs in non-coding transcripts such as the miR-15a/16-1 human cluster, localized within the fourth intron of the non-coding gene, *DLEU2*. Intronic miRNAs in protein-coding transcripts, such as the miR-106b/93/25 human cluster, which is embedded in the thirteenth intron of DNA replication licensing factor *MCM7* transcript (variant 1, which encodes isoform 1). The mouse miR-106b/93/25 orthologue is also found in the thirteenth intron of the mouse *MCM7* orthologous gene. The hairpins indicate the miRNA stem-loops. Orange boxes indicate protein-coding sequences (Adapted from Kim 2005).

As a single miRNA can randomly target many genes, the emergence of novel miRNAs could entail severe detrimental consequences by repressing important genes required in the cell. Nevertheless, younger miRNAs are usually expressed at low levels and restricted to specific tissues, while older and more conserved miRNAs show broad patterns of expression (Liang and Li 2009; Shen et al. 2011). Based on these

observations, it has been proposed that evolutionary conserved miRNAs have experienced increased transcriptional control, being initially scarce and probably targeting a high proportion of genes with adverse effects. Gradually, miRNAs with neutral or advantageous roles, for which purifying selection could eliminate deleterious target sites, expanded their expression to higher levels and numbers of tissues (Chen and Rajewsky 2007; Wang et al. 2010; Berezikov 2011; Shen et al. 2011). In fact, older miRNAs (those that emerged early in evolution) seem to be under stronger selective pressures since they are more ubiquitously expressed and conserved, capable of targeting many genes and involved in diseases with higher probabilities. Moreover, older miRNAs tend to localize in specific configurations such as clusters or particular chromosomes (for example, old human miRNAs are preferentially confined along chromosomes 7, 9, 15, 18 and 22). All these patterns suggest an increase of the selective forces acting on miRNA genes and their regulatory activities over time (Liang and Li 2009; Wang et al. 2010; Shen et al. 2011; Zhu et al. 2012; Iwama et al. 2013; Meunier et al. 2013; Santpere et al. 2016).

3.2. Diversification of microRNAs

Once a miRNA gene has emerged it can further evolve and acquire new regulatory functions. Mutations in miRNA genes may lead to changes in target gene recognition if they occur within the seed or at any position along the mature region involved in the mRNA binding. Also changes in the miRNA primary and precursor sequences could modify the hairpin structure and consequently the capability of Drosha, Dicer, RISC and other maturation proteins, which could ultimately alter the proportion of mature synthesized miRNAs (Berezikov 2011). Nevertheless, apart from nucleotide changes at the DNA level, post-transcriptional variation has also contributed to the diversification of miRNA functions.

3.2.1. Structural shifts

Different mature miRNAs can be generated from the same hairpin precursor depending on the cleavage position used by Drosha and Dicer proteins during the maturation process, as well as the non-template addition of terminal nucleotides. All mature miRNAs produced from the same precursor molecule that differ in a few nucleotides at their 5' and/or 3' extremes are considered as isomiRs (Landgraf et al. 2007; Morin et al. 2008; Chiang et al. 2010). Generally, only one of the isomiRs is produced at higher rates and it is annotated as the reference form; nevertheless, all of them could potentially have relevant functions regardless of their expression levels. IsomiRs varying at the 3' extreme are not expected to significantly change the miRNA function since this region is not directly involved in the target gene recognition. However, the addition or loss of nucleotides at the 5' extreme would redefine a completely new seed sequence (determined by nt 2-8 in the mature miRNA molecule) and totally change the target gene repertoire. That is probably why 5' isomiRs have been less commonly found than 3' isomiRs. The dominance of one particular isomiR can change among orthologous and paralogous miRNAs in a process known as *seed shifting* (Grimson et al. 2008; Wheeler et al. 2009; Chiang et al. 2010; Marco et al. 2010).

One of the most common mechanisms of miRNA diversification relies on the mature strand (-5p or -3p) that exerts the target gene repression. It was initially assumed that only one of the mature strands constituting the miRNA duplex (known as the *major strand* and usually attributed to the -5p strand) was incorporated into the RISC complex with a regulatory function, while the other one, referred to as the *minor* or *passenger strand*, was degraded. Nevertheless, increasing studies, mainly based on deep-sequencing data, have shown that both strands can play important regulatory roles and that even the proportion of functional -5p and -3p strands of certain miRNA may differ among different species, tissues or developmental stages due to a process called *arm switching* (Wheeler et al. 2009; Yang and Lai 2010; Berezikov 2011). Although the preference for one or the other mature strands to

be functional could be determined by the stability of the mature miRNA duplex or by the sequence encoded in the primary or flaking regions, the molecular bases explaining the *arm switching* phenomenon are not yet fully understood (Berezikov 2011; Griffiths-Jones et al. 2011).

Finally, in some cases, as a result of tandem duplication events, paralogous sequences may locate one next to the other. Each of these sequences could partially and complementarily pair with upstream and downstream fragments forming different hairpins. This mechanism, defined as *hairpin shift*, implies that one mature sequence can be part of two different hairpins, both of which probably differ in their secondary structure and maturation process and, therefore, could be the subject of different *seed* and *arm switching* diversification processes (Berezikov 2011).

Although interesting examples of *seed*, *arm* and *hairpin shifting* have been reported across different species and biological conditions, the functional and evolutionary consequences of these events are yet to be determined.

3.2.2. RNA editing

Another post-transcriptional mechanism involved in miRNA variation is RNA editing, which involves the modification of an RNA sequence by nucleotide insertion, deletion or substitution. Through this mechanism, a different product from that encoded at the DNA level can be produced in a range from 0 to 100% depending on what proportion of transcripts are edited. Most RNA editing is the site-specific nucleotide modification performed by ADAR enzymes (Adenosine Deaminases Acting on RNAs), which catalyze the hydrolytic deamination of an adenosine into an inosine (A-to-I) in dsRNAs, or by APOBECs, which deaminate a cytidine into a uridine (C-to-U/T) in both RNA/DNA sequences.

The large family of APOBEC enzymes (at least 11 primary gene products and several alternatively spliced variants have been found in

humans) are present across vertebrates and play a relevant role in the restriction of virus invasion and transposable element introgression, the diversification of endogenous protein functions and the somatic variation of antigen receptor genes (Conticello 2012; Salter et al. 2016). However, the most common RNA editing in metazoans is catalyzed by ADARs. Since inosine is recognized as guanosine by the cellular machinery, ADAR A-to-I editing is also expressed as an A-to-G change (**Figure 10**). The ADAR family is not as diverse as that of APOBECs (until now only the three paralogs ADAR1, ADAR2 and ADAR3 have been identified in vertebrates) (Nishikura 2016), even though they account for more than 90% of the edited sites registered in the human transcriptome (Levanon et al. 2004; Jin et al. 2009; Peng et al. 2012).

The role of RNA editing is still under extensive debate. As many potential APOBEC and ADAR target sequences are expected to be present in the cell, it is plausible that editing randomly occurs with no significant effects at all. Moreover, technical limitations experienced during the analysis of editing modifications have reported substantial differences in the editing estimations, which has generated various hypotheses regarding the extent and functionality of this phenomenon (Bass et al. 2012). Nevertheless, high-throughput sequencing analyses have shown that a large proportion of transcript sites are subject to editing (Peng et al. 2012), and that some editing events seem to have functional consequences that could be *visible* for natural selection (Salter et al. 2016). Also, some genomic regions have shown clear editing patterns across tissues, species and development (Gommans et al. 2009; He et al. 2011; Li et al. 2013; Warnefors et al. 2014). Editing appears to be enriched in brain regions and to affect genes involved in nervous system function (J.B. Li et al. 2009). Consequently, altered editing patterns have appeared in epileptic mice, patients with chronic depression and malignant gliomas (Maas et al. 2010). On the other hand, RNA editing has been connected to the stability of the transcripts and their subsequent degradation or stabilization when an A-to-I transition respectively creates or

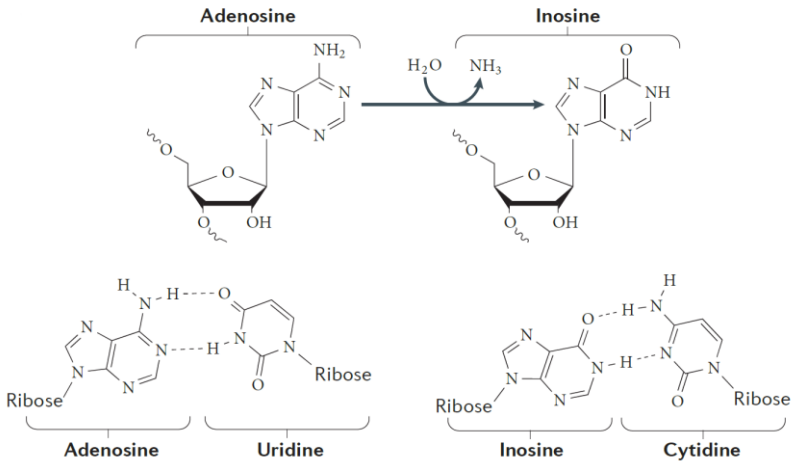


Figure 10. A-to-I editing by ADARs. ADARs catalyze a hydrolytic deamination reaction that converts adenosine to inosine. Adenosine base-pairs with uridine, whereas inosine behaves like a guanosine, as it base-pairs with cytidine in a Watson-Crick-bonding configuration (adapted from Nishikura 2016).

---

eliminates a base-pair mismatch (Levanon et al. 2004). Interestingly, RNA editing is a crucial mechanism required for normal animal development and survival, since disruption of ADAR and APOBEC enzymes have proven to be lethal or produce severe developmental disorders in mice (Nakamuta et al. 1996; Nishikura 2016). Overall, these observations suggest that, at least partially, editing may have an important biological function in animal evolution.

a) RNA editing in protein-coding and non-coding regions

Editing changes can alter protein functions when they occur in protein-coding sequences. For instance, A-to-I editing within the mRNA coding sequence of the Glutamate receptor subunit B (*GluR-B*) is essential for normal reception function. Changes in the editing patterns of GluR-B have been associated with epilepsy and premature death in mice, as well as with malignant brain tumour invasion in humans (Maas et al. 2001). Another famous RNA editing example is the C-to-U editing in the Apolipoprotein B (*ApoB*) mRNA. Editing at

cytidine 6666 of ApoB occurs in the small intestine of mammals but not in the liver, producing transcripts with different lengths and lipid transportable capacities. Editing alterations in this position have been associated with an increase of cholesterol levels in the blood (Greeve et al. 1993; Salter et al. 2016). Remarkably, editing has also been connected to alternative splicing processes which expand the universe of diversity that can be generated through post-transcriptional modification events. A-to-I editing at position 1087 in the *Nova1* transcript (a brain-specific regulator of alternative splicing) does not change the splicing regulatory activity, but rather the stability of the protein. Furthermore, *Nova1* editing seems to be conserved across amniotes and to increase during embryonic development in mice and chicken (Irimia et al. 2012).

Although relevant modifications can arise by editing within coding regions, most RNA editing takes place along non-coding sequences, whose functional implications still need to be determined in the majority of cases (J.B. Li et al. 2009). Some interesting examples have shown that A-to-I editing within mRNA untranslated and intronic regions are related to transcript nuclear retention (Prasanth et al. 2005), alteration of splicing sites (Rueter et al. 1999; Athanasiadis et al. 2004; Lev-Maor et al. 2007) and changes in the target sequences recognized by miRNAs and other post-transcriptional regulators (Liang and Landweber 2007; Peng et al. 2012).

Among non-coding transcripts, miRNAs gather a huge proportion of the editing. Since miRNAs form dsRNA structures along all their biogenesis, these regulators can be bound and edited by ADARs in the primary, precursor and mature stages. At least 16% of miRNA transcripts are subject to A-to-I editing in humans (Kawahara et al. 2008; Peng et al. 2012) and in many cases miRNA editing has shown interesting patterns. Certain regions, such as brain tissues, have registered higher levels of editing activity while others, such as liver tissue, barely accumulate edited miRNAs. Also, editing at specific miRNA sites seems to be conserved across evolutionary distal species and to increase in frequency throughout development (**Figure 11**).

Moreover, certain diseases, like different types of cancer, have been significantly associated with changes in the miRNA editing patterns when compared with healthy control cases (Wheeler et al. 2009; Gong et al. 2014; Warnefors et al. 2014). Although largely unexplored, miRNA editing could dramatically change the miRNA function if the former takes place within the seed and alters the spectrum of target genes (Kawahara, Zinshteyn, Sethupathy, et al. 2007) or, if occurring along the primary or precursor regions, it could modify the secondary structure of the miRNA hairpin and its processing by Drosha or Dicer (Yang et al. 2006; Kawahara, Zinshteyn, Chendrimada, et al. 2007). Further studies have shown that the ADAR binding to the miRNA can block the RISC loading. In this case, ADAR would affect the processing and function of the miRNA by physical disruption of the RISC binding instead of by A-to-I nucleotide change (Heale et al. 2009; Yang et al. 2013). In addition, as some studies have shown, an excess of editing activity could be interpreted as a detrimental feature by the cell machinery. For example, hyperedited primary and precursor miRNAs that were not recognized by the proteins involved in miRNA biogenesis were shown to be degraded by the activity of the nuclease Tudor-SN in the oocyte and zygote embryonic stage in mice (García-López et al. 2013).

Apart from miRNAs, repetitive elements constitute another important source of RNA editing as these sequences tend to pair, forming double-stranded structures. Primate-specific Alu repeats are one class of short interspersed elements (SINEs) that cover around 10% of the whole human genome (Batzer and Deininger 2002). Primates account for the highest editing rates found in metazoans, which is explained by the presence of Alu repeats along their genomes. For instance, in humans, over 90% of editing happens in Alu sequences (Athanasiadis et al. 2004; Kim et al. 2004; Levanon et al. 2004; Peng et al. 2012; Li et al. 2013; Picardi et al. 2015). Interestingly, depending on their location, editing in Alu elements could trigger different consequences such as the nuclear retention of the transcripts or the alternative splicing of

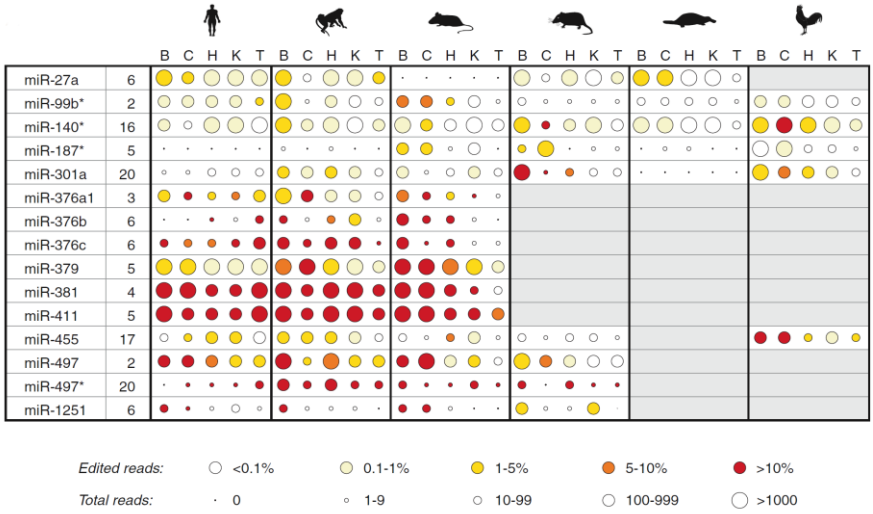


Figure 11. Frequency of miRNA editing across tissues and species. Estimated miRNA frequencies in the brain (B), cerebellum (C), heart (H), kidney (K) and testis (T). From left to right, the animal silhouettes represent human, macaque, mouse, opossum, platypus and chicken. miRNAs in the left-hand column are followed by the position of the edited site within the mature miRNA (asterisk refers to the minor strand). The colour of each circle corresponds to the proportion of edited reads, while the size of the circle corresponds to the total number of reads for each miRNA. Data was not normalized and therefore expression levels should not be compared across samples. Gray shading indicates the absence of an annotated miRNA ortholog in that particular species (adapted from Warnefors et al. 2014).

---

the precursor mRNAs when they respectively occur in UTRs or introns (Rueter et al. 1999; Chen and Carmichael 2008). The functional and evolutionary consequences of the widespread distribution of Alu repeats in primates together with their tendency towards being edited are still largely unknown. It has been argued that highly-edited Alu sequences may have driven genome evolution through exonization processes (the inclusion of a new exon in a transcript) (Lev-Maor et al. 2007; Möller-Krull et al. 2008). Moreover, when linking the profuse editing achieved in primates and brain tissues, some authors have speculated the possible role of Alu editing in the evolution of brain functions in primates (Eisenberg et al. 2005),

as well as the compensation for a supposed Alu invasion in primate genomes (Levanon and Eisenberg 2015).

b) RNA editing as a driving force in evolution

RNA editing has been revealed as a potential source for genetic variation and evolutionary change since it allows for the exploration of a new sequence space, reducing the cost generated by other irreversible mechanisms such as DNA mutations (Paz-Yaacov et al. 2010). As the wild-type transcripts (the non-edited ones) continue to be present in the cell, a newly emerged edited variant that is usually lowly expressed (most sites in humans show editing levels below 2%) could be tested under relaxed selective pressures. In this regard, if we assume that RNA editing of transcripts continuously emerges at low frequencies and with no beneficial effects, under a scenario of environmental change, such spuriously edited products could then become crucial for the organism, resulting in its acclimatization. Hence, the possibility of generating a huge range of genetic variation through a low-cost mechanism could provide a selective advantage in certain organisms and environments by promoting phenotypic plasticity and evolvability (Gommans et al. 2009; J.B. Li et al. 2009). RNA editing has also been considered as a contributor to the buffering of external noise and the robustness of gene networks by adding a layer of genetic regulation and increasing the capacity of response in the presence of external variation (Gommans et al. 2009; Paz-Yaacov et al. 2010).

Hence, although recurrence of RNA editing, and in particular, miRNA editing, has been widely demonstrated and proven to be crucial for animal survival, the consequences of editing at specific genetic sites, developmental stages, tissue identities, animal species and other biological conditions remains largely undetermined. *"How and why do some tissues show higher editing frequencies than others?"; "to what extent is editing conserved among and within species and populations?";* or "*has miRNA editing been a relevant mechanism underlying phenotypic variation?"* are some of the questions that still need to be addressed.

4. The roles of microRNAs in animal evolution

4.1. microRNAs over time and animal complexity

Although the precise time remains uncertain, miRNA surely emerged early in animal evolution. At least 34 miRNA families are conserved across bilaterians and, even though Porifera miRNAs show specific sequence and structural features, the presence of Drosha homologous proteins in this group suggests a common origin of the miRNA machinery across all metazoans (Grimson et al. 2008; Berezikov 2011). When looking at large time-scales and comparing many different taxa, as well as when focusing on particular species and populations, it is easy to realize that miRNAs, as a whole, have experimented different patterns of evolution over time depending on their molecular ages, genomic locations, regulatory functions and levels of conservation. Two general aspects of miRNA evolution in the context of animal evolution have been proposed.

On the one hand, the emergence of miRNAs is considered as an ongoing process. New miRNAs are continuously being added to animal genomes and, although the majority of them tend to be detrimental and eliminated by purifying selection, once they integrate into the regulatory cell machineries, they have low chances of being lost (Hertel et al. 2006; Wheeler et al. 2009; Campo-Paysaa et al. 2011; Meunier et al. 2013). This pattern of genomic accumulation is based on the fact that most new miRNAs are presented in a particular *phylum* and do not show homologous sequences in other *phyla*, while ancient miRNAs are highly conserved and generally preserved in all groups deriving from a common ancestor (Hertel et al. 2006; Wang et al. 2010; Campo-Paysaa et al. 2011).

On the other hand, miRNAs have experienced different evolutionary rates over time. During metazoan evolution, some periods have been related to miRNA bursts and expansions, while other time spans seem to have been less fruitful. Although several studies have estimated different diversification rates and times for miRNAs depending on the

molecular clock analyses applied and the set of miRNAs under consideration, certain periods seem to have been particularly crucial in miRNA evolution. In particular, regarding the emergence of vertebrates, high raw numbers of newly evolved miRNAs and high miRNA acquisition rates have been registered (Heimberg et al. 2008). Also, the vast majority of human miRNAs (genes and families) emerged at four main times coinciding with the advent of vertebrates, mammals, primates and hominoid lineage (Hertel et al. 2006; Wang et al. 2010; Iwama et al. 2013; Meunier et al. 2013) (**Figure 12**).

The constant increase in miRNA repertoires has been directly linked with animal complexity, since it is assumed that the emergence of new miRNAs is connected to the acquisition of new cell types and tissue identities in organisms (Wang et al. 2010; Meunier et al. 2013). Indeed, miRNA expansion periods have been correlated to advents of phenotypic complexity in animal *phyla* throughout evolution, which reinforce the argument that non-coding regulatory elements are key contributors of animal diversification (Heimberg et al. 2008). Despite several interesting examples having been reported (Christodoulou et al. 2010), more functional approaches are still needed in order to demonstrate the contribution of the rise of miRNA to bauplan innovations. Intriguing questions connecting animal complexity and miRNA emergence are still under debate. For example, the significant bias in the number of reported miRNA genes among species (particularly non-human species and species that do not belong to model organisms), as well as the lack of low-quality genome annotations for most of them, have probably distorted the real number of conserved and young miRNAs within each group, and thus the potential roles of these regulators in the evolution of organism organization (Hertel et al. 2006). Additionally, differences in organism complexity should also be conceived through geological time and not only through different co-existing groups (Peterson et al. 2009). Since each metazoan lineage has experienced its own increase in the number of miRNAs, it is assumed that miRNA-mediated regulation did not only vary among current living species, but also among organisms

from the same clade at different periods such as, for instance, an arthropod from the current era and another one from the Cambrian ages.

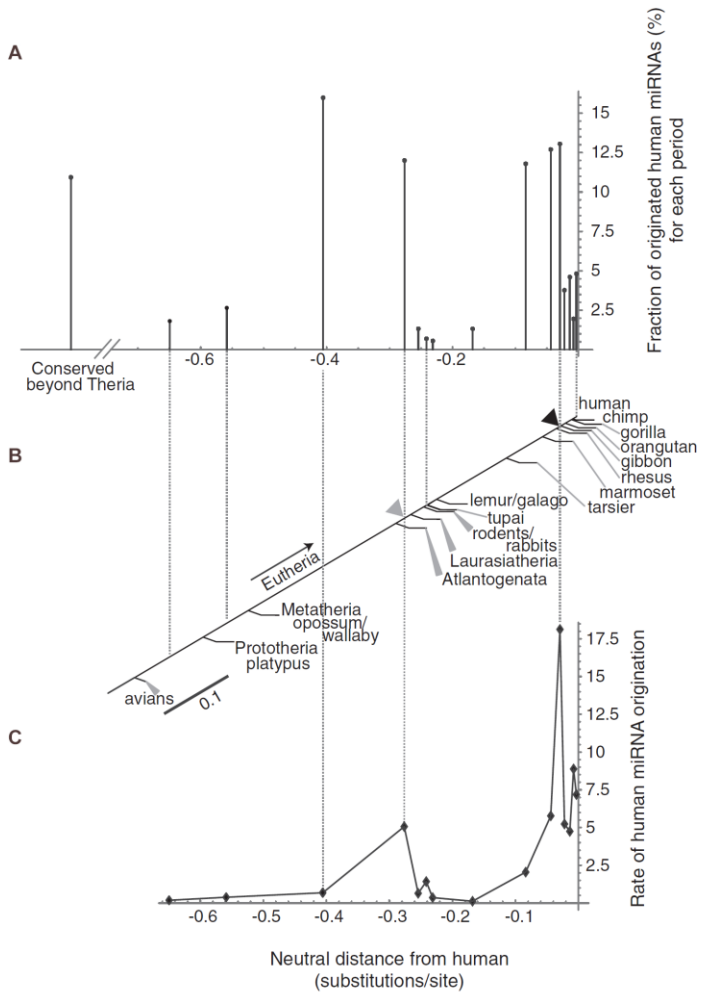


Figure 12. Evolutionary periods and rates of human miRNA emergence over time. (A) Percentage of human miRNAs originating during each period shown as bar charts in which the *x* axis stands for the neutral evolutionary distance from humans. **(B)** Relationship of each emergence bar with the divergence of a species or clade. **(C)** Rate of emergence of human miRNAs, in which the *y* axis indicates the percentage of miRNAs originating during each period, with a scale of 0.01 neutral substitution per site (adapted from Iwama et al. 2013).

4.2. Implications of microRNA genetic variation in animal function

It has been proven that miRNA containing regions are, in general, selectively constrained compared with neutral evolving flanking regions (Saunders et al. 2007; Quach et al. 2009; Carbonell et al. 2012). As mentioned, evolutionary older miRNAs, shared among different taxonomic groups, show high levels of sequence conservation, even higher than 18S rDNA, one of the most conserved genes in the metazoan genome. These features, together with the continuous addition of miRNAs within animal genomes and the presence of species-specific miRNAs in all clades, have permitted the use of these regulatory molecules as phylogenetic markers in evolutionary studies (Hertel et al. 2006; Wheeler et al. 2009; Campo-Paysaa et al. 2011). It has been argued that due to the small size of miRNAs, their multi-targeting properties and their implications in complex networks affecting crucial regulatory processes, changes in miRNA regions must be extremely detrimental (Li et al. 2012; Li and Zhang 2013). However, as widely demonstrated, variability in animal genomes is extensive and it affects miRNA regions as well. Moreover, given that miRNAs can target thousands of genes, genetic changes in their sequence are expected to cause important pleiotropic effects.

The burst of genomic data generated in the past few years thanks to the proliferation of high-throughput sequencing technologies has made it possible to obtain detailed information about the genetic variation presented in our genomes. This variation ranges from changes at a single base to fragment rearrangements of several megabases, and it is not only present at the genomic DNA, but also at the RNA through post-transcriptional modifications such as alternative splicing, maturation alterations or RNA editing modifications. One of the most popular catalogues compiling genomic variation data is the 1000 Genomes Project, based on whole-genome sequencing analyses performed on more than 2,000 individuals from 26 worldwide populations. This publicly available database provides a comprehensive picture of the genetic variation

present in the human genome including SNPs, short insertions and deletions, and structural variation (The 1000 Genomes Project Consortium 2015). There are many other platforms in charge of collecting valuable data such as dbSNP, which includes diverse variant annotations from 53 different species (Sherry et al. 2001) or the International HapMap Project centred on developing a map of nearby SNPs inherited as a single block (haplotypes) in the human genome (The International Hapmap Consortium 2003). Other databases have focused on providing information on variation specifically related to miRNAs like miRNA SNiPer, a catalogue of SNPs present in vertebrate miRNAs (Zorc et al. 2012); PolymiRTS, a database of polymorphisms in predicted target sites and miRNA genes (Bhattacharya et al. 2014); or the YM500, which includes information about miRNA variants (isomiRs) produced by RNA editing and arm switching (Cheng et al. 2013).

Genetic variants in miRNAs can alter their biology at different levels, such as their maturation process, secondary structure and expression levels, or at their base pairing properties and spectrum of recognized target genes. To date, several studies have addressed the association of miRNA genetic variants with relevant functional changes, mainly related to disease (Jazdzewski et al. 2008; Lv et al. 2008), as well as to phenotypic modifications of evolutionary interest (Li et al. 2012; Lopez-Valenzuela et al. 2012). Nevertheless, there is still very little information regarding the contribution of genetic variation to miRNA function, and even more importantly, regarding the extent to which such variation has been evolutionary adaptive in animal diversification.

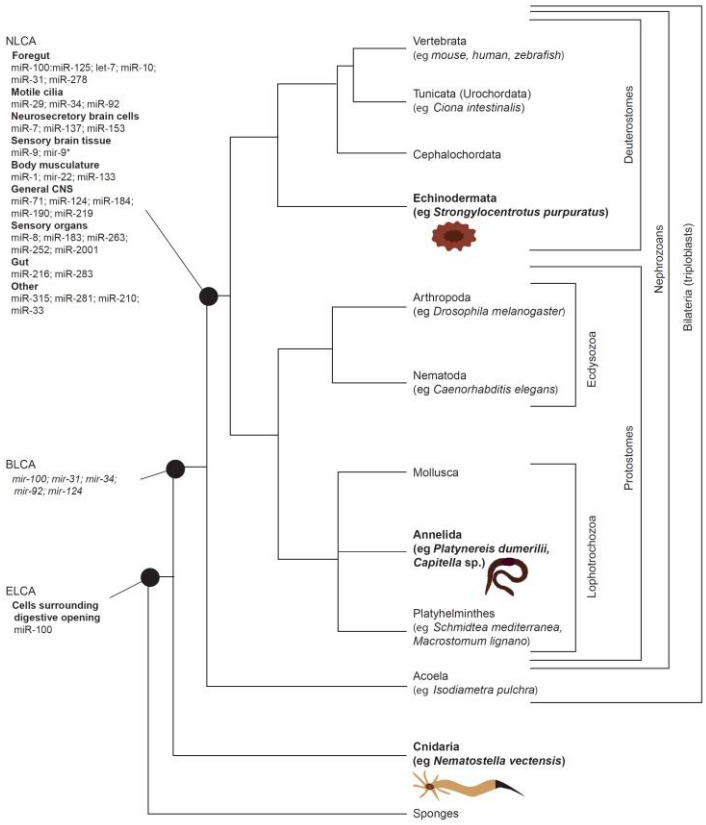
4.2.1. microRNA genetic variation across species

Generally, homologous miRNA genes in distal evolutionary groups that share a common ancestor (orthologous miRNAs), show high levels of evolutionary conservation, especially in the seed region. Some of them date back hundreds of millions of years with no changes at all, highlighting the effect of strong purifying selection over these

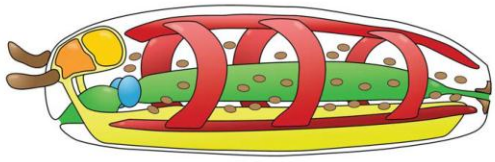
regulators (Berezikov et al. 2005; Heimberg et al. 2008; Li and Zhang 2013). Indeed, phylogenetic analyses have revealed that certain animal miRNAs are very old. For example, miR-100, the oldest known animal miRNA, emerged at the base of eumetazoans before the Cnidaria and Bilateria diversification, with only a few modifications since then (Christodoulou et al. 2010). Interestingly, in many cases these ancient orthologous miRNAs do not only show conservation at their sequence levels but also at their patterns of expression, suggesting that they have significantly conserved their regulatory functions. As an example, continuing with miR-100, this miRNA and its cluster partners, miR-125 and let-7, have been found to be active in the neurosecretory cells surrounding the foregut in all nephrozoan species analyzed (the last common ancestor of protostomes and deuterostomes). These marked conservation patterns, in both sequence and expression, present in several other miRNA examples, have permitted a high-confidence reconstruction of what may have been the cell types and tissues in earlier nephrozoans (**Figure 13**) (Christodoulou et al. 2010; Mulder and Berezikov 2010). It is worth noting that the high level of miRNA sequence conservation does not always correlate to high levels of expression conservation, as shown in an analysis of almost 100 conserved miRNAs from medaka, chicken, zebrafish and mouse. In many cases, miRNA expression timing and location varied among the four species, presumably affecting their regulatory function (Ason et al. 2006). Thus, variations in the expression patterns of conserved miRNAs could also contribute to shaping phenotypic differences throughout evolution.

The assumption that miRNAs show high levels of sequence conservation has somehow biased the identification of orthologous miRNAs in different species by means of comparative genomic approaches (Bentwich et al. 2005; Lin et al. 2010). Conserved miRNAs that have accumulated significant variation in certain clades and species-specific miRNAs are usually masked through gene homology searches. In fact, significant genetic changes among orthologous miRNAs do exist, and they have attracted special interest because they

A



B



Ancient bilaterian miRNAs

Foregut	Motile cilia	Neurosecretory brain tissue	Sensory brain tissue	General musculature	General CNS	Sensory organs	Gut	Other
miR-100 let-7^ miR-125^ miR-375^ miR-10 miR-278 miR-31	miR-29 miR-92 miR-34	miR-7 miR-137 miR-153	miR-9 miR-9*	miR-22# miR-1 miR-133	miR-71 miR-124 miR-184 miR-190 miR-219	miR-8 miR-183 miR-263 miR-252^ miR-2001	miR-216 miR-283	miR-315 miR-281 miR-210^ miR-33

◄ **Figure 13. miRNA conservation in metazoan evolution (A)** Phylogenetic relationships among major phylogenetic clades and reconstruction of ancestral tissue types based on conserved miRNA expression patterns. NLCA, BLCA and ELCA: the Nephrozoan, Bilaterian and Eumetazoan last common ancestor, respectively (adapted from Mulder & Berezikov 2010). **(B)** Reconstruction of tissue identities present in the last common ancestor of protostomes and deuterostomes based on the tissue-specificity of ancient miRNAs (adapted from Christodoulou et al. 2010).

---

may be involved in evolutionary innovations. miRNA genetic novelties could emerge from the acquisition of new hairpins with a regulatory capacity. Previous studies demonstrated human-specific expression of 31 miRNAs across different brain regions that were absent in chimpanzee and macaque. These miRNAs were enriched in neuronal functions and, at least, a proportion of them could have contributed to shaping neural functions in our species (Hu et al. 2011). miRNA genetic novelties could also arise from changes in their expression levels, nucleotide sequences, (e.g. single nucleotides variants or SNVs) and genomic dosage (e.g. miRNA copy number variants or CNVs). Nevertheless, the specific contribution of these changes to animal evolution remains to be determined.

4.2.2. microRNA genetic variation within species: human populations

Stringent miRNA sequence conservation has been repeatedly reported not only among species, but also when looking at the intraspecific level. In different studies including individual genomic sequences from different human populations and based on the allele frequency of SNP variants, the miRNA precursor regions presented high conservation compared with the 5' and 3' flanking sequences; and within the precursor, the sequences corresponding to the mature strands were significantly more conserved than the rest of the hairpin, particularly at the first 14 nt which include the seed (Saunders et al. 2007; Quach et al. 2009). Only around 10% of the human miRNAs showed SNPs (dbSNP annotations) in the precursor region and less than 1% in the mature region (Saunders et al. 2007). More recently, studies based on

14 human populations using data from the 1000 Genomes Project estimated an unexpected amount of variability. Among 720 miRNAs, 44% had SNVs, most of which, however, presented very low allele frequencies in most of the populations studied (Carbonell et al. 2012). Although these estimations may substantially change as more miRNAs and polymorphic variants are discovered, it seems that generally miRNAs have been depleted of genetic variation, especially in the functional mature sequences. Moreover, it is likely that a substantial proportion of miRNA variants have recessive or neutral effects and that most of them segregate at very low allele frequencies in human populations (Carbonell et al. 2012; Lu and Clark 2012; Torruella-Loran et al. 2016).

Though scarce, existing genetic variants among close evolutionary relatives have provided interesting evidence on miRNA function diversification. An exhaustive study performed on the primate-specific miRNA, miR-1304, reported that an ancestral allele located in the seed region, found in Neanderthals and segregating at very low frequency in Asian human populations, permitted the miRNA-mediated repression of *ENAM* and *AMTN*, two target genes involved in teeth enamel formation. On the contrary, the human derived allele did not reduce the expression of these genes, but rather significantly increased the number of putative target genes, chiefly associated with behaviour and nervous system development and function (Lopez-Valenzuela et al. 2012). These findings illustrate the presumed contribution of miRNA nucleotide changes to shaping phenotypic traits with evolutionary relevance among ancient and present humans.

4.2.3. Genetic variation in target sites

Another important source of miRNA regulatory variation is reflected at the target sites of miRNA regulated genes. Since miRNAs strongly affect gene expression, it is believed that changes in miRNAs have an influence on target site evolution and vice versa. Interestingly, genes showing overlapped spatial and temporal expression patterns with a particular miRNA tend to be depleted of target sites for that miRNA

(Farh et al. 2005; Stark et al. 2005). Moreover, ubiquitously expressed genes involved in basic cellular processes seem to evade miRNA regulation having short 3' UTRs that are specifically depleted of miRNA binding sites. In contrast, tissue-specific target genes tend to be co-expressed with their corresponding miRNAs, as well as to be involved in developmental processes and tissue-identity configuration (Stark et al. 2005). In addition, genes with evolutionary conserved target sites are usually co-expressed with their miRNAs, while genes with non-conserved sites (which are almost 10 times more abundant) are not. Altogether, these patterns suggest that the mutual exclusion between miRNAs and housekeeping-conserved genes may confer robustness in ancient and basal developmental programs, while recently evolved miRNAs would target specific genes at particular times and locations, being implicated in the fine-tuning of gene expression and developmental transitions. These observations provide evidence for the reciprocal influence between miRNA and target gene evolution (Farh et al. 2005; Stark et al. 2005; Berezikov 2011).

The effects of genetic variants within target genes have attracted most of the attention over the past few years. miRNA target sites show a high proportion of SNP variants, which could be interpreted as a relaxation of the selective constraints acting in these regions. Nevertheless, SNP density in target sites should be interpreted with caution since many predicted target genes could be false positives (Li and Zhang 2013). Computational analyses predicted that more than 400 target sites mutated in the human lineage compared with mouse orthologous regions, and that some of them were fixed by positive selection probably contributing to the evolution of human-specific traits (Gardner and Vinther 2008). At the population level, natural variation in target sites has also been registered. An interesting work observed that the ancestral and derived allele variants in the SNP rs683, which is located at the 3' UTR of *TYR1* (a gene associated with skin pigmentation), differentially affect miR-155-mediated regulation. The derived allele of rs683 is virtually fixed in the African and Asian populations, while the SNP is polymorphic in Europeans. Luciferase

assays demonstrated a significant reduction of *TYR1* expression in the presence of the derived allele, whose frequency negatively correlates with the latitude. These analyses suggest a possible role of miR-155 in the regulation of skin pigmentation and a potential local positive selection event in the *TYR1* target site in response to differential UV radiation (Li et al. 2012).

Although other interesting examples have been studied, and the link between genetic variation in miRNA-mediated regulation and evolutionary changes, and occasionally adaptation, has been demonstrated, the functional consequences of miRNA changes across and within species have not been studied in-depth yet. Also, extensive research beyond human populations concerning other primate and non-primate species is still required in order to better understand the specific contribution of miRNAs throughout animal evolution.

5. microRNAs in primates

This thesis is based on the functional implications of miRNA variation among great apes and their close primate relatives, rhesus macaque. Fossil and extant species records suggest a complex evolutionary history beyond the Primate Order. Indeed, primate taxonomy has been a controversial issue since it involves many diverse groups constituting more than 450 species, which have usually experienced recurrent events of geographical isolation, migration, inbreeding and gene flow, followed by substantial variations in their effective population sizes (Prado-Martinez et al. 2013; Rogers and Gibbs 2014). Current studies situate the origin of primates in the Cretaceous-Paleocene boundary around 80-90 million years ago (MYA). Further expansions during the Eocene gave rise to the present clades of 1) Strepsirrhini, including Lorisiformes (galagos, pottos, lorises), Chiromyiformes (Malagasy aye-aye) and Lemuriformes (Malagasy lemurs); 2) Tarsiiformes (tarsiers); and 3) Simiiformes, including Platyrrhini (New World monkeys) and Catarrhini, formed by Cercopithecoidea (Old World monkeys) and Hominoidea (great apes and gibbons) (Perelman et al. 2011; Rogers and Gibbs 2014) (**Figure 14**). Morphological and genetic data have placed the human lineage in the taxonomic group of great apes, being closely related to chimpanzees, bonobos, gorillas and orangutans. Genomic comparative analyses support that humans are more evolutionary related to chimpanzees than to any other species and that both shared a common ancestor around 4-7 MYA, which split with gorillas 6-10 MYA, and with orangutans 12-16 MYA (Marques-Bonet, Ryder, et al. 2009; Perelman et al. 2011; Prado-Martinez et al. 2013). However, there is still controversy about these estimated dates since they could substantially differ depending on the mutation rates and generation times assumed for each species (Langergraber and Prüfer 2012).

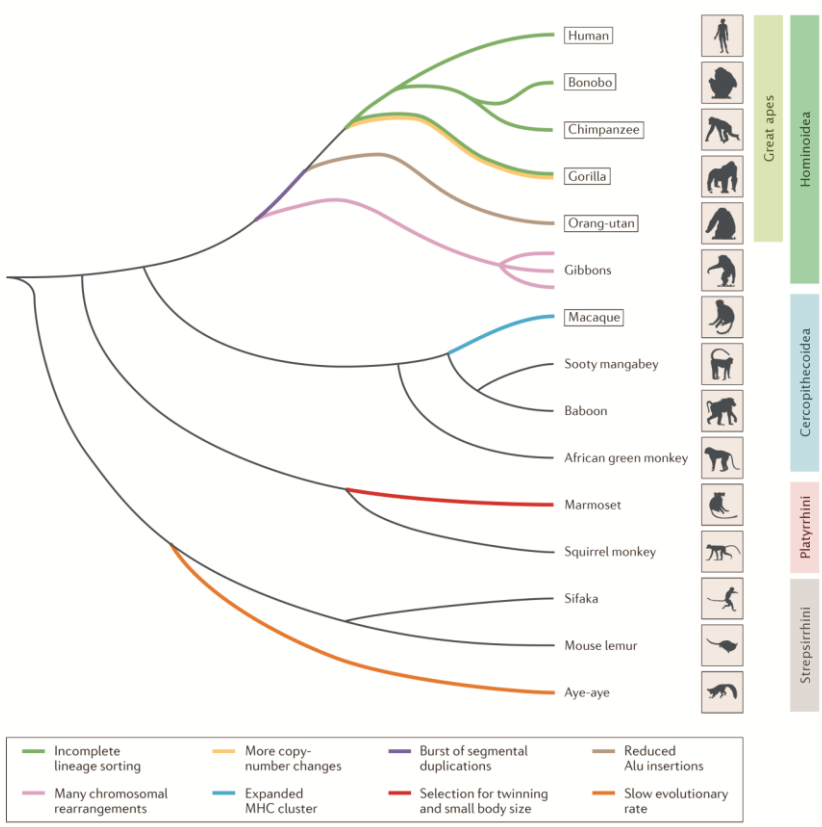


Figure 14. Primate phylogenetic tree. Evolutionary relationships among the most representative primate species. The superfamily Cercopithecoidea comprises the "Old World monkeys", native to Asia and Africa. The parvorder Platyrrhini consists of the "New World monkeys", native to South and Central America. The great ape classification includes humans (*Homo sapiens*), chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), gorillas (*Gorilla beringei* and *Gorilla gorilla*) and orangutans (*Pongo abellii* and *Pongo pygmaeus*). Species enclosed in boxes were analyzed in the results section of this thesis, and encompass the great ape and macaque (*Macaca mulatta*) species. The coloured branches indicate some interesting specific genomic features in different groups. MHC, major histocompatibility complex (modified from Rogers & Gibbs 2014).

In spite of the rich evolutionary history of the non-human great ape species, all of them have recently experienced a dramatic decline in their natural habitats and local population sizes. Extant species and subspecies of chimpanzees, bonobos and gorillas are currently located throughout central Africa; while the two orangutan species reside on the Borneo and Sumatran islands. Gibbons (Hylobatidae family), the other group that together with great apes constitutes the superfamily Hominoidea, are restricted to south-eastern tropical and southern subtropical regions of Asia; while rhesus macaque has larger extant populations spread across Asia (Prado-Martinez et al. 2013; Rogers and Gibbs 2014) (**Figure 15**). Although great ape species have undergone high population size reductions, they show more genetic diversity than humans. This phenomenon seems to be a result of the recent and rapid human expansion from a small founder population (Kaessmann et al. 2001). The current human dominance in both range and population size is very recent and, in fact, there were more non-human great apes than humans over 50,000 years ago (Wall 2013).

5.1. Why study primates and microRNAs?

The study of great ape and primate evolution has provided key insights into the understanding of human evolution and human disease (Calvignac-Spencer et al. 2012). Great apes have been the focus over the past few years because they are the only extant species that could provide information about recent human evolution. Consequently, one of the main objectives in the genomic era has been the whole-genome sequencing of all great ape genera and other closely related species such as rhesus and cynomolgus macaques (Lander et al. 2001; Consortium 2005; Marques-Bonet, Ryder, et al. 2009; Wall 2013; Rogers and Gibbs 2014). These efforts have recently been extended with the sequencing of representative numbers of living humans from different world populations, as well as with the development of paleogenomics and the sequencing of extinct hominins such as Neanderthal, Denisovan and ancient modern humans (Durbin et al. 2010; Green et al. 2010; Reich et al. 2010; Meyer et al. 2012; Wall

2013; Rogers and Gibbs 2014). Thus, the study of great ape evolution and the identification of the genetic and genomic particularities within each group could be useful to better trace the history of our origins and to identify the traits that are responsible for our *humanness*, in the hope of contributing to answer, at least from a biological point of view, an old question in human history: *what makes us humans?* (Jobling et al. 2013)

Additionally, from a less-anthropocentric perspective, the study of primate evolution has made it possible to compile a comprehensive amount of morphological, biogeographical, reproductive, behavioural, paleontological and genomic data from wild ape populations. This information has proven invaluable for endangered species conservation (Kohn et al. 2006; Avise 2010; Prado-Martinez et al. 2013), as well as for clarifying the putative evolutionary mechanisms involved in the processes of speciation and diversification, not only in the Primate Order, but also across all metazoans.

Whole-genome comparisons analyses have shown substantial similarities among humans and our closest primate relatives. For instance, the genomic difference between humans and chimpanzees has been estimated at only 1.4% when looking at nucleotide substitutions in orthologous regions (Scally et al. 2012; Wall 2013; Rogers and Gibbs 2014). The magnitude of this difference may considerably vary when based on other genomic features. For instance, differences in the proportion and type of insertions/deletions, repetitive elements, segmental duplications, non-protein-coding regions, copy-number variants, loss-of-function variants, gene expression or methylation patterns have provided various estimations of the resemblance between human and other primates (Khaitovich et al. 2006; Jiang et al. 2007; Marques-Bonet, Kidd, et al. 2009; Wall 2013; Rogers and Gibbs 2014; Sonay et al. 2015).

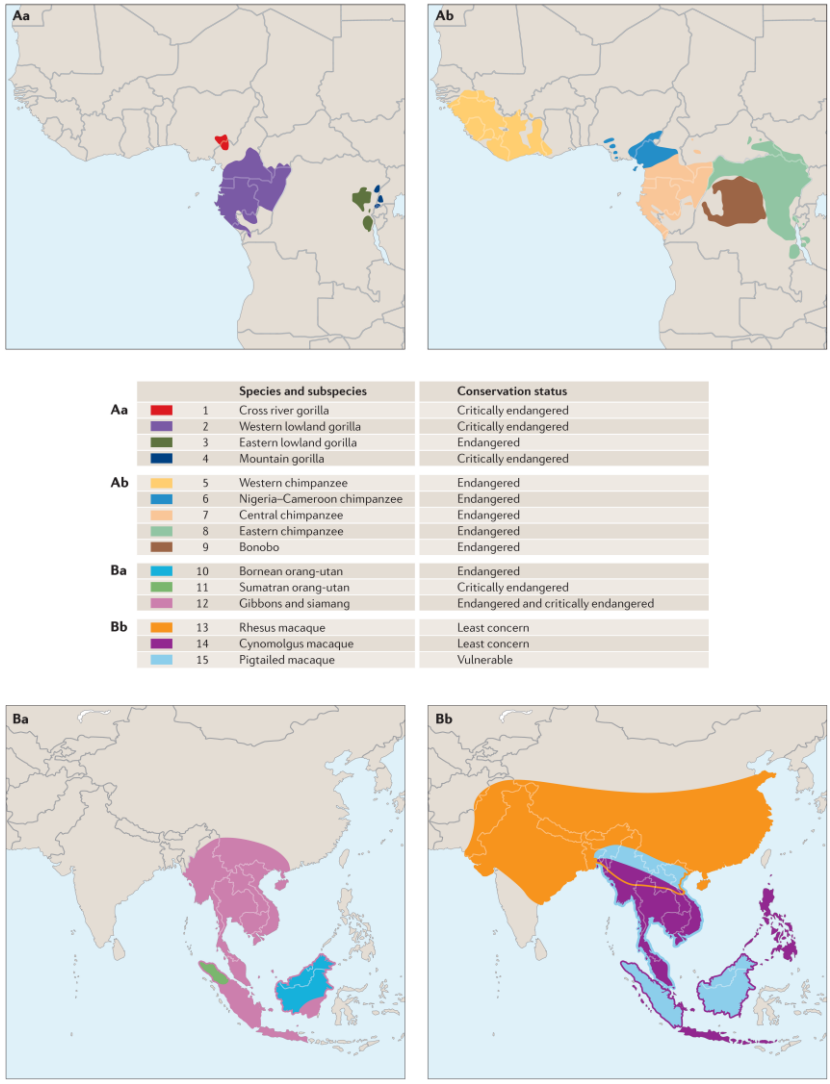


Figure 15. Geographical distribution of great apes and macaques. Approximate species and subspecies population distribution of gorilla **(Aa)**; chimpanzee and bonobo **(Ab)**; orangutan, gibbon and siamang **(Ba)**; and rhesus, cynomolgus and pigtailed macaque **(Bb)**. The conservation status of each group is indicated, based on the current classification of the Red List of Endangered Species from the International Union for Conservation of Nature and Natural Resources (modified from Rogers & Gibbs 2014).

Nevertheless, regardless of which genetic properties are taken into account, genomic similarities among great apes and primates are widely extended. This likeness at the genomic level is of special interest regarding the marked phenotypic differences easily observed among populations and species. The genetic mechanisms that could explain these phenotypic disparities still remain largely unknown, but it seems clear that diversity among great apes is probably due to a small number of changes. Gene expression variation may play a relevant role as well. For example, the most significant difference in the expression patterns between humans and chimpanzees was observed in testis. Also, genes that are highly expressed in the brain have changed more rapidly during human evolution (Khaitovich et al. 2006; Marques-Bonet, Ryder, et al. 2009). Nevertheless, the biological relevance of these changes is difficult to infer. Some authors have highlighted the important contribution of more complex features such as the structure, dynamics and function of genomes (Rogers and Gibbs 2014). In particular, much attention has been focused on the role that regulatory elements and the non-coding genome may have had in shaping animal diversification and evolution (King and Wilson 1975; Taft et al. 2007; Nowick et al. 2009; Morris and Mattick 2014; Perdomo-Sabogal et al. 2014). It is worth noting that, although comparative genomics have provided fundamental information about primate evolutionary processes, most of these approaches still need to be supported by functional and expression analyses.

Being part of both categories (regulatory elements and non-protein-coding RNAs), and as briefly reviewed in these chapters, there is substantial evidence to suggest that miRNAs are crucial factors that influence phenotypic variation and are involved in disease and evolution. miRNA acquisitions and expansions have been dated at the same time as important bauplan modifications that occurred during animal evolution (Hertel et al. 2006; Sempere 2006; Heimberg et al. 2008; Christodoulou et al. 2010). Remarkably, the emergence of primate species has been linked to a dramatic increase of the miRNA repertoires within this group (Du et al. 2013; Iwama et al. 2013). In

this regard, it has been estimated that more than two thirds of all annotated human miRNAs are primate-specific (Iwama et al. 2013; Santpere et al. 2016). Furthermore, several studies have shown that miRNA regions have been constrained and have evolved under purifying selection (Saunders et al. 2007; Quach et al. 2009; Wheeler et al. 2009). Curiously, in spite of their biological relevance, little is known about the functional and evolutionary consequences of miRNA variation and evolution in the human and primate lineages.

In summary, the study of miRNA variation among primates and great apes could shed light in two different senses. On the one hand, it could assist with the identification of particular genetic factors and mechanisms that play important roles in human and non-human species evolution (Hertel et al. 2006; Heimberg et al. 2008; Quach et al. 2009; Brameier 2010; Dannemann et al. 2012). On the other hand, it could be useful for understanding how these gene regulators evolve and how changes at molecular scales (i.e. those occurring within miRNAs) may involve changes at larger scales (i.e. those happening at cellular, tissue, organismal or population levels), in an attempt to understand how micro and macroevolution are interrelated (Farh et al. 2005; Ason et al. 2006; Krol et al. 2010; Berezikov 2011).

II. OBJECTIVES

6. General Objective

The main objective of this thesis was to investigate the functional implications of microRNA nucleotide sequence variation, produced either by DNA genomic changes or RNA editing post-transcriptional modifications, and its contribution on primate evolution. Motivated by this aim, we defined the following specific objectives.

7. Specific Objectives

The first chapter of the Results section is based on the study of microRNA variation produced by DNA genomic nucleotide changes in great apes. The specific objectives of this section are listed below:

- To compare patterns of intra- and inter-species genomic sequence conservation in human microRNAs present in great apes, both among individual microRNA regions and among groups of microRNAs classified according to different genomic and functional features.
- To construct a comprehensive catalogue of microRNA species-specific nucleotide substitutions present in great apes based on the genomic sequences of 82 individuals including gorillas, orangutans, bonobos, chimpanzees and humans.
- To analyze the expression and functional implications of human-specific nucleotide substitutions located within the mature or seed microRNA regions of several microRNAs present in great apes.

The second chapter of the Results section is based on the study of microRNA variation produced by RNA editing modifications in great apes and other primates. The specific objectives of this section are listed below:

- To explore the frequency of A-to-I microRNA editing changes in whole small RNA samples from different primate species and tissues including human placenta and macaque, gorilla, chimpanzee and human brain cortex.

- To compare the A-to-I RNA editing frequency of the microRNA mir-376a1, both in the primary and mature microRNA stages, in samples from different primate tissues at the intra and inter-species level.
- To investigate the possible causes of differences in the editing patters among samples, in the case said differences are observed.
- To evaluate the possible epistatic effects among edited sites in mir-376a1 and the contribution of the A-to-I transitions to the stability of the microRNA hairpin.

III. RESULTS

8. Functional implications of human-specific changes in great ape microRNAs

Alicia Gallego, Marta Melé, Ingrid Balcells, Eva García-Ramallo, Ignasi Torruella-Loran, Hugo Fernández-Bellon, Teresa Abelló, Ivanela Kondova, Ronald Bontrop, Christina Hvilsom, Arcadi Navarro, Tomàs Marquès-Bonet, Yolanda Espinosa-Parrilla

Published

Functional Implications of Human-Specific Changes in Great Ape microRNAs.

PLoS ONE 11(4): e0154194. doi:10.1371/journal.pone.0154194

Supplementary data for this paper can be found at the Supplementary Material section of this thesis. A copy of this manuscript and the supplementary material are available in the CD attached to the thesis booklet and also on the web: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0154194

9. microRNA editing across human tissues and primate species: the case of mir-376a1

Alicia Gallego, Diego A. Hartasánchez, Marina Brasó-Vives, Eva Garcia-Ramallo, Maria Lopez-Valenzuela, Neus Baena, Miriam Guitart, Hugo Fernández-Bellon, Ivanela Kondova, Ronald Bontrop, Yukio Kawahara, Yolanda Espinosa-Parrilla

Submitted for publication

Supplementary data for this paper can be found at the Supplementary Material section of this thesis. A copy of this manuscript is also available in the CD attached to the thesis booklet.

microRNA editing across human tissues and primate species:the case of mir-376a1

Alicia Gallego[1], Diego A. Hartasánchez[1], Marina Brasó-Vives[1], Eva Garcia-Ramallo[1], Maria Lopez-Valenzuela[1], Neus Baena[2], Miriam Guitart[2], Hugo Fernández-Bellon[3], Ivanela Kondova[4], Ronald Bontrop[4], Yukio Kawahara[5], Yolanda Espinosa-Parrilla[1,6,*]

[1] Institute of Evolutionary Biology (IBE) (Universitat Pompeu Fabra-CSIC), Barcelona, Spain.
[2] Genetics Laboratory, UDIAT-Centre Diagnòstic, Corporació Sanitària Parc Taulí-Institut Universitari, Barcelona, Spain.
[3] Parc Zoològic de Barcelona, Barcelona, Spain.
[4] Biomedical Primate Research Centre, Rijswijk, Netherlands.
[5] Graduate School of Medicine, Osaka University, Osaka Japan.
[6] School of Medicine, University of Magallanes, Punta Arenas, Chile.

* Corresponding author: yolespinosa@gmail.com

ABSTRACT

microRNAs are crucial regulators of gene expression that are highly conserved among species. Post-transcriptional modifications such as RNA editing may have important roles in introducing genetic variation and determining phenotypic differences among tissues, individuals and species. However, defining what proportion of the total microRNA editing is functional is still under debate. To better understand the biological relevance of microRNA editing we performed an exhaustive characterization of A-to-I site-specific editing patterns on mir-376a1, a mammalian microRNA for which RNA editing is involved in development and disease. Thorough an interdisciplinary approach based on high throughput small RNA sequencing, Sanger sequencing and *in silico* simulations we explored microRNA editing in primate samples including human placenta and macaque, gorilla, chimpanzee and human brain cortex. We observed that mir-376a1 editing is a common phenomenon in brain compared to placenta, showing more variability among human tissues than among primate species. Primary mir-376a1 is edited at three positions, -1, +4 and +44, from which the +4 site, located at the seed region of the mature miR-376a1-5p, reached the highest editing frequency in all tissues and species analyzed, and conferred the highest stability to the hairpin molecule. We suggest that molecular stability might partly explain the editing recurrence observed in certain microRNAs and that editing events conferring new functional regulatory roles in particular tissues and species could have been conserved along evolution, as it might be the case of mir-376a1.

INTRODUCTION

RNA editing is the site-specific modification of an RNA sequence generating a different product from that encoded in the DNA. Most RNA editing in metazoans is the conversion from adenosione to inosine, which has similar base-pairing properties as guanosine, A-to-I(G), catalyzed by adenosine deaminases (ADARs) acting on double-stranded RNAs (dsRNAs). [1,2] RNA editing can alter the encoded amino acid sequence in certain messenger RNAs (mRNAs) giving isoforms with novel functions. [1,3–6] Despite the biological impact that these changes may introduce at the protein level, RNA editing mainly occurs in non-protein-coding regions, although its functional consequences remain unclear for most of the cases. [1,7] microRNAs (miRNAs), a small class of non-coding RNAs, are crucial regulators of gene expression involved in almost all biological processes, that are highly conserved among species. [8–10] They are transcribed in a primary miRNA which is processed by Drosa giving a ~70 nt precursor hairpin molecule that is cleaved in the cytoplasm by Dicer in a miRNA duplex formed by two ~22 nt mature miRNA strands, -5p and -3p. Target gene repression is based on the sequence interaction between one of the mature strands (especially nucleotides 2-8, known as "seed region") and the mRNA, [11] which leads to mRNA translational repression or degradation. [12–14] Post-transcriptional modifications such as miRNA editing may have an important role in introducing genetic variation and determining phenotypic differences among tissues, individuals and species. [9] Since miRNAs adopt a dsRNA secondary structure along their maturation process, they are naturally ADARs' substrates and could be edited with no significant effects. Nevertheless, previous studies have shown that miRNA editing can alter the biogenesis from primary to mature molecules [15–17] and remodel the spectrum of regulated target genes when they occur at the seed region. [2,18]

Quantification studies of miRNA A-to-I editing have been chiefly based on two approaches. On the one hand, most recent literature has

taken advantage of the large amount of data generated by small RNA next generation sequencing (NGS) technologies, which have permitted to draw a general picture of miRNA editing in the transcriptome and to characterize editing patterns across tissues, species and several biological conditions. [19–23] Regardless of the enormous information these studies have provided, little consistency is observed among results. One of the main restrictions of this approach is the necessity of applying stringent filters during the RNA-DNA mapping in order to conserve genuine editing changes and to exclude other sources of variation such as sequencing errors, single nucleotide polymorphisms (SNPs) or somatic mutations. [19,24–27] Due to these constraints, editing frequencies are normally underestimated and lowly expressed and/or lowly edited miRNAs remain undetectable. [28] On the other hand, approaches based on direct Sanger sequencing of specific genomic regions have estimated that miRNA editing may be widespread, reporting strong evidences for several miRNAs. [17,25,29] Although these low-throughput methods are highly laborious, they allow for a more accurate calculation of editing percentages in one or few nucleotide sites. [25] In fact, direct sequencing has been commonly used for validating editing patterns observed by NGS studies. [19,28,30,31] Other methods based on pyrosequencing technologies have also been used for RNA editing quantification. [32] Even though they are fast and accurate tools for quantification of relatively long fragments, current platforms are limited for short fragment analysis, as mature miRNA molecules.

Analyses of miRNA editing frequencies have evidenced the relevance of these changes, nevertheless the particular molecular mechanisms originating and regulating editing events are still largely unknown. Examination of edited sequences has revealed that most edited adenosines residues are located within U<u>A</u>G triplets. [2,17,19,20] It has also been demonstrated that the competition of ADAR binding proteins for the access to the dsRNA sequences can regulate the editing efficiency at a given adenosine position. [33] Additionally, comparative analyses of miRNA editing have evidenced the existence of regular

patters of editing among closely related species and tissues, suggesting a role of these post-transcriptional modifications in evolution, tissue differentiation, development and disease. [18,23,34–36] For instance miRNA editing appears to be enriched in neural tissues, in older individuals and in healthy compared to cancerous samples. At least six miRNAs (mir-376a-1, mir-376b, mir-376c, mir-379, mir-381 and mir-411) showed editing conservation across placental mammals and two miRNAs (mir-301 and mir-455) between mammals and bony fishes. [23] Within these miRNAs, mir-376a1 is a placental mammalian miRNA whose RNA editing has been demonstrated to be involved in development and disease. [34,37] It is located in the large mir-376 cluster in human chromosome 14, expressed in developing embryos and adult tissues, such as placenta and brain, and involved in neural regulation. [38,39] Alike other members of this cluster, mir-376a1 is mainly edited at three adenosine residues: the -1 site located in the precursor sequence, and the +4 and +44 sites in the seed regions of -5p and -3p strands, respectively. Editing at these three sites seems to be tissue-specific and it can completely redirect the set of recognized target genes when it occurs at the +4 site of the -5p strand. [18]

In this work went deeper into the biological significance of miRNA editing through an exhaustive characterization of A-to-I site-specific editing patterns of mir-376a1 using whole small RNA NGS and Sanger sequencing in a set of primate samples including human placenta and macaque, gorilla, chimpanzee and human brain cortex. We observed different patterns of RNA editing depending on tissue identity and miRNA maturation stage. Editing frequency estimations and transition probability simulations revealed that certain mir-376a1 sites are more frequently edited than others and that this could be explained because they confer more stability to the miRNA hairpin molecule. We suggest that, although miRNA editing could have neutrally emerged as a consequence of ADARs' editing activity over dsRNAs, some of these changes could have been maintained if they conferred a particular molecular advantage to the miRNA such as a better thermodynamic stability. Additionally, editing events that would

involve new functional regulatory roles to their miRNAs in particular tissues and species could have been conserved at the pos-transcriptional level.

RESULTS AND DISCUSSION

mir-376a1-5p and -3p show significant expression in human placenta and primate brains

Expression levels of miR-376a1 mature forms were first evaluated by RT-qPCR in samples of human placenta and human, chimpanzee, gorilla and macaque brain cortex. Both mature miRNAs, miR-376a1-5p and miR-376a1-3p (henceforth -5p and -3p), were similarly expressed between them and among all primate brains, while -3p was 2.9 times more expressed than -5p in human placenta. Expression of -3p in placenta was 3.7 and 3.6 times higher than expression of -5p and -3p in brain, respectively (**Table S1**). Note that mature miR-376a1-3p is identical to miR-376a2-3p and has high sequence similarity with miR-376b-3p, therefore observed expression of -3p is expected to include expression of these three miRNAs (**Fig. S1**). We next examined expression of miR-376a1 mature molecules through analysis of small RNA sequencing (small RNA-seq) data including the samples mentioned above. Although whole small RNA-seq data came from only one representative individual (human placenta and non-human primate brain cortex samples) or two (human brain cortex samples) and general conclusions should not be extracted, the observed expression patterns of miR-376a1 mature molecules were consistent with those observed by RT-qPCR. Except in one human brain sample, in which miR-376a1-5p was almost two times more expressed than -3p, expression of both matures was more similar among primate brain samples compared to placenta (with read counts ranking from 86 to 495). In placenta, both mature miRNAs showed higher expression than in brain, particularly the -3p strand with 1,318 read counts (**Table S2**). Even though miR-376a1-3p is considered the main functional strand (or major strand) and that to our knowledge -5p has not yet been described as a functional molecule in chimpanzee and gorilla, [40]

our data shows that both mature miRNAs, -5p and -3p, are significantly expressed in brain cortex of the studied species and in human placenta, suggesting regulatory roles for both mature forms in these species.

Editing patterns in mature mir-376a1 strands among different tissues and species

A-to-I miRNA editing frequencies in mature molecules were first evaluated in the small RNA-seq dataset. Since inosine in the RNA is transcribed into guanosine during reverse-transcription, RNA edited sites can be identified as A-to-G mismatches in the RNA-DNA mapping. The ratio of edited miRNAs (represented by at least two sequencing reads) among the total number of miRNAs was about 0.12 (SD = 0.01) in all samples, showing that miRNA editing was similarly frequent among all tissues and species analyzed. We then focused in mir-376a1 editing and observed that editing frequencies in the -5p strand were relatively low in all primate brains and particularly in placenta (editing frequencies between 0.7 and 2.6%). Editing in the -3p appeared to be widespread (frequencies between 69.3 and 86.6%), although sequencing reads for this miRNA were probably including both, the miR-376a1-3p and miR-376a2-3p, mature miRNAs (**Table 1**).

In general, the observed editing frequencies for miR-376a1-5p were similar to previous studies based on NGS that reported a miR-376a1-5p editing frequency of 1-5% in human brain and in a human dataset including 410 samples from 24 different tissues, disorders and cell lines. [21,23] However, other studies show an editing frequency of 18.9% for miR-376a1-5p in human brain [20] and 20-29% in mouse brain. [20,34] On the other hand, as far as we know, no information has been published regarding editing percentages in the mature miR-376a1-3p observed by NGS. Discrepancies among editing studies based on NGS analysis have been largely discussed and may be due to differences in the filtering criteria for calling RNA editing, in the sequencing depth of the RNA-seq data, in the reference genomes used during the RNA-DNA mapping, in the type of aligner applied, or the

consideration or not of paralogous sequences, among others. [19,24,25,27,41] To overcome these possible limitations we next analyzed the editing frequencies of miR-376a1 in our samples using Sanger sequencing. Even though Sanger sequencing is more time-consuming (RT-PCR product cloning and Sanger sequencing of a large number of clones are needed for every miRNA sequence), this approach is thought to achieve more precision in the estimation of RNA editing frequencies when a small number of nucleotide sites are being considered. [19,25,28,30,31]

Sanger sequencing was performed on miR-376a1-5p since it has a unique sequence among related paralogs (**Fig. S1**), which allows for a more confident analysis. Furthermore, functional studies have also evidenced the crucial role of editing in miR-376a1-5p, whose edited and non-edited versions specifically silence different target genes such as *SFRS11* (involved in mRNA splicing), *SLC16A1* (a gene with central roles in metabolic pathways), *TTK* (associated with cell proliferation), *PRPS1* (implicated in nucleotide synthesis), *ZNF513* (involved in retinal development), *SNX19* (playing crucial roles in intracellular trafficking), and *RAP2A* and *AMFR* (involved in glioma cell invasion) . [18,37] Moreover, we observed expression of miR-376a1-5p in the brain cortex of chimpanzee and gorilla brain, two species in which this miRNA had not been described before, and for which the observation of editing changes could provide insights into its biological relevance in brain functions. We thus carried out Sanger sequencing over 20 colonies in samples from two human brains and one human placenta and observed that both brain samples reached 25% of editing at the +4 site located at the miR-376a1-5p seed region. The placenta sample was edited in 16% of the cases at this +4 site. Interestingly, both tissues presented higher editing frequencies when they were analyzed using Sanger sequencing instead of NGS, considering both, our data (**Fig. 1**) and published work. [20,21,23] Although there is no available information of RNA editing frequencies analyzed using Sanger sequencing for this miRNA neither in brain cortex nor in placenta, previous studies observed 41% of editing at +4

in miR-376a1-5p in human medulla oblongata, [18] which is in agreement with our finding of higher editing frequencies when Sanger sequencing is used instead of whole small NGS, at least for this mature miRNA. However, other authors found direct Sanger sequencing an inefficient method for editing quantification when editing levels are low or miRNAs are weakly expressed. [20,34] Overall, there is still an important debate on how extensive the phenomenon of RNA editing is in the whole genome and in particular genes and miRNAs, as well as which are the most accurate methods for editing quantification. [24]

Primary mir-376a1 is differentially edited in brain and human placenta

In order to elucidate the editing frequency discrepancies observed between NGS and direct sequencing and to investigate a possible relationship between editing in the mature and primary molecules, we analyzed the patterns of editing at different sites of the primary mir-376a1 molecule. Since primary miRNAs were not present in our small RNA-seq dataset, we evaluated editing in the primary molecule using Sanger sequencing over 20 colonies in six human brain cortex and six human placenta samples, as well as in brain cortex samples from two chimpanzees, one gorilla and one macaque. Editing was detected at the three previously reported positions (-1, +4 and +44) and no new editing sites were observed. Although we saw large inter-individual variation, general patters of editing could be described. Human placenta was shown to be less edited (5-10.8%) than all primate brain cortex samples (15.3-40.5%) and, regarding the different editing sites, we observed that the adenosine located at position -1 was the least edited one, while the adenosine at +4 was the most edited one in all the samples (**Fig. 2** and **Table S3**).

Previous works based on Sanger sequencing from different clones of the primary mir-376a1 reported editing frequencies in human brain of 16%, 46% and 41%, respectively at -1, +4 and +44 sites. [18] Other studies have estimated 21-30% and 51-60% of editing respectively at the +4 and +44 sites in human brain and 21-30% at both sites in

human placenta, [2] although these estimations were calculated comparing the A-G peak heights overlapping at the same site in the sequencing chromatograme, a method that may introduce a significant quantification error. Despite existing differences in editing frequencies observed in the primary mir-376a1 with respect to previous studies based also on Sanger sequencing in several clones, [18] which are probably due to inter-individual variation, the editing progression along the three sites remains constant. For all primate brain cortex and placenta samples, +4 is the most frequently edited site followed by +44 and -1 sites. This pattern is also maintained in human medulla, amygdala and lung. [18] In addition, according to the Sanger sequencing, editing frequencies at the +4 site are similar between the mature -5p and the primary forms. These results could indicate that editing patters are acquired during the first stages of the miRNA maturation process and maintained through it in mir-376a1-5p. Editing frequencies at the +44 site of the -3p strand were lower in the primary mir-376a1 observed by Sanger sequencing (**Fig. 2** and **Table S3**) than in the mature miR-376a1-3p observed by NGS (**Table 1**). Since in the latter we currently do not know which primary mir-376a (a1 or a2) is the main source of the mature miR-376a1-3p, this discrepancy could be explained if the primary mir-376a2 is more expressed and also more edited at the +44 site than the primary mir-376a1. Previous studies have estimated that positions -1, +4 and +44 of the primary mir-376a2 are edited with frequencies of 55%, 90% and 98%, respectively, [18] which would indicate that mir-376a2 sites are more edited than those in mir-376a1, at least in the primary miRNA stages.

Editing at the +4 site confers the highest stability to the hairpin mir-376a1 molecule

Since editing in the primary mir-376a1 occurs at three different sites, we wanted to assess if editing at one position may be related with editing at the other positions. We first analyzed the frequency of the eight possible edited combinations (isoforms) in each sample and observed that the majority of primary isoforms were edited in the primate brain cortex samples (62.9%, on average) while only the

22.5% of isoforms were edited in human placenta. Among edited isoforms, the most common ones were those with only one edited sited (i.e. AAG, AGA and GAA) in all samples. Editing at two positions (i.e. AGG, GAG and GGA) was less frequent; and the isoform with three edited sites (i.e. GGG) was extremely rare and present only in three brain samples from two human (H1, H5) and one chimpanzee (C2) individuals. Again, we observed high inter-individual variation in the number of edited isoforms among samples (**Fig. 3** and **Table S4**). Nevertheless this variation seems to be higher within species (brain and placenta human samples) than across species (all studied brain samples), a pattern that should be confirmed through broader analyses including more tissue samples (e.g. placenta samples from non-human primates) that could not be studied here due to the difficulty in obtaining these samples. Previous analyses on A-to-I editing in Alu sequences among primates revealed that these regions are more frequently edited in humans than in chimpanzees and macaques. [42] As authors argue, the inter-species editing variation in this case could be the result of differences in the location and orientation of Alu elements in the genomes and the dsRNA structures that they form in each species. [42] In the particular case of mir-376a1, the hairpin structure resemblance among the studied primates might explain the observed editing frequency similarities across species. Nevertheless, some other mechanism, as tissue specificity in the expression levels of ADARs, [43] may explain editing differences among tissues observed by us and other studies. [15,18,21,23]

Given that specific isoforms appear to be more frequent than others and that the relative frequencies also differ among samples, we next carried out computer simulations in order to identify which are the editing rates at each of the three positions (-1, +4 and +44) that could best account for the observed editing patterns. Our model considered that editing occurs at site-specific constant rates and independent of the previous edition of another site. For each sample (placenta from human and brain cortex from human, chimpanzee, gorilla and macaque) we ran 4,851 sets of simulations (each set consisting of

1,000 independent runs) to cover the whole range of possible editing rates. Each simulation involved a population of 1,000 primary mir-376a1 sequences and a set of editing rates (adding up to one) assigned to each of the three positions. At each time step one randomly chosen primary sequence was edited at one of the three positions according to the assigned rates. We then calculated the proportion of edited isoforms obtained for every sample (and set of editing rates), and selected the set of rates that minimized the difference between the simulated and the observed data. Simulations were performed for each species and tissue; and also for the group of all non-human primate brain cortex samples together in order to have a similar number of individuals as human brain cortex and placenta cases.

In all samples except in human placenta, the editing rate at position +4 was higher than that at positions -1 and +44. On the other hand, in all cases except in chimpanzee brain cortex, the editing rate at position -1 was the lowest edited site (**Fig. 4** and **Table 2**). Having observed that editing at the three positions occurred at different rates, we further tested for possible epistatic effects among these sites. In order to asses if editing at one site may be related with editing in the remaining sites we compared the editing rates in each site and sample calculated in two ways. First, we considered the fraction of isoforms edited at only one particular position (i.e. "AAG", "AGA", and "GAA") over the total number of isoforms edited in one position, which gave us the editing probability independently for each position (**Table S5A**). Second, we considered the fraction of editing events (i.e. the number of A-to-G transitions) that occurred in a particular position over the total number of editing events in all positions, which gave us the editing probability for each position taking into account the editing probabilities in the remaining positions (**Table S5B**). Finally, we analyzed the relative difference between both estimations (**Table S5C**). If editing probabilities were not dependent among sites we would expect to find no significant differences between both estimations. Although we could only analyze a limited number of individuals and no general conclusions should be made, we observed

that the relative difference values for the three positions were very low for all samples (**Table S5C**), especially for human and chimpanzee brain cortex. Altogether, these observations suggest that the three sites are independently edited in the primary mir-376a1, although the simulated data suggest that they are edited with different probabilities. Variations among editing rates at the three sites could be explained if, for instance, ADARs would preferentially edit +4 and +44 sites instead of -1. As previous studies have shown, edited adenosines are most often located within UAG triplets, [2,17,19,20] which is the case of the +4 and +44 sites of mir-376a1. These two positions are probably more suitable target sequences for ADAR recognition than adenosine at -1, present in an AAG triplet. Nevertheless, some other features should be affecting these rates since position +4 is more frequently edited than position +44. As it has been shown, primary miRNAs are more often edited at the -5p than at the -3p strand and more distal to the Dicer cleavage site next to the hairpin loop. [17] Also, edited adenosine residues tend to locate in front of a cytosine, [2,17] which has been proposed as a mechanism that confers molecular stability to the miRNA. [44] These facts may provide feasible explanations as to why +4 is the most edited site; however, the mechanisms behind these observed patterns remain poorly understood. Observed editing tendencies could be additionally explained by some still unknown regulatory mechanisms of the editing process that could dictate the editing efficiency at a given adenosine in a tissue-specific manner, as has been recently suggested. [45] The noted differences in the mir-376a1 editing rates among sites could also be a reflection of the primary molecule's lifetime, being those isoforms edited at the +4 site more abundant than those edited at the -1. Primary lifetimes mostly depend on the Drosha processing speed in order to yield precursor miRNAs; as well as on the degradation time, which also depends on the secondary structure stability of the precursor hairpin. To address this, we calculated the minimum free energy (MFE) of all eight isoforms of human precursor mir-376a1 secondary structures using RNAfold. [46] Although some studies have reported that inosine does not exactly have the same base-pairing properties as guanosine, [47] MFE values

were calculated using the guanosine as the edited nucleotide since RNAfold does not take into account the significant existing similarities between inosine and guanosine bases. Remarkably, all isoforms in which the +4 site was edited showed the highest stabilities, even higher than the non-edited precursor molecule (**Table 3**). This is a feasible observation since the +4 adenosine is in an A-C mismatch, and its editing would confer a more stable G-C pair (**Fig. S2**).

We hypothesize that, since editing at position +4 confers much more stability to the precursor miRNA, these isoforms would be more abundant in the cell, which would be in agreement with previous work indicating that editing might play a role in controlling dsRNA stability. [28] In fact, molecular stability could be the main reason explaining why simulations pointed out that the editing rate at the +4 site was the highest among the three sites. ADARs' editing activity itself cannot explain the difference between editing rates at +4 and +44 sites since they are both present within U<u>A</u>G triplets. Regarding MFE values we observed that editing at the +44 site reduces molecular stability in the same proportion as editing at the -1 site (**Table 3**), however in all samples except in chimpanzee brain, editing rates were higher for +44 than -1 (**Fig. 4** and **Table 2**), which could be a consequence of the adenosine at +44 being in a more suitable triplet for ADAR's recognition than adenosine at -1. Regarding brain cortex, which is significantly edited compared to placenta, simulations in both human and non-human primate sets of samples showed that the "AGA" isoform reached the highest frequency among the isoforms edited at one site, while the "AGG" and "GGA" isoforms were the most common ones among those edited at two sites, which again reveals that the +4 site is the main determinant of the primary mir-376a1 editing. Looking at human brain samples, for which we had a more confident dataset, the observed "AGG" isoform appeared at higher frequencies than expected. Since editing rates at positions +4 and +44 seem to be independent of each other (**Fig. 4**), the prevalence of this isoform could be due to slight epistatic effects that could not be

detected by us, an issue that might be confirmed if more individual samples could be analyzed.

In conclusion, although whole small NGS sequencing allows for the analysis of general genomic and miRNA editing patterns in a large set of samples, direct sequencing provided higher editing frequency estimations at different positions and miRNA stages according to our results. Discrepancies among methods in the analysis of miRNA editing modifications stress the necessity of exploring external factors such as the surrounding sequence, the secondary structure stabilities and the possible epistatic relationships among nucleotide residues, which may differ among miRNAs and samples. In this work we aimed to shed light on the limitations of the study of miRNA editing based on an interdisciplinary approach and focusing in the particular case of mir-376a1. We argue that miRNA editing might have arisen neutrally at the first stages of miRNA maturation as post-translational modifications. However, we observed that the most frequent edited isoforms were those that confer a higher stability to the miRNA hairpin molecules, and that these were maintained across species in a tissue-specific manner. We suggest that some editing changes might be conserved if they provide genetic variation and regulatory diversification at particular tissues, individual and/or species. Finally, as other studies have speculated, [42,48,49] if positive selection is favouring edited changes over non-edited ones, the editing patterns that we observed at the RNA level could be foreshadowing a future scenario in which the A-to-I changes would be finally fixed in the genome, raising the question of RNA editing as a potential force for DNA recoding by genetic mechanisms that still would remain to be identified.

METHODS

RNA samples and RNA extraction

Gorilla brain cortex tissue was provided by the Barcelona Zoo (Spain). Chimpanzee and macaque brain cortex tissues were provided by the Biomedical Primate Research Centre (Netherlands). Human placenta

samples were provided by the Corporació Sanitària Parc Taulí-Institut Universitari (Spain). This work was conducted according to relevant Spanish and International guidelines. Primate samples were collected after natural death and thus, in a non-invasive way, without disturbing, threatening or harming the animals. Human placenta samples were collected after delivery from healthy donors who signed an informed consent. Human brain cortex RNA samples were purchased at Ambion (Foster City, CA, USA). Total RNA was extracted with the RNeasy Mini Kit (QIAGEN, Valencia, CA, USA).

Small RNA-sequencing analysis

One μg of total RNA showing RNA integrity numbers between 7-9 were processed using an Illumina protocol considering the following steps: acrylamide gel purification of RNA bands corresponding to whole small RNA ranging 17–30 nt long, ligation of 5' and 3' adapters to the RNA in two separate subsequent steps each followed by acrylamide gel purification, cDNA synthesis followed by acrylamide gel purification, and a final step of PCR amplification to generate template libraries that were sequenced using Illumina HiSeq2000 (Axeq services, Seoul, South Korea). Small RNA-seq data was analyzed using Chimira software, [50] by mapping read sequences against all annotated human precursor miRNAs registered in miRBase [40] using the Standard Nucleotide Blast (BLASTn v.2.2.24+) and allowing up to two mismatches for each sequence and discarding any anti-sense hits. Total sequencing reads were normalized across samples using DESeq2. Sequencing reads mapping to multiple precursor miRNA paralogs were assigned only to the first optimal alignment call returned by BLASTn. To simplify, orthologous miR-376a1 names from all primate species analyzed were named as the human annotation (see corresponding species annotations in **Table S6**). The miRNA editing frequencies in mature molecules were evaluated in the small RNA-seq dataset using the modifications tool analysis of Chimira. [50]

miRNA expression analysis by real time quantitative reverse transcription PCR (RT-qPCR)

Expression levels of mir-376a1-5p and mir-376a1-3p in the different primate tissues were analyzed by RT-qPCR starting from 150 ng of total RNA in a RT final volume of 10 µl, using specific primers (**Table S8**) and based on the protocol previously described. [51] The qPCR was performed using LightCycler® 480 SYBR Green I Master (Roche Diagnostics) according to manufacturer's protocol. On average, qPCR efficiencies were 1.85, 1.84 and 1.72 for mir-376a1-5p, mir-376a1-3p and miR-25-3p specific primers, respectively. Expression level comparisons between mature -5p and -3p strands were calculated based on Pfaffl, [52] which quantifies the relative expression ratio of the studied miRNAs based on the qPCR efficiency and their quantification cycle deviation (ΔCq), comparing with the control reference miRNA, miR-25-3p as previously described. [53]

miRNA editing percentage analysis by Sanger sequencing

Editing at primary mir-376a1 was determined based on the method previously described. [29] First-strand cDNA was synthesized from one µg of total RNA in a volume of 11 µl using the pri-mir-376a1 RV primer (**Table S8**) and 200 U of M-MuLV Reverse Transcriptase (New England). cDNA was PCR amplified using pri-mir-376a1 FW and RV primers (**Table S8**) and specific pri-mir-376a1 products were isolated from agarose gel bands using QIAEX II Gel Extraction Kit (QIAGEN). Editing at mature miR-376a1-5p was determined using Ion Total RNA-Seq Kit v2 (Ion Torrent Life Technologies) following manufacturer's protocol. First-strand cDNA was synthesized from 500 ng of total RNA in a volume of three µl and PCR amplified using Ion 5' PCR kit primer and miR-376a1-5p RV primer (**Table S8**). Mature miR-376a1-5p products were isolated from TBE 5X polyacrylamide gel (Bio Rad) bands using QIAEX II Gel Extraction Kit (QIAGEN). All RT-PCR products were cloned using TOPO® TA Cloning® Kit for Sequencing (Invitrogen) following manufacturer's protocol and PCR amplified using Phusion® Hot Start Flex DNA polymerase (New England) and M13 FW and RV primers

(**Table S8**). Once the presence of the RT-PCR products was confirmed by electrophoresis gel screening, these products were Sanger sequenced by GATC-Biotech using M13 RV primer. After sequencing 50 samples (colonies) of H1 human brain sample and to determine the minimum reliable samples that would return the same proportion of true edited sites in the three edited positions (-1, +4 and +44) than the original set, we performed a standard resampling of 10,000 times in sets of 10 to 40 sequences in steps of 10, and calculated the proportion of editing at the three positions sites in each step. We found that a similar average of edited sites could be ascertained when using a minimum of 20 sequences, thus 20 colonies were sequenced and analyzed in the remaining samples.

Transition probability simulations

Editing rates were calculated based on computer simulations with different sets of editing rates for each position. For each sample and set of editing rates, we ran 1,000 simulations, which involved a population of N=1,000 primary mir-376a1 sequences. At every time step, a randomly chosen miRNA was edited at one of the three described positions (-1, +4, +44) according to the assigned probabilities of each set. Simulations were stopped once the proportion of non-edited miRNAs equalled the observed proportion in each sample. Then, a sample of size equal to the experimental samples was randomly extracted from the population and edited isoforms (seven possibilities) were counted. The proportion of each edited isoform with respect to the total number of all edited isoforms was calculated for each simulation, as well as the proportion average across the 1,000 simulations. For each set of probabilities (and sample), we therefore extracted seven fractional numbers (adding up to one) corresponding to each edited isoform. Difference (Diff) between our simulated data and our observed data was calculated with the following formula:

$$Diff = \sum_{i=AAG}^{isoforms} (observed - simulated)^2$$

The set of simulations was performed 4,851 times to span the whole phase space of possible editing rates in 0.01 intervals. Diff was calculated for each set (and sample) and the set with the minimum Diff was selected for each sample (**Table 2**).

Epistatic relationship simulations

Editing probabilities at each position were first evaluated considering the fraction of isoforms edited at one position over the total number of isoforms edited in one position (1EditProp) (e.g. $p_{+44} = f_{AAG} / (f_{GAA} + f_{AGA} + f_{AAG})$, for position +44) and second, considering the fraction of editing events at a particular position over all editing events (AllEditProp) (e.g. $p'_{+44} = (f_{AAG} + f_{AGG} + f_{GAG} + f_{GGG}) / (f_{GAA} + f_{AGA} + f_{AAG} + 2f_{GGA} + 2f_{GAG} + 2f_{AGG} + 3f_{GGG})$, for position +44). The relative difference (W-value) between both estimations was calculated with the following formula:

$$W = \sum^{sites} \frac{(1EditPop - AllEditProp)^2}{AllEditProp}$$

W-values close to zero would indicate low differences between estimations and lack of epistatic effects among edited sites.

Secondary structure predictions of miRNAs

Minimum free energy (MFE) and secondary structures of mir-376a1 precursor edited isoforms were predicted using RNAfold from ViennaRNA Package (release 2.1.9) [46] and human sequences annotated in miRBase (release, 21).

ACKNOWLEDGEMENTS

This work was funded by "Spanish National Institute for Bioinformatics" (www.inab.org); "Ministerio Economía y Competitividad de España" [grant numbers BFU2012-38236, BFU2010-18477, BFU2009-06974 and CGL2009-09013]; "Direcció General de Recerca de la Generalitat de Catalunya" [grant numbers 2009SGR-1101, 2014SGR-866 and SGR2014-1311]; "European Union Seventh Framework Programme" [grant numbers PIOF-GA-

2009-236836 and PIRSES-GA-2013-612583]; "FEDER European Regional Development Fund: A way to build Europe"; "Ministerio de Educación, Cultura y Deporte de España" [grant FPU-MEC to AG]; "Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan for Scientific Research on Innovative Areas" [grant number 15H01471 to YK] and "Fundació Parc Taulí-Institut Universitari UAB" [grant number FPT 07/2244 to NB and MG].

We would like to thank Teresa Abelló from Barcelona Zoo for helping in the gorilla brain samples collection, and Juan Antonio Rodríguez for his support in the data analysis and the interpretation of the results.

REFERENCES

1. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem 2010; 79:321–49.
2. Blow MJ, Grocock RJ, van Dongen S, Enright AJ, Dicks E, Futreal PA, Wooster R, Stratton MR. RNA editing of human microRNAs. Genome Biol 2006; 7:R27.
3. Burns CM, Chu H, Rueter SM, Hutchinson LK, Canton H, Sanders-Bush E, Emeson RB. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature 1997; 387:303–8.
4. Hoopengardner B, Bhalla T, Staber C, Reenan R. Nervous system targets of RNA editing identified by comparative genomics. Science 2003; 301:832–6.
5. Rosenthal JJC, Bezanilla F. Extensive editing of mRNAs for the squid delayed rectifier K+ channel regulates subunit tetramerization. Neuron 2002; 34:743–57.
6. Ohlson J, Pedersen JS, Haussler D, Ohman M. Editing modifies the GABA(A) receptor subunit alpha3. RNA 2007; 13:698–703.
7. Li JB, Levanon EY, Yoon J-K, Aach J, Xie B, Leproust E, Zhang K, Gao Y, Church GM. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science 2009; 324:1210–3.

8. Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H. Birth and expression evolution of mammalian microRNA genes. Genome Res 2013; 23:34–45.

9. Wheeler BM, Heimberg AM, Moy VN, Sperling E a, Holstein TW, Heber S, Peterson KJ. The deep evolution of metazoan microRNAs. Evol Dev 2009; 11:50–68.

10. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science 2001; 294:853–8.

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

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

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

14. Kim VN. MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol 2005; 6:376–85.

15. Kawahara Y, Zinshteyn B, Chendrimada TP, Shiekhattar R, Nishikura K. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO Rep 2007; 8:763–9.

16. Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R, Nishikura K. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nat Struct Mol Biol 2006; 13:13–21.

17. Kawahara Y, Megraw M, Kreider E, Iizasa H, Valente L, Hatzigeorgiou AG, Nishikura K. Frequency and fate of microRNA editing in human brain. Nucleic Acids Res 2008; 36:5270–80.

18. Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science 2007; 315:1137–40.

19. Peng Z, Cheng Y, Tan BC-M, Kang L, Tian Z, Zhu Y, Zhang W, Liang Y, Hu X, Tan X, et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol 2012; 30:253–60.

20. Alon S, Mor E, Vigneault F, Church GM, Locatelli F, Galeano F, Gallo A, Shomron N, Eisenberg E. Systematic identification of edited microRNAs in the human brain. Genome Res 2012; 22:1533–40.

21. Gong J, Wu Y, Zhang X, Liao Y, Sibanda VL, Liu W, Guo A-Y. Comprehensive analysis of human small RNA sequencing data provides insights into expression profiles and miRNA editing. RNA Biol 2014; 11:1375–85.

22. Nigita G, Veneziano D, Ferro A. A-to-I RNA Editing: Current Knowledge Sources and Computational Approaches with Special Emphasis on Non-Coding RNA Molecules. Front Bioeng Biotechnol 2015; 3:1–7.

23. Warnefors M, Liechti A, Halbert J, Valloton D, Kaessmann H. Conserved microRNA editing in mammalian evolution, development and disease. Genome Biol 2014; 15:R83.

24. Lee J, Ang JK, Xiao X. Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. RNA 2013;

25. Bass B, Hundley H, Li JB, Peng Z, Pickrell J, Xiao XG, Yang L. The difficult calls in RNA editing. Nat Biotechnol 2012; 30:1207–9.

26. Park E, Williams B, Wold BJ, Mortazavi A. RNA editing in the human ENCODE RNA-seq data. Genome Res 2012; 22:1626–33.

27. Levanon EY, Eisenberg E. Does RNA editing compensate for Alu invasion of the primate genome? BioEssays 2015; 37:175–81.

28. Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D, et al. Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol 2004; 22:1001–5.

29. Kawahara Y. Quantification of adenosine-to-inosine editing of microRNAs using a conventional method. Nat Protoc 2012; 7:1426–37.

30. Bahn JH, Lee J, Li G, Greer C, Peng G. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. 2012; :142–50.

31. He T, Wang Q, Feng G, Hu Y, Wang L, Wang Y. Computational detection and functional analysis of human tissue-specific A-to-I RNA editing. PLoS One 2011; 6:e18129.

32. Iwamoto K, Bundo M, Kato T. Estimating RNA editing efficiency of five editing sites in the serotonin 2C receptor by pyrosequencing. RNA 2005; 11:1596–603.

33. Washburn MC, Hundley HA. Trans and cis factors affecting A-to-I RNA editing efficiency of a noncoding editing target in C. elegans. RNA 2016; 22:722–8.

34. Ekdahl Y, Farahani HS, Behm M, Lagergren J, Ohman M. A-to-I editing of microRNAs in the mammalian brain increases during development. Genome Res 2012; :1477–87.

35. Nemlich Y, Greenberg E, Ortenberg R, Besser MJ, Barshack I, Jacob-Hirsch J, Jacoby E, Eyal E, Rivkin L, Prieto VG, et al. MicroRNA-mediated loss of ADAR1 in metastatic melanoma promotes tumor growth. J Clin Invest 2013; 123:2703–18.

36. Shoshan E, Mobley AK, Braeuer RR, Kamiya T, Huang L, Vasquez ME, Salameh A, Lee HJ, Kim SJ, Ivan C, et al. Reduced adenosine-to-inosine miR-455-5p editing promotes melanoma growth and metastasis. Nat Cell Biol 2015; 17.

37. Choudhury Y, Tay FC, Lam DH, Sandanaraj E, Tang C, Ang BT, Wang S. Attenuated adenosine-to-inosine editing of microRNA-376a*promotes invasiveness of glioblastoma cells. J Clin Invest 2012; 122:4059–76.

38. Glazov EA, McWilliam S, Barris WC, Dalrymple BP. Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. Mol Biol Evol 2008; 25:939–48.

39. Seitz H. A Large Imprinted microRNA Gene Cluster at the Mouse Dlk1-Gtl2 Domain. Genome Res 2004; 14:1741–8.

40. Kozomara A, Griffiths-Jones S. MiRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 2014; 42:68–73.

41. Pachter L. A closer look at RNA editing. Nat Biotechnol 2012; 30:246–7.

42. Paz-Yaacov N, Levanon EY, Nevo E, Kinar Y, Harmelin A, Jacob-hirsch J, Amariglio N, Eisenberg E, Rechavi G. Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. Proc Natl Acad Sci U S A 2010; 107:12174–9.

43. Picardi E, Manzari C, Mastropasqua F, Aiello I, D'Erchia AM, Pesole G. Profiling RNA editing in human tissues: towards the inosinome Atlas. Sci Rep 2015; 5:14941.

44. Wong SK, Sato S, Lazinski DW. Substrate recognition by ADAR1 and ADAR2. RNA 2001; 7:846–58.

45. Washburn MC, Hundley H a. Trans and cis factors affecting A-to-I RNA editing efficiency of a noncoding editing target in C. elegans. RNA 2016; 22:722–8.

46. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res 2008; 36:70–4.

47. Kume H, Hino K, Galipon J, Ui-Tei K. A-to-I editing in the miRNA seed region regulates target mRNA selection and silencing efficiency. Nucleic Acids Res 2014; 42:10050–60.

48. Qureshi IA, Mehler MF. Non-coding RNA networks underlying cognitive disorders across the lifespan. Trends Mol. Med.2011; 17:337–46.

49. Tian N, Wu X, Zhang Y, Jin Y. A-to-I editing sites are a genomically encoded G: implications for the evolutionary significance and identification of novel editing sites. RNA 2008; 14:211–6.

50. Vitsios DM, Enright AJ. Chimira: Analysis of small RNA sequencing data and microRNA modifications. Bioinformatics 2015; 31:3365–7.

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

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

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

TABLES

Table 1. Total plain read counts and editing ratios in the mature miR-376a1-5p and -3p. Sequencing reads were mapped against all human miRNAs annotated in miRBase. These data are not normalized, thus total number of sequencing reads are not comparable among samples.

Sequencing reads	Human placenta	Human cortex 1	Human cortex 2	Chimpan zee cortex	Gorilla cortex	Macaque cortex
Sequencing reads -5p	813	189	669	158	234	134
Editing ratio -5p	0.7%	2.6%	1.9%	2.5%	6.4%	2.2%
Sequencing reads -3p	1017	153	374	116	179	88
Editing ratio -3p	86.2%	71.9%	80.2%	77.6%	86.6%	69.3%

Table 2. Editing rates associated to each site and sample based on computer simulations. Editing rates of positions -1, +4 and +44, in the primary mir-376a1, that better reproduce the observed data.

Edited site	Human placenta	Human cortex	Chimpanz ee cortex	Gorilla cortex	Macaque cortex	Non-human primate cortex[1]
-1	0.13	0.17	0.29	0.17	0.11	0.2
+4	0.41	0.46	0.45	0.51	0.64	0.49
+44	0.46	0.37	0.26	0.32	0.25	0.31

[1]Brain cortex samples from chimpanzee, gorilla and macaque.

Table 3. Minimum free energy (MFE) values. MFE values in kcal/mol predicted by RNAfold for each human precursor mir-376a1 edited isoforms. Isoforms are represented by their three editable sites where first, second and third positions correspond to -1, +4 and +44 sites, respectively. Smaller MFE values are associated with higher stabilities.

Edited isoform	MFE
A**G**A	-24.80
GGA	-24.50
A**GG**	-24.50
GGG	-24.20
AAA	-18.20
GAA	-17.90
AA**G**	-17.90
GA**G**	-17.60

FIGURES

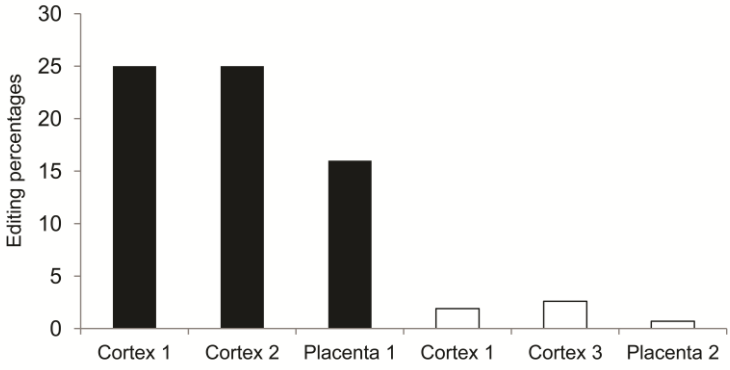


Figure 1. Editing frequencies in the mature miR-376a1-5p measured by Sanger sequencing and NGS. Frequencies obtained by Sanger sequencing of 20 colonies (black) and NGS of small RNA-seq (white) in human brain cortex and placenta samples based on our data.

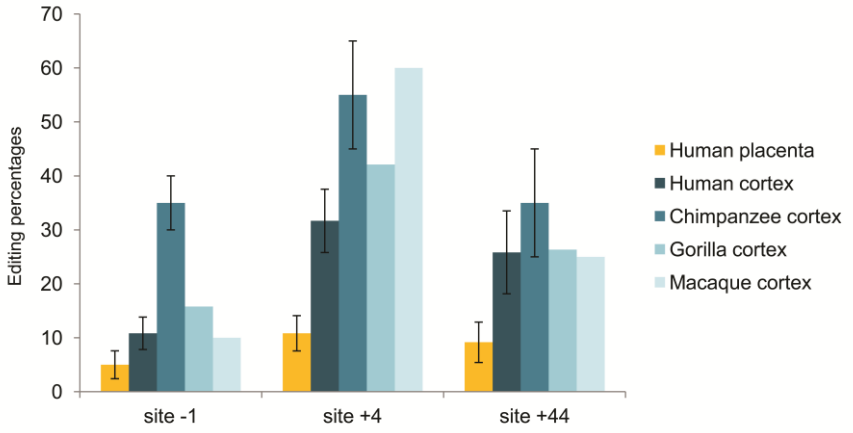


Figure 2. Editing frequencies in the primary mir-376a1 measured by Sanger sequencing. Editing percentage averages in the primary mir-376a1 at the three sites -1, +4 and +44. Error bars represent the standard error in sets of samples including more than one sample.

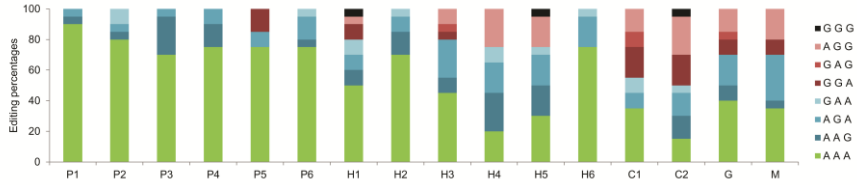


Figure 3. Edited isoform frequencies per individual. Editing frequencies of primary mir-376a1 isoforms evaluated by Sanger sequencing in 20 colonies and shown as percentages. P = human placenta (six individuals); H = human brain cortex (six individuals); C = chimpanzee brain cortex (two individuals); G = gorilla brain cortex (one individual); M = macaque brain cortex (one individual). Isoforms are represented by their three editable sites where first, second and third positions correspond to -1, +4 and +44 sites, respectively.

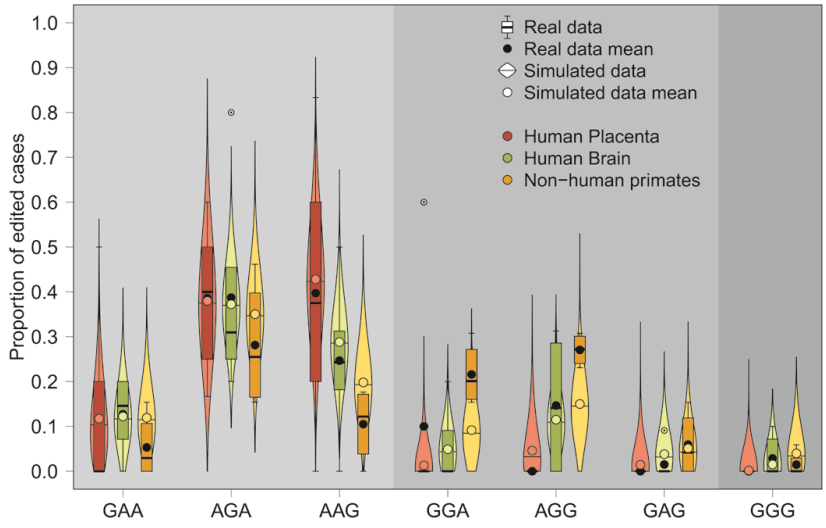


Figure 4. Proportion of observed and expected edited isoforms across samples. Data are shown for human placenta, human brain cortex and non-human primate (chimpanzee, gorilla and macaque) brain cortex samples. Box plots show the proportion of observed edited isoforms with respect to the total number of edited isoforms per sample; outliers are shown as dotted circles; whiskers extend to the 1.5 interquantile range; medians are shown as a short horizontal black line; means are shown as filled black circles. Violin plots show the distributions of the corresponding proportion of edited isoforms from 1,000 simulation runs. The editing rates for each site used in these simulations correspond to those displayed in **Table 2** and refer to those rates that minimize the sum of the differences between the observed and simulated means of each edited isoform; simulation medians are shown as a thin horizontal line; simulation means are shown as empty circles. Isoforms are represented by their three editable sites where first, second and third positions correspond to -1, +4 and +44 sites, respectively.

IV. DISCUSSION

The contribution of miRNA-mediated regulation to phenotypic diversity and the fact that miRNA genes are affected by different sources of variation have been widely demonstrated. Nevertheless, little is known about how miRNA nucleotide changes are ultimately related to diversification in miRNA function and to the evolutionary processes that shaped animal divergence. The motivation of this thesis has been to try to address these questions from two perspectives. In the first chapter of the Results, we investigated the miRNA nucleotide changes that exist among great ape populations and individuals at the genomic level. In the second chapter, the study of miRNA variation centred on the post-transcriptional modifications that emerged through A-to-I RNA editing substitutions across different individuals and primate tissues. Following and expanding upon the line of previous studies, the work in this thesis further supports the strong effect of natural selection on miRNA genes throughout great ape evolution. On the other hand, our research shows the impact that specific nucleotide changes, arising from DNA mutations or RNA post-transcriptional modifications, may have on the stability of the miRNA secondary structure, the expression of the miRNA molecule, and/or the recognition of target genes, which are crucial determinants of the miRNA function.

Beyond the seed region

An important part of our results is the analysis of genomic data from more than 80 individuals representing ten great apes populations, sequenced at high coverage and reported at the time of starting this thesis (Prado-Martinez et al. 2013). Thanks to this valuable information, we have been able to analyze miRNA variation, both among and within populations, something that had previously not been possible to achieve for great apes. After analyzing the annotated human miRNAs (miRBase, release 19) present in these newly available genomes, we unexpectedly observed no differences in the conservation patterns between the miRNA seed region and the whole mature miRNA sequence; looking either inter- or intra-species. The seed region, which shows perfect sequence complementarity with its

corresponding target site in the mRNA, is known to be essential for the miRNA function and different studies have evidenced significant conservation in this particular region (Lewis et al. 2003; Saunders et al. 2007; Selbach et al. 2008; Peterson et al. 2014). Some authors have also proposed that other specific positions along the mature region may also be functionally crucial, based on a lack of variation at certain sites that are mainly situated at the downstream nucleotides next to the seed (Quach et al. 2009; Wheeler et al. 2009). Nevertheless, our results try to go one step further and suggest that the entire mature miRNA molecule could constitute a single unit of selection, being significantly more conserved than the rest of the precursor hairpin. This observation could somewhat change our perspective in the study of miRNA function and evolution by taking the general view beyond the relevance of the seed. Over and over, the miRNA field has established its own *central dogma* regarding the seed region. Certainly, it is accepted that some parts of the mature sequence seem to be indispensable, either because they could compensate seed mismatches with the target site or because they provide optimal stability to the hairpin structure (Grimson et al. 2007b). However, all these arguments have revolved around the idea of supporting the activity of the seed. As recently reviewed, the simplification of miRNA molecules as fixed structures constituted by well-defined functional elements remains incomplete and needs to move into a more heterogeneous space that includes all the diversity and variability that is in fact associated with miRNAs (Desvignes et al. 2015). Remarkably, a single miRNA gene could give rise to the canonical -5p and -3p mature strands and their multiple related isomiRs, as well as to other processed products like loRs or moRs (sequences derived from, respectively, the loop or the 5' and 3' extremes of the precursor miRNA hairpin), all of which are potentially functional (Berezikov 2011; Desvignes et al. 2015) (**Figure 16**). Additionally, as briefly mentioned in the Introduction of this thesis, increasing reports have demonstrated that, apart from canonical miRNA targeting, almost all types of miRNA-mRNA interactions can be functionally relevant (Zisoulis et al. 2010; Schnall-Levin et al. 2011; Kosik 2013; Hausser and Zavolan 2014; Zhou and Rigoutsos 2014).

We think that our observations highlight the importance of considering the entire mature region in the study of the miRNA function, and perhaps more relevantly, enrich the increasing perception of miRNA genes as dynamic regulators, as opposed to simple "seed carriers", that can be differentially processed and bind a wide range of target sites through different domains.

We were interested in further exploring if any nucleotide position within the mature miRNA region shows higher levels of conservation than the resting sites. We could not test for this possibility before

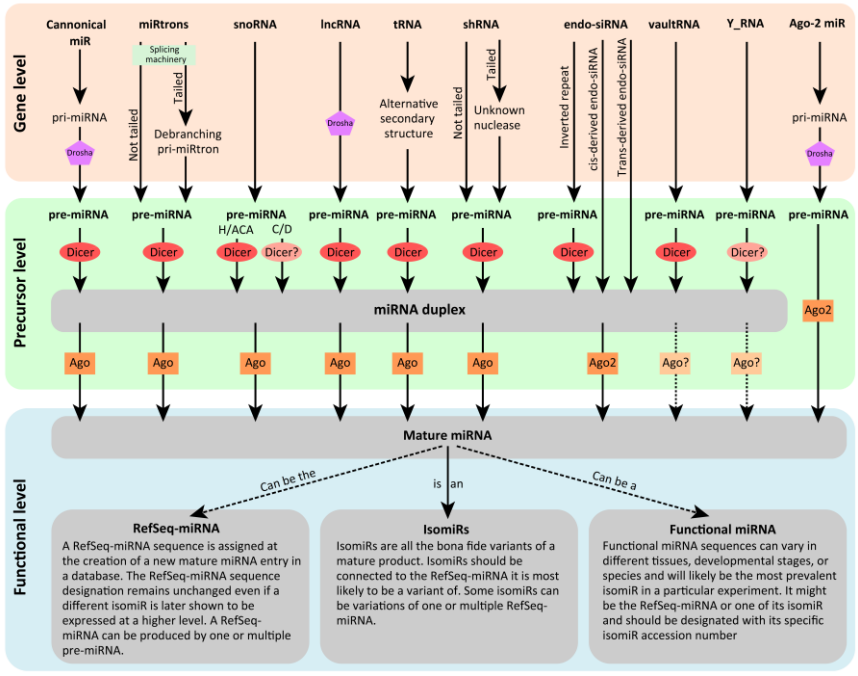


Figure 16. Diverse genetic origins and maturation pathways leading to functional miRNAs. Apart from canonical miRNAs, many different types of ncRNA genes, such as snoRNAs, lncRNAs and tRNAs, can be transcribed and differentially processed producing miRNA mature duplexes with regulatory functional capacities. The authors of this work propose a new miRNA classification system that includes all the miRNA variability. Functional experiments are crucial in order to distinguish true miRNAs from RNA degradation products (adapted from Desvignes et al. 2015).

completing this thesis; in any case, further analyses could clarify if the whole mature sequence is evolutionarily relevant or, in contrast, if just a few sites account for a huge selective constrain. Moreover, as previously observed in humans, by analysing many more individuals from the same population, we could better characterize the frequency of every allele variant, increasing our knowledge about the relevance of the studied nucleotide changes and possibly local adaptation processes (Quach et al. 2009; Carbonell et al. 2012).

Trying to infer the evolutionary history of microRNAs

Laboring under the assumption that there are high levels of conservation in miRNA regions, we were very curious to examine how the scarce but existing nucleotide variation is distributed across miRNA genes. By looking at the human miRNAs that show orthologous sequences in the remaining great ape genomes, we saw that the level of sequence conservation is mainly explained by the time of emergence, i.e. more conserved miRNAs tend to be older. Despite being not statistically significant, we also noticed that these ancient and conserved miRNAs are often more highly expressed (in terms of levels of expression), associated with disease, located within miRNA clusters and duplicated (these last two features could be intricately linked).

The patterns we observed might support recent studies that have contributed to tracing the evolutionary history of miRNAs. The current outlook maintains that, in general terms, a new-born miRNA (emerged *de novo* or by neo-functionalization of a previous existing miRNA) is probably expressed at low levels, in only some tissues and targeting a limited numbers of genes with modest functional consequences (Liang and Li 2009; Shen et al. 2011). Introns and tandem duplications are a common source of miRNA generation, thus it is not odd that younger miRNAs are usually intronic and the oldest ones form larger clusters since they have had more time to spread by duplication and to experiment chromosomal rearrangements that may join miRNAs with related functions together. Also, it is likely that

younger miRNAs are under more relaxed constraint, evolving faster and, in many cases, eliminated by purifying selection, probably due to the accumulation of deleterious changes (Zhang et al. 2007; Meunier et al. 2013; Santpere et al. 2016). These early life stages of miRNAs have been proposed as a trial period for the miRNA establishment (Liang and Li 2009). Those miRNAs that are not eliminated would acquire more relevant functions being eventually more broadly expressed, at higher levels, regulating more genes and consequently, playing a role in diseases with higher probabilities and being subjected to stronger selective pressure (Landgraf et al. 2007; Liang and Li 2009; Shen et al. 2011; Meunier et al. 2013).

This well-supported picture, however, still raises many questions since we have just started to understand some of the most general patterns of miRNA conservation, organization and function. For example, it remains to be determined which specific molecular and functional features permit miRNAs to evolve and to be conserved in animal genomes over time; why and how some miRNAs are broadly expressed while others are tissue-specific regardless of their conservation levels (Landgraf et al. 2007); or how we can test the age of ancient miRNAs that experimented high mutation rates within specific lineages. The latter, in particular, gives rise to a great debate in the field, since most molecular age estimations of miRNAs are based on the presence of homologous and highly conserved sequences across evolutionary distal species (Hertel et al. 2006; Wheeler et al. 2009; Iwama et al. 2013; Meunier et al. 2013). Nevertheless, *is this characterization of an old miRNA reliable?* As already proposed, we are most probably seeing a partial picture of the entire ancient miRNA repertoire that only includes the genes that experienced the highest and most extensive levels of conservation (Hertel et al. 2006; Wheeler et al. 2009). If a significant portion of the ancient miRNA genes had been under less-intense selective pressures and had substantially diversified throughout the animal phylogeny, then currently they could be considered as recently emerged miRNAs; however, they would have, indeed, more ancient origins. In addition, the study of miRNA

genes present in metazoans has often been based on their high levels of sequence conservation. This kind of approach has probably led to a bias in miRNA evolutionary studies, since only the most conserved regions are detected while many species-specific miRNAs are excluded. Hopefully, increasing studies based on not only sequence conservation comparative analyses but also the conservation of secondary structures, and involving more representative individuals from the same population, as well as from evolutionary close and distal species, will allow us to draw up a clearer map of the metazoan miRNA spectrum in order to better understand the history of these old and continuously emerging post-transcriptional regulators.

Sequence variation as a source of miRNA function diversification

In order to elucidate the evolutionary history of miRNAs in the context of primate evolution, apart from analyzing how nucleotide changes are spread along miRNA genes, we also wanted to investigate which could be some of the functional consequences associated with this variation. Hence, we studied the miRNAs mir-299, mir-508, mir-541 and mir-503 (first chapter of Results), which present nucleotide changes within the mature (the first three) or seed region (the last one), fixed in all great ape populations and differing in humans. We further studied mir-376a1 (second chapter of Results) in human, chimpanzee, gorilla and macaque, whose primary sequence has been reported to be edited at three different positions located in the precursor region and in both the -5p and -3p seed regions.

As reflected in both chapters, by means of DNA mutation or RNA editing modification, nucleotide changes in the mature miRNA regions of mir-299, mir-508, mir-541 (emerged at the DNA level) and mir-376a1 (emerged at the RNA level) could seriously affect the hairpin stability. Moreover, comparative analysis of mir-299 and mir-508 between human and non-human great apes revealed that variations in the length of the precursor sequences may also affect the stability of the hairpins. Interestingly, we observed that the hairpins with higher

molecular stabilities were significantly more expressed, at least in the studied mir-299, mir-508 and mir-541. Our results suggest that nucleotide changes affecting the mature miRNA region and/or the lengths of the precursor miRNA molecule could dramatically alter the stability of the hairpins and, consequently, their expression levels. Changes in the expression levels of the hairpins could increment or reduce the final proportion of mature miRNA molecules, possibly triggering functional consequences, as previously observed for a human SNP present in mir-146a (Jazdzewski et al. 2008). In order to test if the differences in the expression levels of the hairpins could involve new regulatory functions, it would be interesting to analyze which genes appear deregulated taking into account this variable. This could be done by transcriptome analyses after transfection of the precursor molecules rather than the mature miRNAs. Nevertheless, in this case, we would be analyzing the regulatory effects of both the -5p and -3p mature strands. To overcome this limitation we could instead transfect the mature miRNA variants (represented as mimic miRNAs) but at different concentrations in order to check for the effect of miRNA dosage on the whole transciptome.

In addition to molecular stability, the studied nucleotide changes seem to affect the spectrum of target genes, in some cases with interesting novel functions associated. For instance, as reported, editing at position +44 of mir-376a1 can redefine the set of target genes, some of which are associated to complete new phenotypes in human and mouse (Kawahara, Zinshteyn, Sethupathy, et al. 2007; Choudhury et al. 2012). The most relevant case that we found was the human-specific genomic nucleotide substitution located in the mature region of miR-541-3p. Deregulated gene analyses and target gene predictions revealed that at least five candidate genes implicated in brain and neural function (*KCTD11*, *SLC12A6*, *PMP22*, *UQCC* and *TMEM183B*) seem to be directly regulated by the human variant of this miRNA but not by the non-human primate one. Our work highlights how species-specific nucleotide substitutions in mature miRNA regions could promote the regulation of a different spectrum

of target genes, which may play important roles, among others, in brain function and brain regulation in the human lineage.

Overall, our results emphasize how nucleotide changes occurring in the mature miRNA region (at the DNA or RNA level) may act as a source of miRNA functional diversification by redirecting the repertoire of target genes. Additionally, these nucleotide changes and/or changes in the length of the precursor miRNA sequence may significantly affect the stability of the hairpin, presumably altering its expression levels and the proportion of expressed mature molecules. Both sources of variation (nucleotide sequence and hairpin stability changes) may provide new regulatory functions to the miRNAs and contribute to shaping phenotypic traits in great apes (**Figure 17**).

What we have learnt about RNA editing in mir-376a1

In the second part of the Results section, we performed a thorough analysis of miRNA editing patterns in mir-376a1. Our outcomes highlight the vast amount of variation associated with a single miRNA gene. Specifically, we observed that eight different edited isoforms of mir-376a1 (i.e. AAA, AAG, AGA, AGG, GAA, GAG, GGA and GGG, where the three nucleotides represent the editable positions -1, +4 and +44 of the precursor sequence, respectively) were expressed at significant levels in all the samples that we analyzed. This reinforces the importance of considering all the products derived from one miRNA gene in the study of miRNA function since, initially, all of them could be functionally relevant and regulate different sets of target genes.

As discussed in the second chapter, we observed interesting tissue-dependent patterns of A-to-I editing in mir-376a1, which is more frequently edited in primate brain cortex samples than in human placenta. Moreover, in all the examined samples, position +4 was the most edited, followed by positions +44 and -1. The conservation of RNA editing patterns across different primate tissues, as well as miRNA sites, suggests a functional role of RNA editing in primate brains, at least in the case of mir-376a1. The observed editing profiles

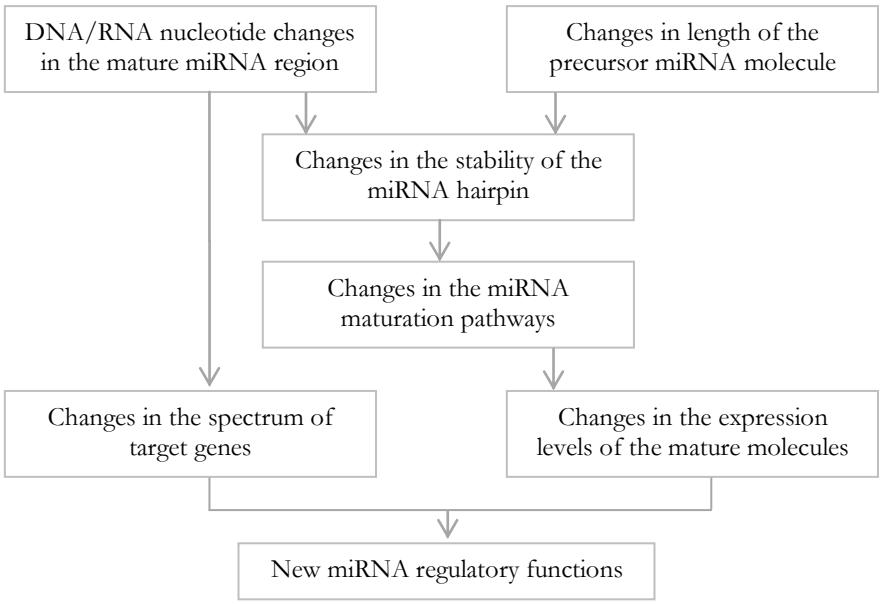


Figure 17. Proposed pathways of miRNA diversification. Based on previous studies and on our results, we suggest that nucleotide changes in the mature miRNA regions (both generated at the DNA or RNA levels) may redefine a new set of target genes. These latter changes and/or changes in the length of the precursor miRNA regions could modify the stability of the hairpins leading to alteration in the miRNA maturation pathways and in the synthesis of mature molecules. Both processes, nucleotide sequence and hairpin stability changes, could ultimately implicate new miRNA regulatory functions.

---

reflect previous observations of brain tissues reaching the highest editing frequencies and, despite a possible role of RNA editing in brain plasticity and intricate brain regulation has been suggested, the causes and functional consequences of the higher editing values in brain regions are still unknown (Blow et al. 2006; Li et al. 2013; Chen et al. 2014; Warnefors et al. 2014; Picardi et al. 2015). Considering that we analyzed the same miRNA across diverse tissues, species and individuals, it would be particularly interesting to explore the reasons behind the observed editing differences between brain cortex and

placenta. One possible explanation could be that ADAR enzymes are not equally expressed in these two tissues, making editing activity significantly higher in the brain. It could have been pertinent to look at the expression of ADAR in the different tissues and species analyzed; however, this was not possible to test at the time of writing this thesis. Interestingly, there is evidence both in favour and against this hypothesis. While some authors have observed a positive correlation between the levels of ADAR expression and the frequency of RNA editing across different human and macaque tissues (Chen et al. 2014; Picardi et al. 2015), others did not find differences in the ADAR levels among several primate and mammalian tissues that, on the contrary, were differentially edited (Kawahara, Zinshteyn, Chendrimada, et al. 2007; Paz-Yaacov et al. 2010). Even though changes in the ADAR expression profile could somehow explain editing differences among tissues, it seems very likely that additional regulatory mechanisms are also involved in this process.

Thus, based on our results and supporting previous studies, we have observed that RNA editing significantly differs among tissues (Blow et al. 2006; Kawahara, Zinshteyn, Chendrimada, et al. 2007; Warnefors et al. 2014). Nevertheless, how RNA editing patterns vary among individuals has been poorly investigated. One of the most valuable contributions of our research is the analysis of RNA editing frequencies across different individuals (from the same species and from different primate species). Interestingly, although we could describe reliable editing patterns among tissues in the studied samples, the eight edited isoforms of mir-376a1 considerably varied among individuals. In fact, once we were able to globally examine our results for the first time, this remarkable variability caught our attention. The question remains: *although we can describe clear editing patterns among tissues, to what extent can we infer a functional role of miRNA editing giving this huge level of variation?* This is something that still concerns us and, as mentioned in the second chapter of the Results section, we think that the analysis of more individuals and in particular of more tissue samples (e.g. placenta samples from non-human primates) or

particular conditions (e.g. brain samples from neurodegenerative disorders), could provide valuable information about the actual magnitude of the inter-individual variability in different circumstances. In this regard, two recent studies have also reported considerable variation in the RNA editing frequencies among individuals when comparing high-throughput datasets of various studies and tissue samples from three different humans, not only in miRNA regions but also in the whole transcriptome, suggesting that RNA editing could be a highly dynamic mechanism determined by individual-specific features (Chen et al. 2014; Picardi et al. 2015). Motivated by the fact that RNA editing can vary along a wide range of frequencies depending on the case study, a very enlightening research has recently analyzed the APOBEC-mediated C-to-U RNA editing patterns derived from single cells (macrophages and dendritic cells). Interestingly, this approach showed that some transcripts were similarly edited across cells, but others exhibited a huge editing rate variation (Harjanto et al. 2016). Although these analyses were performed on specific gene transcripts and cell types, and they were based on C-to-U changes, this work brings to light the importance of understanding the specific RNA editing profiles associated with each transcript; not just at the species, individual or tissue-level, but even at the cellular-level. The advent of more precise technologies, such as single-cell and single-molecule sequencing, could provide key insights into the RNA editing patterns of particular transcripts under different conditions, giving us a better understanding of the roles of this post-transcriptional process, based on the biological context in which it emerges.

On the other hand, previous reports show that certain transcript domains are more frequently edited than others. In particular, adenosine residues located next to the 5' extreme of the miRNA hairpin, within UAG triplets and pairing whit cytosine bases, are more often edited (Blow et al. 2006; Kawahara et al. 2008; Alon et al. 2012; Peng et al. 2012). Nevertheless, these conditions are not indispensable and they do not explain, by themselves, why and when certain

adenosines are deaminated. Our results suggest the role of A-to-I transitions in the stabilization of the miRNA hairpins. Following previous works and based on our analyses, we hypothesize that editing changes at very specific positions could confer higher stability to the dsRNA structures in which they occur, ultimately increasing their activity as a result of increasing their lifetimes (Levanon et al. 2004; Irimia et al. 2012). We propose that A-to-I RNA editing could have been a selected post-transcriptional mechanism in metazoans that confers high genetic variability and increases the molecular stability of certain transcripts.

Finally and for the first time in the study of RNA editing changes, we also explored the possible epistatic relationships among edited sites. Considering that the primary molecules of mir-376a1 can be edited at three different sites, it was of our interest to investigate if editing at one position may alter the editing at the remaining ones. Although we did not observe significant epistatic effects, larger sample size would likely provide more confidence to our results. Remarkably, we think that this approach may be useful for future studies in order to better understand possible connections among editable sites, particularly in those transcripts showing highly conserved editing patterns.

Scopes and limitations

The major part of the work presented here was based on data from experiments that were possible thanks to the efforts of many people. As in the *in silico* analyses, all the *wet lab* experiments were discussed at length before being performed with the aim of designing the most reliable methods to test our inquiries, as well as minimizing time and costs. As is to be expected, there were many times when the experiments failed or were not as accurate as we had hoped. To give an example, we had a lot of trouble during the RNA editing quantification studies and we had to implement several protocols and to have extensive discussions with one of the most relevant groups working in the field, until we were finally able to precisely evaluate RNA editing changes in very small sequences as mature miRNA

molecules are formed. Now that this thesis is drawing to a close, it is also a good time to recapitulate the limitations we experienced and to think about how our work could be improved in the next future.

A very stringent limitation on the study of human evolution within the context of primates and great apes is the bias in the genomic data that have been generated over the years, clearly favoring human species over the rest. For instance, 2588 mature miRNA sequences have been reported in the human genome, while only 587, 357, 660 and 914 have been annotated for, respectively, chimpanzee, gorilla, orangutan and macaque genomes (Kozomara and Griffiths-Jones 2014). This inequality has also affected some of our studies. Expression analyses of the non-human variants of mir-299, mir-503, mir-508 and mir-541 (first chapter of Results) were performed on human cell lines, probably altering the results concerning species-specific genes. Although all the genes that were candidates for being regulated by the non-human miRNA variants were confirmed to be present in their corresponding primate genomes, it would be important to further explore which particular genes are deregulated when the non-human miRNA variants are present in cell lines from the different great ape species. Moreover, in this first section of the Results, we studied the set of miRNA genes that have been annotated in the human genome (miRBase, release 19) and for which we found an orthologous miRNA gene in the rest of the great apes. However, this method did not take into account those conserved miRNAs that could have been lost in any of the great ape lineages, as well as those species-specific miRNAs that could have exclusively emerged in every lineage. By excluding both groups of miRNA genes, we probably lost relevant information about the putative contribution of certain miRNAs to the phenotypic diversification processes in great apes.

In addition, cloning amplification studies, performed during the RNA editing analyses of the primary and mature mir-376a1 molecules (second chapter of the Results section) could somehow alter the natural environment in which these molecules are expressed. Although no ADAR expression has been reported beyond metazoans (Jin et al.

2009), bacterial ecosystems might favor or disfavor certain edited isoforms, modifying the editing frequency estimations. As discussed in the corresponding manuscript of this work, we also observed significant changes between NGS and Sanger sequencing approaches. These discrepancies are also evident in the literature, increasing our doubts regarding the extensiveness and functional relevance of the editing changes that we (and others) have observed. We can only assume that each method will provide different and valuable information. High throughput technologies may eventually cover a wider spectrum of editable sites, which would permit more ambitious comparative genomic analyses with many samples. On the contrary, direct Sanger sequencing may provide higher editing estimations, which could become crucial in the analyses of lowly edited/expressed transcripts and specific genes/sites. Overall, there are several reasons to suspect the reliability of all the approaches we, and other researchers, have adopted as the basis of our observations. Nevertheless, we think it is important to be aware of these limitations, trying to be conservative among different methods, i.e. not drawing definitive conclusions regarding RNA editing changes when comparing results obtained by different approaches.

An obvious bias of our research was the use of artificial and simplistic systems to infer biological processes that occur in complex and variable physiological conditions. In particular, overexpression experiments in HeLa and SH-SY5Y neuroblastoma cell lines (first chapter of the Results section) could have distorted the real levels of expression and/or regulatory functions of each studied miRNA variant. Immortalized cell lines constitute a feasible approach to working in a living environment avoiding the use of living organisms; however, these cells have important disadvantages such as their increased chromosomal rearrangements and mutation rates, and the expression of many cancer-related genes. A possible solution could have been the use of primary neuronal cultures in order to analyze the behavior of the studied miRNAs in their natural cellular environment.

However, these cell cultures are extremely difficult to maintain, particularly when cell transfection experiments must be carried out.

As with all studies based on transcriptomic data, extreme care must be taken in the extraction and preservation of the samples. Small changes in their manipulation could produce significant deviations in the observed expression profiles. Also, miRNA expression patterns may considerably vary under minor environmental changes or across different tissue regions (even cell types). Despite following the same guidelines when performing all the RNA extractions and experimental replicates, as well as always trying to compare equivalent/homologous conditions (e.g. tissue sections from corresponding regions and healthy adult donors or cell cultures grown under the same criteria), some variables could have escaped our control introducing certain (maybe relevant) noise in our comparative analyses. Through the analyses of different individuals and replicates, we expect to have overcome most of these difficulties, although the inclusion of more samples would probably contribute to distinguishing between reliable patterns and stochastic variability.

Finally, the study of the functional implications associated with miRNA variation usually requires several gene validation assays. By means of transcriptome analyses performed after miRNA overexpression and target gene prediction, several genes were identified as candidates for being regulated by the studied miRNAs during work on this thesis. Further gene validations could confirm if the proposed genes are indeed directly deregulated by these miRNAs. These studies could be based, for example, on luciferase reporter gene assays in which the miRNA-mediated gene repression would be confirmed if the expression of the reporter luciferase gene is reduced. Another informative approach would be the analysis of endogenous protein production, after the overexpression and/or knockdown of the studied miRNAs.

Overcoming the aforementioned limitations and future work on the functional implications associated with miRNA genetic variation could shed light on how molecular changes, occurring at the genomic and

post-transcriptional level in miRNA genes, may have contributed to shaping phenotypic traits among primates, along with improving our understanding of the evolutionary history of our species.

V. CONCLUSIONS

- The microRNA seed region has been largely considered the most important region for microRNA function; nevertheless the conservation and functional analyses performed emphasize the importance of the entire mature microRNA sequence.

- More conserved microRNAs are generally older. They also tend to be clustered, duplicated, more expressed and associated with disease.

- We found a total of 263 human-specific nucleotide substitutions located in 235 precursor microRNAs, from which 179 were in the precursor excluding the mature sequence, 61 were in the mature excluding the seed and 23 were in the seed.

- The human-specific nucleotide substitutions located in the mature region of miR-299-3p, miR-508-3p and miR-541-3p and/or changes in the lengths of their respective precursor microRNA molecules are related to variations in their expression levels.

- The human-specific nucleotide substitution located in the seed region of miR-503-3p does not change the expression levels of its corresponding precursor molecule, but it is expected to significantly alter the spectrum of target genes.

- The human-specific nucleotide substitution located in the mature region of miR-541-3p might be associated with changes in the expression of five predicted target genes (*KCTD11*, *SLC12A6*, *PMP22*, *UQCC* and *TMEM183B*) previously shown to be involved in brain and neural functions.

- The ratio of edited microRNAs to the total number of microRNAs was very similar among all primate tissues and species analyzed, reaching values of about 0.12.

- Eight different edited isoforms derived from mir-376a1 primary molecules are detected at significant levels in several primate tissues with potentially different regulatory functions.

- The frequency of A-to-I mir-376a1 editing shows differences among tissues, species and individuals, being more frequent in primate cortex brain than in human placenta.

- Position +4 is the most edited site of the primary mir-376a1 in all the tissues and primate species analyzed, which may be due to the fact that it is the editing position that provides the highest stability to the precursor molecule.

- No significant epistatic effects were detected among the three edited sites (-1, +4 and +44) of mir-376a1.

- A-to-I RNA editing changes occurring at higher frequencies could have been selected if they gave higher molecular stabilities to specific double-stranded RNA transcripts that play crucial roles in certain tissues and/or species.

VII. REFERENCES

Agarwal V, Bell GW, Nam JW, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian mRNAs. Elife 4:1–38.

Alon S, Mor E, Vigneault F, Church GM, Locatelli F, Galeano F, Gallo A, Shomron N, Eisenberg E. 2012. Systematic identification of edited microRNAs in the human brain. Genome Res. 22:1533–1540.

Amaral PP, Mattick JS. 2008. Noncoding RNA in development. Mamm. Genome 19:454–492.

Ambros V. 2004. The functions of animal microRNAs. 431.

Ason B, Darnell DK, Wittbrodt B, Berezikov E, Kloosterman WP, Wittbrodt J, Antin PB, Plasterk RH a. 2006. Differences in vertebrate microRNA expression. Proc. Natl. Acad. Sci. U. S. A. 103:14385–14389.

Athanasiadis A, Rich A, Maas S. 2004. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. 2.

Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature 444:171–178.

Avise JC. 2010. Perspective: Conservation genetics enters the genomics era. Conserv. Genet. 11:665–669.

Axtell MJ, Westholm JO, Lai EC. 2011. Vive la différence: biogenesis and evolution of microRNAs in plants and animals. Genome Biol. 12:221.

Bahn JH, Lee J, Li G, Greer C, Peng G. 2012. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. :142–150.

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 116:281–297.

Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. Cell 136:215–233.

Baskerville S, Bartel DP. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA 11:241–247.

Bass B, Hundley H, Li JB, Peng Z, Pickrell J, Xiao XG, Yang L. 2012. The difficult calls in RNA editing. Nat. Biotechnol. 30:1207–1209.

Batzer M a, Deininger PL. 2002. Alu Repeats and Human Genomic Diversity. Nat. Rev. Genet. 3:370–379.

Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, et al. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet. 37:766–770.

Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. 2007. Mammalian Mirtron Genes. Mol. Cell 28:328–336.

Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH a, Cuppen E. 2005. Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120:21–24.

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

Betel D, Wilson M, Gabow A, Marks DS, Sander C. 2008. The microRNA.org resource: Targets and expression. Nucleic Acids Res. 36:149–153.

Bhattacharya A, Ziebarth JD, Cui Y. 2014. PolymiRTS Database 3.0: Linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res. 42:86–91.

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

Brameier M. 2010. Genome-wide comparative analysis of microRNAs in three non-human primates. BMC Res. Notes 3:64.

Brenner S, Jacob F, Meselson M. 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature 190:576–581.

Britten RJ, Davidson EH. 1969. Gene Regulation for Higher Cells: A Theory. Science (80-.). 165:349–357.

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. 1992. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71:515–526.

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. 1992. The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71:527–542.

Burns CM, Chu H, Rueter SM, Hutchinson LK, Canton H, Sanders-Bush E, Emeson RB. 1997. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature 387:303–308.

Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL, Raj A. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biol. 16:20.

Calin G a, Croce CM. 2006. MicroRNA signatures in human cancers. Nat. Rev. Cancer 6:857–866.

Calvignac-Spencer S, Leendertz S a J, Gillespie TR, Leendertz FH. 2012. Wild great apes as sentinels and sources of infectious disease. Clin. Microbiol. Infect. 18:521–527.

Campo-Paysaa F, Sémon M, Cameron RA, Peterson KJ, Schubert M. 2011. microRNA complements in deuterostomes: origin and evolution of microRNAs. Evol. Dev. 13:15–27.

Carbonell J, Alloza E, Arce P, Borrego S, Santoyo J, Ruiz-Ferrer M, Medina I, Jiménez-Almazán J, Méndez-Vidal C, González-Del Pozo M, et al. 2012. A map of human microRNA variation uncovers unexpectedly high levels of variability. Genome Med. 4:62.

Chen C, Ridzon D a., Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, et al. 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. 33:1–9.

Chen JY, Peng Z, Zhang R, Yang XZ, Tan BCM, Fang H, Liu CJ, Shi M, Ye ZQ, Zhang YE, et al. 2014. RNA Editome in Rhesus Macaque Shaped by Purifying Selection. PLoS Genet. 10:8–12.

Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. Nat. Rev. Genet. 8:93–103.

Chen LL, Carmichael GG. 2008. Gene regulation by SINES and inosines: Biological consequences of A-to-I editing of Alu element inverted repeats. Cell Cycle 7:3294–3301.

Cheng W-C, Chung I-F, Huang T-S, Chang S-T, Sun H-J, Tsai C-F, Liang M-L, Wong T-T, Wang H-W. 2013. YM500: a small RNA sequencing (smRNA-seq) database for microRNA research. Nucleic Acids Res. 41:D285–D294.

Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, et al. 2010. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 24:992–1009.

Choudhury Y, Tay FC, Lam DH, Sandanaraj E, Tang C, Ang BT, Wang S. 2012. Attenuated adenosine-to-inosine editing of microRNA-376a*promotes invasiveness of glioblastoma cells. J. Clin. Invest. 122:4059–4076.

Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, Klaus S, Snyman H, Hannon GJ, Bork P, Arendt D. 2010. Ancient animal microRNAs and the evolution of tissue identity. Nature 463:1084–1088.

Consortium C. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.

Conticello SG. 2012. Creative deaminases, self-inflicted damage, and genome evolution. Ann. N. Y. Acad. Sci. 1267:79–85.

Costa FF. 2010. Non-coding RNAs: Meet thy masters. BioEssays 32:599–608.

Crick F. 1958. On Protein Synthesis. Symp. Soc. Exp. Biol.:138–166.

Crick FHC. 1970. Central Dogma of Molecular Biology. Nature 227:561–563.

Cullen BR. 2010. Five questions about viruses and microRNAs. PLoS Pathog. 6:2–4.

Dannemann M, Nickel B, Lizano E, Burbano HA, Kelso J. 2012. Annotation of primate miRNAs by high throughput sequencing of small RNA libraries. BMC Genomics 13:116.

Derrien T, Johnson R, Bussotti G, Tanzer a, Djebali S, Tilgner H, Guernec G, Merkel a, Gonzalez D, Lagarde J, et al. 2012. The GENCODE v7 catalogue of human long non-coding RNAs : Analysis of their structure , evolution and expression. Genome Res. 22:1775–1789.

Desvignes T, Batzel P, Berezikov E, Eilbeck K, Eppig JT, McAndrews MS, Singer a., Postlethwait JH. 2015. MiRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants. Trends Genet. 31:613–626.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. 2012. Landscape of transcription in human cells. Nature 489:101–108.

Doench JG, Petersen CP, Sharp P a. 2003. siRNAs can function as miRNAs. Genes Dev. 17:438–442.

Drakaki A, Iliopoulos D. 2009. MicroRNA Gene Networks in Oncogenesis. Curr. Genomics 10:35–41.

Du Z-Q, Yang C-X, Rothschild MF, Ross JW. 2013. Novel microRNA families expanded in the human genome. BMC Genomics 14:98.

Durbin RM, Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Collins FS, Vega D La, Francisco M, et al. 2010. A map of human genome variation from population-scale sequencing.

Eisenberg E, Nemzer S, Kinar Y, Sorek R, Rechavi G, Levanon EY. 2005. Is abundant A-to-I RNA editing primate-specific? Trends Genet. 21:77–81.

Ekdahl Y, Farahani HS, Behm M, Lagergren J, Ohman M. 2012. A-to-I editing of microRNAs in the mammalian brain increases during development. Genome Res.:1477–1487.

Esau C, Davis S, Murray SF, Yu XX, Pandey SK, Pear M, Watts L, Booten SL, Graham M, McKay R, et al. 2006. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. Cell Metab. 3:87–98.

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

Faller M, Guo F. 2008. MicroRNA biogenesis: there's more than one way to skin a cat. Biochim. Biophys. Acta - Gene Regul. Mech. 1779:663–667.

Farh KK-H, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP. 2005. The widespread impact of

mammalian MicroRNAs on mRNA repression and evolution. Science 310:1817–1821.

Filipowicz W, Bhattacharyya SN, Sonenberg N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9:102–114.

Franklin RE, Gosling RG. 1953. Molecular Configuration in Sodium Thymonucleate. Nature 171:740–741.

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, et al. 2016. Functional Implications of Human-Specific Changes in Great Ape microRNAs. PLoS One 11:e0154194.

Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. 2011. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat. Struct. Mol. Biol. 18:1139–1146.

García-López J, Hourcade JDD, Del Mazo J. 2013. Reprogramming of microRNAs by adenosine-to-inosine editing and the selective elimination of edited microRNA precursors in mouse oocytes and preimplantation embryos. Nucleic Acids Res. 41:5483–5493.

Gardner PP, Vinther J. 2008. Mutation of miRNA target sequences during human evolution. Trends Genet. 24:262–265.

Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: an expanding universe. Nat. Rev. Genet. 10:94–108.

Ghosh Z, Mallick B, Chakrabarti J. 2009. Cellular versus viral microRNAs in host-virus interaction. Nucleic Acids Res. 37:1035–1048.

Glazov EA, McWilliam S, Barris WC, Dalrymple BP. 2008. Origin, evolution, and biological role of miRNA cluster in DLK-DIO3

genomic region in placental mammals. Mol. Biol. Evol. 25:939–948.

Gommans WM, Mullen SP, Maas S. 2009. RNA editing: a driving force for adaptive evolution? Bioessays 31:1137–1145.

Gong J, Wu Y, Zhang X, Liao Y, Sibanda VL, Liu W, Guo A-Y. 2014. Comprehensive analysis of human small RNA sequencing data provides insights into expression profiles and miRNA editing. RNA Biol. 11:1375–1385.

Gould SJ. 1989. Wonderful life: The Burgess shale and the nature of history.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, et al. 2010. A draft sequence of the Neandertal genome. Science 328:710–722.

Greeve J, Altkemper I, Dieterich JH, Greten H, Windler E. 1993. Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins. J. Lipid Res. 34:1367–1383.

Griffiths-Jones S, Hui JHL, Marco A, Ronshaugen M. 2011. MicroRNA evolution by arm switching. EMBO Rep. 12:172–177.

Grimson A, Farh KKH, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007a. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. Mol. Cell 27:91–105.

Grimson A, Farh KKH, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007b. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. Mol. Cell 27:91–105.

Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature 455:1193–1197.

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. 2008. The Vienna RNA websuite. Nucleic Acids Res. 36:70–74.

Guo L, Zhao Y, Zhang H, Yang S, Chen F. 2014. Integrated evolutionary analysis of human miRNA gene clusters and families implicates evolutionary relationships. Gene 534:24–32.

Gupta R a, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai M-C, Hung T, Argani P, Rinn JL, et al. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464:1071–1076.

Gutschner T, Hämmerle M, Eißmann M, Hsu J, Kim Y, Hung G, Revenko A, Arun G, Stentrup M, Groß M, et al. 2013. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res. 73:1180–1189.

Haas BJ, Wortman JR, Ronning CM, Hannick LI, Smith RK, Maiti R, Chan AP, Yu C, Farzad M, Wu D, et al. 2005. Complete reannotation of the Arabidopsis genome: methods, tools, protocols and the final release. BMC Biol. 3:7.

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

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013. Natural RNA circles function as efficient microRNA sponges. Nature 495 VN -:384–388.

Harjanto D, Papamarkou T, Oates CJ, Rayon-Estrada V, Papavasiliou FN, Papavasiliou A. 2016. RNA editing generates cellular subsets with diverse sequence within populations. Nat. Commun. 7:12145.

Hattori M. 2005. Finishing the euchromatic sequence of the human genome. Tanpakushitsu Kakusan Koso. 50:162–168.

Hausser J, Zavolan M. 2014. Identification and consequences of miRNA-target interactions - beyond repression of gene expression. Nat. Rev. Genet. 15:599–612.

He T, Wang Q, Feng G, Hu Y, Wang L, Wang Y. 2011. Computational detection and functional analysis of human tissue-specific A-to-I RNA editing. PLoS One 6:e18129.

Heale BSE, Keegan LP, McGurk L, Michlewski G, Brindle J, Stanton CM, Caceres JF, O'Connell M a. 2009. Editing independent effects of ADARs on the miRNA/siRNA pathways. EMBO J. 28:3145–3156.

Hébert SS, Horré K, Nicolaï L, Papadopoulou AS, Mandemakers W, Silahtaroglu AN, Kauppinen S, Delacourte A, De Strooper B. 2008. Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. Proc. Natl. Acad. Sci. U. S. A. 105:6415–6420.

Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ, Peterson KJ. 2008. MicroRNAs and the advent of vertebrate morphological complexity. Proc. Natl. Acad. Sci. U. S. A. 105:2946–2950.

Herranz H, Cohen SM. 2010. MicroRNAs and gene regulatory networks: Managing the impact of noise in biological systems. Genes Dev. 24:1339–1344.

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.

Hoagland MB, Stephenson ML, Scott JF, Hecht LI, Zamecnik PC. 1958. A soluble ribonucleic acid intermediate in protein synthesis. J. Biol. Chem. 231:241–257.

Hobert O. 2008. Gene regulation by transcription factors and microRNAs. Science 319:1785–1786.

Hoopengardner B, Bhalla T, Staber C, Reenan R. 2003. Nervous system targets of RNA editing identified by comparative genomics. Science 301:832–836.

Hu HY, Guo S, Xi J, Yan Z, Fu N, Zhang X, Menzel C, Liang H, Yang H, Zhao M, et al. 2011. MicroRNA expression and regulation in human, chimpanzee, and macaque brains. PLoS Genet. 7:e1002327.

Hutvágner G, Zamore PD. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. Science 297:2056–2060.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. Nature 436:793–800.

Iorio M V., Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M, et al. 2005. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 65:7065–7070.

Irimia M, Denuc A, Ferran JL, Pernaute B, Puelles L, Roy SW, Garcia-Fernàndez J, Marfany G. 2012. Evolutionarily conserved A-to-I editing increases protein stability of the alternative splicing factor Nova1. RNA Biol. 9:12–21.

Ishizu H, Siomi H, Siomi MC. 2012. Biology of Piwi-interacting RNAs: New insights into biogenesis and function inside and outside of germlines. Genes Dev. 26:2361–2373.

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.

Iwamoto K, Bundo M, Kato T. 2005. Estimating RNA editing efficiency of five editing sites in the serotonin 2C receptor by pyrosequencing. RNA 11:1596–1603.

Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3:318–356.

Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A, 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.

Jiang Z, Tang H, Ventura M, Cardone MF, Marques-Bonet T, She X, Pevzner P a, Eichler EE. 2007. Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. Nat Genet 39:1361–1368.

Jin Y, Zhang W, Li Q. 2009. Origins and evolution of ADAR-mediated RNA editing. IUBMB Life 61:572–578.

Kaessmann H, Wiebe V, Weiss G, Paabo S. 2001. Great ape DNA sequences reveal a reduced diversity and an expansion in humans. Nat Genet 27:155–156.

Karreth F a., Reschke M, Ruocco A, Ng C, Chapuy B, Léopold V, Sjoberg M, Keane TM, Verma A, Ala U, et al. 2015. The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. Cell 161:319–332.

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.

Kawahara Y, Zinshteyn B, Chendrimada TP, Shiekhattar R, Nishikura K. 2007. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO Rep. 8:763–769.

Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. 2007. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science 315:1137–1140.

Kawahara Y. 2012. Quantification of adenosine-to-inosine editing of microRNAs using a conventional method. Nat. Protoc. 7:1426–1437.

Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. 2007. The role of site accessibility in microRNA target recognition. Nat. Genet. 39:1278–1284.

Khaitovich P, Enard W, Lachmann M, Pääbo S. 2006. Evolution of primate gene expression. Nat. Rev. Genet. 7:693–702.

Kim DDY, Kim TTY, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A. 2004. Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res. 14:1719–1725.

Kim VN. 2005. MicroRNA biogenesis: coordinated cropping and dicing. Nat. Rev. Mol. Cell Biol. 6:376–385.

King M, Wilson A. 1975. Evolution at two levels in humans and chimpanzees. Science (80-.). 188:107–116.

Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, Ambesi-Impiombato A, Califano A, Migliazza A, Bhagat G, et al. 2010. The DLEU2/miR-15a/16-1 Cluster Controls B Cell Proliferation and Its Deletion Leads to Chronic Lymphocytic Leukemia. Cancer Cell 17:28–40.

Kleinjan D a, Van Heyningen V. 2005. Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease. Am. J. Hum. Genet 76:8–32.

Kloosterman WP, Wienholds E, de Bruijn E, Kauppinen S, Plasterk RH a. 2006. In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. Nat. Methods 3:27–29.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK. 2006. Genomics and conservation genetics. Trends Ecol. Evol. 21:629–637.

Kosik KS. 2013. Circles reshape the RNA world. Nature 495:4–6.

Kozomara A, Griffiths-Jones S. 2014. MiRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42:68–73.

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

Krützfeldt J, Poy MN, Stoffel M. 2006. Strategies to determine the biological function of microRNAs. Nat. Genet. 38 Suppl:S14–S19.

Kume H, Hino K, Galipon J, Ui-Tei K. 2014. A-to-I editing in the miRNA seed region regulates target mRNA selection and silencing efficiency. Nucleic Acids Res. 42:10050–10060.

Kwon C, Han Z, Olson EN, Srivastava D. 2005. MicroRNA1 influences cardiac differentiation in Drosophila and regulates Notch signaling. Proc. Natl. Acad. Sci. U. S. A. 102:18986–18991.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. Science 294:853–858.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129:1401–1414.

Landthaler M, Yalcin A, Tuschl T. 2004. The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis. Curr. Biol. 14:2162–2167.

Langergraber K, Prüfer K. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc. … 109:15716–15721.

Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294:858–862.

Lee J, Ang JK, Xiao X. 2013. Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. RNA.

Lee JT, Davidow LS, Warshawsky D. 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. Nat. Genet. 21:400–404.

Lee RC. 1993. The C . elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to & II-14. 75:843–854.

Lee Y, Kim M, Han JJ, Yeom KH, Lee S, Baek SH, Kim VN. 2004. MicroRNA genes are transcribed by RNA polymerase II. Embo J. 23:4051–4060.

Leung AKL, Sharp P a. 2010. MicroRNA Functions in Stress Responses. Mol. Cell 40:205–215.

Di Leva G, Garofalo M, Croce CM. 2014. MicroRNAs in Cancer. Annu. Rev. Pathol. Mech. Dis. 9:287–314.

Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D, et al. 2004. Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotechnol. 22:1001–1005.

Levanon EY, Eisenberg E. 2015. Does RNA editing compensate for Alu invasion of the primate genome? BioEssays 37:175–181.

Lev-Maor G, Sorek R, Levanon EY, Paz N, Eisenberg E, Ast G. 2007. RNA-editing-mediated exon evolution. Genome Biol. 8:R29.

Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20.

Lewis BP, Shih I, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian microRNA targets. Cell 115:787–798.

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. 8:1–12.

Li J, Zhang Z. 2013. miRNA regulatory variation in human evolution. Trends Genet. 29:116–124.

Li JB, Levanon EY, Yoon J-K, Aach J, Xie B, Leproust E, Zhang K, Gao Y, Church GM. 2009. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science 324:1210–1213.

Li X, Cassidy JJ, Reinke C a, Fischboeck S, Carthew RW. 2009. A microRNA imparts robustness against environmental fluctuation during development. Cell 137:273–282.

Li Z, Bammann H, Li M, Liang H, Yan Z, Phoebe Chen Y-P, Zhao M, Khaitovich P. 2013. Evolutionary and ontogenetic changes in RNA editing in human, chimpanzee, and macaque brains. RNA 19:1693–1702.

Liang H, Landweber LF. 2007. Hypothesis: RNA editing of microRNA target sites in humans? RNA 13:463–467.

Liang H, Li W-H. 2009. Lowly expressed human microRNA genes evolve rapidly. Mol. Biol. Evol. 26:1195–1198.

Liang Y, Ridzon D, Wong L, Chen C. 2007. Characterization of microRNA expression profiles in normal human tissues. BMC Genomics 8:166.

Liao JY, Ma LM, Guo YH, Zhang YC, Zhou H, Shao P, Chen YQ, Qu LH. 2010. Deep sequencing of human nuclear and cytoplasmic small RNAS reveals an unexpectedly complex subcellular distribution of mirnas and tRNA 3' trailers. PLoS One 5.

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433:769–773.

Lin S, Cheung WKC, Chen S, Lu G, Wang Z, Xie D, Li K, Lin MCM, Kung H-F. 2010. Computational identification and characterization of primate-specific microRNAs in human genome. Comput. Biol. Chem. 34:232–241.

Lopez-Valenzuela M, Ramírez O, Rosas A, García-Vargas S, De La Rasilla M, 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, Clark AG. 2012. Impact of microRNA regulation on variation in human gene expression. Genome Res. 22:1243–1254.

Lu J, Getz G, Miska E a, Saavedra EA, Lamb J, Peck D, Cordero AS, Ebert BL, Mak RH, Ferrando A a, et al. 2005. MicroRNA expression profiles classify human cancers. Nature 435:834–838.

Lv K, Guo Y, Zhang Y, Wang K, Jia Y, Sun S. 2008. Allele-specific targeting of hsa-miR-657 to human IGF2R creates a potential mechanism underlying the association of ACAA-insertion/deletion polymorphism with type 2 diabetes. Biochem. Biophys. Res. Commun. 374:101–105.

Maas S, Kawahara Y, Tamburro KM, Nishikura K. 2010. A-to-I RNA editing and human disease. RNA Biol. 3:1–9.

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. 98:14687–14692.

Marco A, Hui JHL, Ronshaugen M, Griffiths-Jones S. 2010. Functional shifts in insect microRNA evolution. Genome Biol. Evol. 2:686–696.

Marques-Bonet T, Kidd JM, Ventura M, Graves T a, Cheng Z, Hillier LW, Jiang Z, Baker C, Malfavon-Borja R, Fulton L, et al. 2009. A burst of segmental duplications in the genome of the African great ape ancestor. Nature 457:877–881.

Marques-Bonet T, Ryder OA, Eichler EE. 2009. Sequencing primate genomes: what have we learned? Annu Rev Genomics Hum Genet 10:355–386.

Mattick JS. 2007. A new paradigm for developmental biology. J. Exp. Biol. 210:1526–1547.

McShea DW. 1996. Metazoan Complexity and Evolution: Is There a Trend? Evolution (N. Y). 50:477.

Melé M, Ferreira PG, Reverter F, Deluca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, et al. 2015. Tissues and Individuals. Science (80-.). 348:660–665.

Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, et al. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495 VN -:333–338.

Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H. 2013. Birth and expression evolution of mammalian microRNA genes. Genome Res. 23:34–45.

Meyer M, Kircher M, Gansauge M, Li H, Mallick S, Schraiber JG, Jay F, Prüfer K, De C, Sudmant PH, et al. 2012. A High Coverage Genome Sequence From an Archaic Denisovan Individual. 338:222–226.

Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, et al. 2008. Circulating microRNAs as stable blood-based markers for cancer detection. Proc. Natl. Acad. Sci. U. S. A. 105:10513–10518.

Möller-Krull M, Zemann A, Roos C, Brosius J, Schmitz J. 2008. Beyond DNA: RNA Editing and Steps Toward Alu Exonization in Primates. J. Mol. Biol. 382:601–609.

Morin RD, Connor MDO, Griffith M, Kuchenbauer F, Delaney A, Prabhu A, Zhao Y, Mcdonald H, Zeng T, Hirst M, et al. 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res.:610–621.

Morris K V, Mattick JS. 2014. The rise of regulatory RNA. Nat. Rev. Genet. 15:423–437.

Muiños-Gimeno M, Espinosa-Parrilla Y, Guidi M, Kagerbauer B, Sipilä T, Maron E, Pettai K, Kananen L, Navinés R, Martín-Santos R, et al. 2011. Human microRNAs miR-22, miR-138-2, miR-148a, and miR-488 are associated with panic disorder and regulate several anxiety candidate genes and related pathways. Biol. Psychiatry 69:526–533.

Mulder K De, Berezikov E. 2010. Tracing the evolution of tissue identity with microRNAs. Genome Biol.

Murray BS, Choe SE, Woods M, Ryan TE, Liu W. 2010. An in silico analysis of microRNAs: mining the miRNAome. Mol. Biosyst. 6:1853–1862.

Nakamuta M, Chang BHJ, Zsigmond E, Kobayashi K, Lei H, Ishida BY, Oka K, Li E, Chan L. 1996. Complete phenotypic characterization of apobec-1 knockout mice with a wild-type genetic background and a human apolipoprotein B transgenic background, and restoration of apolipoprotein B mRNA editing by somatic gene transfer of apobec-1. J. Biol. Chem. 271:25981–25988.

Nemlich Y, Greenberg E, Ortenberg R, Besser MJ, Barshack I, Jacob-Hirsch J, Jacoby E, Eyal E, Rivkin L, Prieto VG, et al. 2013. MicroRNA-mediated loss of ADAR1 in metastatic melanoma promotes tumor growth. J. Clin. Invest. 123:2703–2718.

Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, Shimizu M, Wojcik SE, Ferdin J, Kunej T, Xiao L, et al. 2010. Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. Cancer Res. 70:2789–2798.

Nigita G, Veneziano D, Ferro A. 2015. A-to-I RNA Editing: Current Knowledge Sources and Computational Approaches with Special Emphasis on Non-Coding RNA Molecules. Front. Bioeng. Biotechnol. 3:1–7.

Nishikura K. 2010. Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem. 79:321–349.

Nishikura K. 2016. A-to-I editing of coding and non-coding RNAs by ADARs. Nat. Rev. Mol. Cell Biol. advance on:83–96.

Niwa R, Slack FJ. 2007. The evolution of animal microRNA function. Curr. Opin. Genet. Dev. 17:145–150.

Nowick K, Gernat T, Almaas E, Stubbs L. 2009. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. Proc. Natl. Acad. Sci. U. S. A. 106:22358–22363.

Ohlson J, Pedersen JS, Haussler D, Ohman M. 2007. Editing modifies the GABA(A) receptor subunit alpha3. RNA 13:698–703.

Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. 2007. The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila. Cell 130:89–100.

Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420:563–573.

Olena AF, Patton JG. 2010. Genomic organization of microRNAs. J. Cell. Physiol. 222:540–545.

Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, et al. 2010. Long

noncoding RNAs with enhancer-like function in human cells. Cell 143:46–58.

Pachter L. 2012. A closer look at RNA editing. Nat. Biotechnol. 30:246–247.

Palade GE. 1955. A small particulate component of the cytoplasm. J. Biophys. Biochem. Cytol. 1:59–68.

Palazzo AF, Gregory TR. 2014. The case for junk DNA. PLoS Genet. 10:e1004351.

Park CY, Choi YS, McManus MT. 2010. Analysis of microRNA knockouts in mice. Hum. Mol. Genet. 19:169–175.

Park E, Williams B, Wold BJ, Mortazavi A. 2012. RNA editing in the human ENCODE RNA-seq data. Genome Res. 22:1626–1633.

Paz-Yaacov N, Levanon EY, Nevo E, Kinar Y, Harmelin A, Jacob-hirsch J, Amariglio N, Eisenberg E, Rechavi G. 2010. Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. Proc. Natl. Acad. Sci. U. S. A. 107:12174–12179.

Peng Z, Cheng Y, Tan BC-M, Kang L, Tian Z, Zhu Y, Zhang W, Liang Y, Hu X, Tan X, et al. 2012. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat. Biotechnol. 30:253–260.

Perdomo-Sabogal A, Kanton S, Walter MBC, Nowick K. 2014. The role of gene regulatory factors in the evolutionary history of humans. Curr. Opin. Genet. Dev. 29:60–67.

Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira M a M, Kessing B, Pontius J, Roelke M, Rumpler Y, et al. 2011. A molecular phylogeny of living primates. PLoS Genet. 7:1–17.

Peterson KJ, Dietrich MR, McPeek M a. 2009. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. Bioessays 31:736–747.

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.

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

Picardi E, Manzari C, Mastropasqua F, Aiello I, D'Erchia AM, Pesole G. 2015. Profiling RNA editing in human tissues: towards the inosinome Atlas. Sci. Rep. 5:14941.

De Pontual L, Yao E, Callier P, Faivre L, Drouin V, Cariou S, Van Haeringen A, Geneviève D, Goldenberg A, Oufadem M, et al. 2011. Germline deletion of the miR-17~92 cluster causes skeletal and growth defects in humans. Nat. Genet. 43:1026–1030.

Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, Macdonald PE, Pfeffer S, Tuschl T, Rajewsky N, Rorsman P, et al. 2004. A pancreatic islet-specific microRNA regulates insulin secretion. Nature 432:226–230.

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. Nature 499:471–475.

Prasanth K V., Prasanth SG, Xuan Z, Hearn S, Freier SM, Bennett CF, Zhang MQ, Spector DL. 2005. Regulating gene expression through RNA nuclear retention. Cell 123:249–263.

Quach H, Barreiro LB, Laval G, Zidane N, Patin E, Kidd KK, Kidd JR, Bouchier C, Veuille M, Antoniewski C, et al. 2009. Signatures of purifying and local positive selection in human miRNAs. Am. J. Hum. Genet. 84:316–327.

Qureshi IA, Mehler MF. 2011. Non-coding RNA networks underlying cognitive disorders across the lifespan. Trends Mol. Med. 17:337–346.

Ramachandran PV, Ignacimuthu S. 2013. RNA interference - A silent but an efficient therapeutic tool. Appl. Biochem. Biotechnol. 169:1774–1789.

Rana TM. 2007. Illuminating the silence: understanding the structure and function of small RNAs. Nat. Rev. Mol. Cell Biol. 8:23–36.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468:1053–1060.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli a E, Bettinger JC, Rougvie a E, Horvitz HR, Ruvkun G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403:901–906.

Rigoutsos I. 2009. New tricks for animal micrornas: Targeting of amino acid coding regions at conserved and nonconserved sites. Cancer Res. 69:3245–3248.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann S a., Goodnough LH, Helms J a., Farnham PJ, Segal E, et al. 2007. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. Cell 129:1311–1323.

Rogers J, Gibbs R a. 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. Nat. Rev. Genet. 15:347–359.

Rosenthal JJC, Bezanilla F. 2002. Extensive editing of mRNAs for the squid delayed rectifier K+ channel regulates subunit tetramerization. Neuron 34:743–757.

Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. Nature 448:83–86.

Rueter SM, Dawson TR, Emeson RB. 1999. Regulation of alternative splicing by RNA editing. Nature 399:75–80.

Saba R, Schratt GM. 2010. MicroRNAs in neuronal development, function and dysfunction. Brain Res. 1338:3–13.

Salehi-Ashtiani K, Lupták A, Litovchick A, Szostak JW. 2006. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. Science 313:1788–1792.

Salter JD, Bennett RP, Smith HC. 2016. The APOBEC Protein Family: United by Structure, Divergent in Function. Trends Biochem. Sci. xx:1–17.

Santpere G, Lopez-Valenzuela M, Petit-Marty N, Navarro A, Espinosa-Parrilla Y. 2016. Differences in molecular evolutionary rates among microRNAs in the human and chimpanzee genomes. BMC Genomics 17:528.

Saunders MA, Liang H, Li W-H. 2007. Human polymorphism at microRNAs and microRNA target sites. Proc. Natl. Acad. Sci. U. S. A. 104:3300–3305.

Sauvageau M, Goff L a., Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hacisuleyman E, Li E, Spence M, et al. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife 2013:1–24.

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. Nature 483:169–175.

Schmitz SU, Grote P, Herrmann BG. 2016. Mechanisms of long noncoding RNA function in development and disease. Cell. Mol. Life Sci. 73:1–19.

Schnall-Levin M, Rissland OS, Johnston WK, Perrimon N, Bartel DP, Berger B. 2011. Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs. Genome Res. 21:1395–1403.

Schratt GM, Tuebing F, Nigh E a, Kane CG, Sabatini ME, Kiebler M, Greenberg ME. 2006. A brain-specific microRNA regulates dendritic spine development. Nature 439:283–289.

Seitz H. 2004. A Large Imprinted microRNA Gene Cluster at the Mouse Dlk1-Gtl2 Domain. Genome Res. 14:1741–1748.

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. 2008. Widespread changes in protein synthesis induced by microRNAs. Nature 455:58–63.

Sempere LF. 2006. The Phylogenetic Distribution of Metazoan microRNAs : Insights into Evolutionary Complexity and Constraint. 588:575–588.

Shabalina S a., Koonin E V. 2008. Origins and evolution of eukaryotic RNA interference. Trends Ecol. Evol. 23:578–587.

Shen Y, Lv Y, Huang L, Liu W, Wen M, Tang T, Zhang R, Hungate E, Shi S, Wu C-I. 2011. Testing hypotheses on the rate of molecular evolution in relation to gene expression using microRNAs. Proc. Natl. Acad. Sci. U. S. A. 108:15942–15947.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29:308–311.

Shoshan E, Mobley AK, Braeuer RR, Kamiya T, Huang L, Vasquez ME, Salameh A, Lee HJ, Kim SJ, Ivan C, et al. 2015. Reduced adenosine-to-inosine miR-455-5p editing promotes melanoma growth and metastasis. Nat. Cell Biol. 17.

Sonay TB, Carvalho T, Robinson MD, Greminger MP, Krützen M, Comas D, Highnam G, Mittelman D, Sharp A, Marques-Bonet T, et al. 2015. Tandem repeat variation in human and great ape populations and its impact on gene expression divergence. Genome Res. 25:1591–1599.

Song JL, Stoeckius M, Maaskola J, Friedländer M, Stepicheva N, Juliano C, Lebedeva S, Thompson W, Rajewsky N, Wessel GM.

2012. Select microRNAs are essential for early development in the sea urchin. Dev. Biol. 362:104–113.

St.Laurent G, Wahlestedt C, Kapranov P. 2015. The Landscape of long noncoding RNA classification. Trends Genet. 31:249–251.

Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. 2005. Animal microRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell 123:1133–1146.

Sun G, Yan J, Noltner K, Feng J, Li H, Sarkis D a, Sommer SS, Rossi JJ. 2009. SNPs in human miRNA genes affect biogenesis and function. RNA 15:1640–1651.

Taft RJ, Pheasant M, Mattick JS. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 29:288–299.

Tanzer A, Stadler PF. 2004. Molecular evolution of a microRNA cluster. J. Mol. Biol. 339:327–335.

Tarver JE, Donoghue PCJ, Peterson KJ. 2012. Do miRNAs have a deep evolutionary history? BioEssays 34:857–866.

Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, Karreth F, Poliseno L, Provero P, Di Cunto F, et al. 2011. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell 147:344–357.

Teixeira A, Tahiri-Alaoui A, West S, Thomas B, Ramadass A, Martianov I, Dye M, James W, Proudfoot NJ, Akoulitchev A. 2004. Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. Nature 432:526–530.

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. Nature 526:68–74.

The International Hapmap Consortium. 2003. The International HapMap Project. Nature 426:789–796.

Tian N, Wu X, Zhang Y, Jin Y. 2008. A-to-I editing sites are a genomically encoded G: implications for the evolutionary significance and identification of novel editing sites. RNA 14:211–216.

Torruella-Loran I, Laayouni H, Dobon B, Gallego A, Balcells I, Garcia-Ramallo E, Espinosa-Parrilla Y. 2016. microRNA Genetic Variation, from Population Analysis to Functional Implications of Three Allele Variants Associated with Cancer. Hum. Mutat.

Treangen TJ, Salzberg SL. 2012. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. Nat. Rev. Genet. 13:36–46.

Válóczi A, Hornyik C, Varga N, Burgyán J, Kauppinen S, Havelda Z. 2004. Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. Nucleic Acids Res. 32:e175.

Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, Newman J, Bronson RT, Crowley D, Stone JR, et al. 2008. Targeted Deletion Reveals Essential and Overlapping Functions of the miR-17-92 Family of miRNA Clusters. Cell 132:875–886.

Vitsios DM, Enright AJ. 2015. Chimira: Analysis of small RNA sequencing data and microRNA modifications. Bioinformatics 31:3365–3367.

Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. Cell 136:669–687.

Wahlestedt C. 2013. Targeting long non-coding RNA to therapeutically upregulate gene expression. Nat. Rev. Drug Discov. 12:433–446.

Wall JD. 2013. Great ape genomics. ILAR J. 54:82–90.

Wang QH, Zhou M, Sun J, Ning SW, Li Y, Chen L, Zheng Y, Li X, Lv SL, Li X. 2010. Systematic analysis of human microRNA

divergence based on evolutionary emergence. FEBS Lett. 585:240–248.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57–63.

Warnefors M, Liechti A, Halbert J, Valloton D, Kaessmann H. 2014. Conserved microRNA editing in mammalian evolution, development and disease. Genome Biol. 15:R83.

Washburn MC, Hundley HA. 2016. Trans and cis factors affecting A-to-I RNA editing efficiency of a noncoding editing target in C. elegans. RNA 22:722–728.

Watson JD, Crick FHC. 1953. Molecular structure of nucleic acids. Nature 171:737–738.

Wheeler BM, Heimberg AM, Moy VN, Sperling E a, Holstein TW, Heber S, Peterson KJ. 2009. The deep evolution of metazoan microRNAs. Evol. Dev. 11:50–68.

Wideman JG. 2016. Losing Complexity: The Role of Simpli fi cation in Macroevolution. xx:1–14.

Wilkins MHF, Stokes a R, Wilson HR. 1953. Molecular structure of deoxypentose nucleic acids. Nature 171:738–740.

Williams a. E. 2008. Functional aspects of animal microRNAs. Cell. Mol. Life Sci. 65:545–562.

Williamson B. 1977. DNA insertions and gene structure. Nature 270:295–297.

Wong SK, Sato S, Lazinski DW. 2001. Substrate recognition by ADAR1 and ADAR2. RNA 7:846–858.

Wu C-I, Shen Y, Tang T. 2009. Evolution under canalization and the dual roles of microRNAs: a hypothesis. Genome Res. 19:734–743.

Xiao C, Srinivasan L, Calado DP, Patterson HC, Zhang B, Wang J, Henderson JM, Kutok JL, Rajewsky K. 2008. Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. Nat. Immunol. 9:405–414.

Yang J-S, Lai EC. 2010. Dicer-independent, Ago2-mediated microRNA biogenesis in vertebrates. Cell Cycle 9:4455–4460.

Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R, Nishikura K. 2006. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nat. Struct. Mol. Biol. 13:13–21.

Yang Y, Zhou X, Jin Y. 2013. ADAR-mediated RNA editing in non-coding RNA sequences. Sci. China. Life Sci. 56:944–952.

Yi R, Qin Y, Macara IG, Cullen BR. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 17:3011–3016.

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

Zhang R, Peng Y, Wang W, Su B. 2007. Rapid evolution of an X-linked microRNA cluster in primates. Genome Res. 17:612–617.

Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 9.

Zhao Y, Ransom JF, Li A, Vedantham V, von Drehle M, Muth AN, Tsuchihashi T, McManus MT, Schwartz RJ, Srivastava D. 2007. Dysregulation of Cardiogenesis, Cardiac Conduction, and Cell Cycle in Mice Lacking miRNA-1-2. Cell 129:303–317.

Zhou H, Rigoutsos I. 2014. MiR-103a-3p targets the 5' UTR of GPRC5A in pancreatic cells. RNA 20:1431–1439.

Zhu Y, Skogerbø G, Ning Q, Wang Z, Li B, Yang S, Sun H, Li Y. 2012. Evolutionary relationships between miRNA genes and their activity. BMC Genomics 13:718.

Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, Yeo GW. 2010. Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. Nat. Struct. Mol. Biol. 17:173–179.

Zorc M, Skok D, Godnic I, Calin GA, Horvat S, Jiang Z, Dovc P, Kunej T. 2012. Catalog of microRNA seed polymorphisms in vertebrates. PLoS One 7.

VI. SUPPLEMENTARY MATERIAL

10. Functional implications of human-specific changes in great ape microRNAs

Alicia Gallego, Marta Melé, Ingrid Balcells, Eva García-Ramallo, Ignasi Torruella-Loran, Hugo Fernández-Bellon, Teresa Abelló, Ivanela Kondova, Ronald Bontrop, Christina Hvilsom, Arcadi Navarro, Tomàs Marquès-Bonet, Yolanda Espinosa-Parrilla

Published

Functional Implications of Human-Specific Changes in Great Ape microRNAs.

PLoS ONE 11(4): e0154194. doi:10.1371/journal.pone.0154194

11. microRNA editing across human tissues and primate species: the case of mir-376a1

Alicia Gallego, Diego A. Hartasánchez, Marina Brasó-Vives, Eva Garcia-Ramallo, Maria Lopez-Valenzuela, Neus Baena, Miriam Guitart, Hugo Fernández-Bellon, Ivanela Kondova, Ronald Bontrop, Yukio Kawahara, Yolanda Espinosa-Parrilla

Submitted for publication

Supplemental Figures

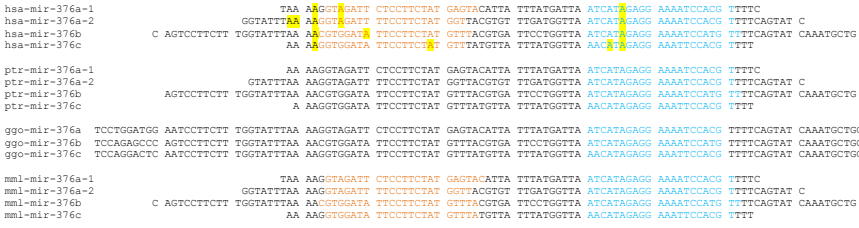


Figure S1. Precursor miRNA sequences of mir-376 family in the primate species analyzed. Orthologous sequences (5' - 3') of mir-376 family (miRBase, release 21) are shown for human (hsa), chimpanzee (ptr), gorilla (ggo) and macaque (mml). Annotated regions processed into mature miRNAs are shown in orange (-5p strand) and blue (-3p strand). Edited adenosine residues previously described in humans are highlighted in yellow. [1]

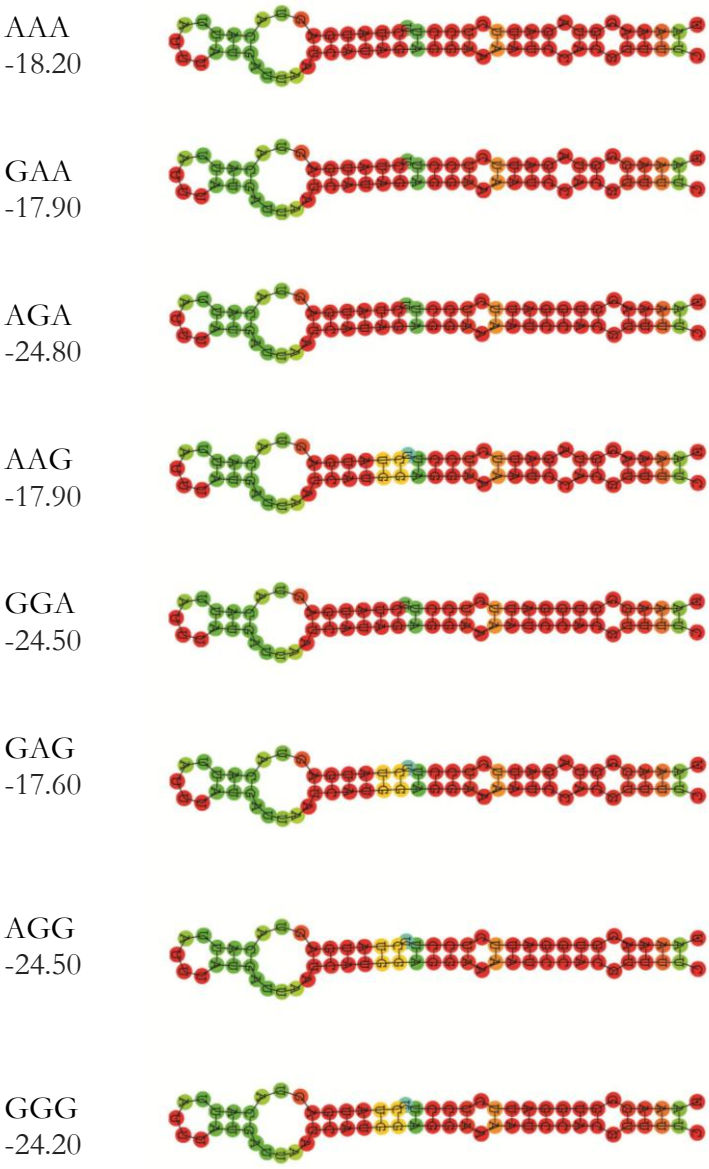


Figure S2. Secondary structures and minimum free energy (MFE). MFE values in kcal/mol and hairpins predicted by RNAfold for each precursor hsa-mir-376a1 edited isoforms.

Supplemental Tables

Table S1. Expression levels of mature miRNAs, miR-376a1-5p and -3p, based on RT-qPCR. Expression levels of miR-376a1-5p, miR-376a1-3p and the reference control gene miR-25-3p, based on the quantification cycle (Cq) in the RT-qPCR. Human brain cortex and placenta data come from six individuals, chimpanzee data from two individuals, while only one individual was evaluated for gorilla and macaque cases. P-values of Student t-test comparisons between -5p and -3p expression relatively to the control miR-25-3p, are shown for species represented by more than two individuals. SD: Standard deviation.

Cq values	miR-376a1-5p	SD	miR-376a1-3p	SD	miR-25	SD	p-value
Human placenta	24.9	1.01	22.9	0.84	20.5	0.40	**0.004**
Human cortex	24.4	0.64	24.5	0.65	21.1	0.45	0.892
Chimpanzee cortex	25.2	0.35	24.7	0.29	21.6	0.11	--
Gorilla cortex	26.2	--	26.1	--	21.6	--	--
Macaque cortex	25.2	--	24.7	--	21.8	--	--

Table S2. Expression levels of miR-376a1-5p and -3p, based on small RNA-seq. Total sequencing reads were normalized across samples using DESeq2. Identical matches between small RNA sequences with more than one precursor sequences (i.e. miRNA paralogs) were assigned only to the first optimal alignment call returned by BLASTn.

Plain counts	Human placenta	Human cortex 1	Human cortex 2	Chimpanzee cortex	Gorilla cortex	Macaque cortex
miR-376a1-5p	1054	185	495	173	224	130
miR-376a1-3p	1318	150	277	127	172	86

Table S3. Editing frequencies of primary mir-376a1 sites based on Sanger sequencing. Editing percentage averages in the primary mir-376a1 at positions -1, +4 and +44. Human brain cortex and placenta data come from six individuals, chimpanzee data from two individuals, while only one individual was evaluated for gorilla and macaque cases. SE: Standard error.

Samples	-1 site		+4 site		+44 site	
	% Editing	SE	% Editing	SE	% Editing	SE
Human placenta	5	2.6	10.8	3.3	9.2	3.7
Human cortex	10.8	3.0	31.7	5.9	25.8	7.7
Chimpanzee cortex	35	5	55	10	35	10
Gorilla cortex	16	--	42	--	26	--
Macaque cortex	10	--	60	--	25	--
All primate cortex[1]	16	11.7	40.5	16.7	28	15.3

[1]Editing percentage average among all human, chimpanzee, gorilla and macaque brain cortex samples

Table S4. Edited isoform frequencies per individual. Editing frequencies of primary mir-376a1 isoforms evaluated by Sanger sequencing in 20 colonies and shown as percentages. P = human placenta (six individuals); H = human brain cortex (six individuals); C = chimpanzee brain cortex (two individuals); G = gorilla brain cortex (one individual); M = macaque brain cortex (one individual). Isoforms are represented by their three editable sites: -1, +4 and +44, of the primary molecule.

Isoforms			Human placenta samples						Brain cortex samples									
-1	+4	+44	P1	P2	P3	P4	P5	P6	H1	H2	H3	H4	H5	H6	C1	C2	G	M
A	A	A	90	80	70	75	75	75	50	70	45	20	30	75	35	15	40	35
A	A	G	5	5	25	15	-	5	10	15	10	25	20	0	-	15	10	5
A	G	A	5	5	5	10	10	15	10	10	25	20	20	20	10	15	20	30
G	A	A	-	10	-	-	-	5	10	5	-	10	5	5	10	5	-	-
G	G	A	-	-	-	-	15	-	10	-	5	-	-	-	20	20	10	10
G	A	G	-	-	-	-	-	-	-	-	5	-	-	-	10	-	5	-
A	G	G	-	-	-	-	-	-	5	-	10	25	20	-	15	25	15	20
G	G	G	-	-	-	-	-	-	5	-	-	-	5	-	-	5	-	-

Table S5. Analysis of epistatic effects among edited sites. (A) Editing rates associated to each site independently calculated for each site. Probability values for positions -1, +4 and +44, in the primary mir-376a1, calculated considering the fraction of isoforms edited at only one position over the total number of isoforms edited in one position (e.g. $p_{+44} = f_{AAG} / (f_{GAA} + f_{AGA} + f_{AAG})$, for position +44). **(B) Editing rates associated to each site considering editing events in all sites.** Probability values for positions -1, +4 and +44, in the primary mir-376a1, calculated considering the fraction of editing events (i.e. the number of A-to-G transitions) that occurred in a particular position over all editing events (e.g. $p'_{+44} = (f_{AAG} + f_{AGG} + f_{GAG} + f_{GGG}) / (f_{GAA} + f_{AGA} + f_{AAG} + 2f_{GGA} + 2f_{GAG} + 2f_{AGG} + 3f_{GGG})$, for position +44). **(C) Editing rates equalities associated to each site.** Relative differences between editing rates calculated for each site independently (1EditProp, **Table S5**) and calculated considering editing events in all sites (AllEditProp, **Table S6**) for every site and for the sum of the three sites (W) in the primary mir-376a1.

A

Edited site	Human placenta	Human cortex	Chimpan zee cortex	Gorilla cortex	Macaque cortex	Non-human primate cortex
-1	0.125	0.159	0.273	0.000	0.000	0.125
+4	0.417	0.477	0.456	0.667	0.857	0.625
+44	0.458	0.364	0.273	0.333	0.143	0.250

B

Edited site	Human placenta	Human cortex	Chimpanzee cortex	Gorilla cortex	Macaque cortex	Non-human primate cortex
-1	0.200	0.159	0.280	0.167	0.105	0.218
+4	0.433	0.463	0.440	0.500	0.632	0.494
+44	0.367	0.378	0.280	0.333	0.263	0.288

C

Edited site	Human placenta	Human cortex	Chimpanzee cortex	Gorilla cortex	Macaque cortex	Non-human primate cortex
-1	0.02812	1.938E-006	0.00019	0.16667	0.02812	0.03994
+4	0.00064	0.00041	0.00048	0.05556	0.00064	0.03459
+44	0.02292	0.00055	0.00019	0.00000	0.02292	0.00486
Total (W)	0.05168	0.00097	0.00086	0.22222	0.24082	0.07938

Table S6. Orthologous miRNA names in the studied primate species. Corresponding orthologous mir-376a1 sequences in all species analyzed and named as their annotations in miRBase (release, 21).

Species	Precursor miRNA	Mature -5p	Mature -3p
Human *(Homo sapiens)*	hsa-mir-376a1	hsa-miR-376a1-5p	hsa-miR-376a1-3p
Chimpanzee *(Pan troglodytes)*	ptr-mir-376a1	ptr-miR-376a1-5p	ptr-miR-376a1-3p
Gorilla *(Gorilla gorilla)*	ggo-mir-376a	ggo-miR-376a-5p	ggo-miR-376a-3p
Macaque *(Macaca mulatta)*	mml-mir-376a1	mml-miR-376a1-5p	mml-miR-376a1-3p

Table S7. Primer sequences. Forward (FW) and reverse (RV) primers (5' - 3') used in this work.

Primer name	Primer sequence	Use
miR-376a1-5p	FW: CGCAGGTAGATTCTCCTTCT RV: GGTCCAGTTTTTTTTTTTTTTTACTC	RT-qPCR
miR-376a1-3p	FW: CGCAGATCATAGAGGAAAATCC RV: TCCAGTTTTTTTTTTTTTTTACGTG	RT-qPCR
miR-25-3p	FW: CATTGCACTTGTCTCGGT RV: GGTCCAGTTTTTTTTTTTTTTTCAGA	RT-qPCR
pri-mir-376a1	FW: GCTTTCCTGGATGGAATCCT RV: CGAGGTTTTCAAAGCAGCAT	RT-PCR
miR-376a1-5p	RV: GTCGGTGATTACTCATAGAAGGAG	RT-PCR
M13	FW: GTAAAACGACGGCCAG RV: CAGGAAACAGCTATGAC	PCR screening and Sanger sequencing

REFERENCES

1. Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. 2007. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science. 315:1137–40.