eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

ZTF-FCT
Zientzia eta Teknologia Fakultatea
Facultad de Ciencia y Tecnología
Euskadi 2006

Degree's Final Project / Gradu Amaierako Lana
Degree in Biology / Biologiako Gradua

Sequencing of snoRNAs from multiple sclerosis association regions and characterization of SNPs./Esklerosi anizkoitzaren asoziazio guneetako snoRNAen sekuentziazioa eta SNPen karakterizazioa.

Author/Egilea:
Mikel Muñoz Oreja

Directors/Zuzendariak:
Dr. David Otaegui Bichot (IIS BioDonostia)
Dr. Begoña Marina Jugo Orrantia (UPV/EHU)

Leioa, 2016 June 28th /Leioa, 2016ko ekainaren 28a

ABSTRACT

Multiple sclerosis (MS) is a common inflammatory and neurodegenerative disease that causes neurological disability. Transcriptome regulation has been seen affected in MS patients' blood cells. Recently, non-coding RNAs (ncRNAs) have emerged as prominent transcriptome regulators, and in turn, they could play an important role in MS pathogenesis. From this point of view, the multiple sclerosis group in IIS BioDonostia has already analyzed the expression of protein-coding RNAs and some ncRNAs, especially miRNAs, in search of putative MS-related genes. These studies highlighted some possible candidates, amongst others, the snoRNA SNORA40. Thus, the aim of this work is to analyze snoRNAs located in MS-associated regions of the genome and further investigate the possible MS relationship of SNORA40. For that purpose, after detecting the candidate snoRNAs and optimizing the conditions for amplifying them, we did the sequencing of 10 snoRNAs in 14 siblings from 7 Gipuzkoan families and the sequencing of SNORA40 in 47 samples previously included in a RNA expression study. That way, we could identify both already described and novel genetic variants in some of the snoRNAs, but could not relate those variants to the disease as to our samples, ruling out a possible Mendelian inheritance way of those genetic variants.

LABURPENA

Esklerosi anizkoitza (EA) ohiko gaixotasun inflamatorio eta neuroendekatzaile bat da, eta ezgaitasun neurologikoa sortzen du. Transkriptomaren erregulazioan aldaketak behatu izan dira EA duten gaixoen odoleko zeluletan. Azken aldian ikusten ari garenez, RNA ez kodetzaileek (ncRNAs) garrantzia dute transkriptomaren erregulazioan, eta baliteke eragin handia izatea EAren patogenesian. Ildo horretatik, IIS BioDonostiako esklerosi anizkoitzaren taldeak dagoeneko ikertu ditu RNA proteina-kodetzaileen eta zenbait ncRNAen adierazpenak, bereziki miRNAenak, EAri erlazionaturiko gene posibleen bila. Ikerketa horiek gene hautagai batzuen izenak eman zituzten, hala nola, SNORA40 snoRNArena. Horrela, lan honen helburua da EAri asoziaturiko genomako guneetan dauden snoRNAen sekuentziak aztertzea eta SNORA40k EArekin izan dezakeen erlazioa gehiago ikertzea. Horretarako, snoRNA hautagaiak detektatu, eta horiek anplifikatzeko baldintzak prestatu ondoren, 10 snoRNA eta SNORA40 sekuentziatu genituen; lehena, 7 familia gipuzkoarren 14 anai-arrebetan, eta, bigarrena, aurrez RNAren adierazpen maila aztertu zen 47 laginetan. Era horretan, snoRNA horietako batzuetan, identifikatu ahal izan genituen jada deskribaturiko eta berriak ziren aldaki genetikoak, baina ezin izan genituen aldaki horiek gaixotasunarekin erlazionatu gure laginetan, eta, hala, ustezko herentzia mendeliar posible bat baztertu genuen.

INDEX

Page

INTRODUCTION

Disease description

Multiple sclerosis (MS) is a central nervous system (CNS) autoimmune disease. This means that an aberrant immune response is induced in genetically susceptible persons. However, the etiology of the illness still remains unknown and the exact causes for having multiple sclerosis are not fully understood. It is believed to be a complex disease, meaning different factors make the disease to come out, mainly environmental factors and a complex genetic background (Tullman, 2013). Among the environmental factors, high sunlight and ultraviolet radiation exposure, vitamin D deficiency, diet, geomagnetism, cigarettes, Epstein-Barr virus and other viruses or infective agents have been described (Compston and Coles 2008).

The main steps in the pathogenesis of MS start with inflammation driven by lymphocytes. This inflammation causes the blocking of the propagation of the action potential impeding the saltatory conduction of the electrical current through the axons in the CNS; mainly, because of the destruction of the myelin sheets that encircle the axons. To drive the inflammatory process myelin-reactive T cells must activate and have a memory phenotype. This is indeed what differentiates from the same type of cell found in controls because non-activated myelin-specific autoreactive T cells are found in the peripheral blood and cerebrospinal fluid (CSF) both in MS patients and in controls with similar frequencies (Comabella and Khoury 2012). The activated state of those T cells is thought to be associated with an upregulation of adhesion molecules that make these cells more devoted to interact with the blood-brain barrier (BBB) and hence to enter from the blood to the CNS. However, the way these autoreactive cells become activated in the peripheral blood remains elusive, but the fact that these cells from MS patients are activated in the periphery has given rise to the hypothesis that MS is an autoimmune disease and particularly, that there is a deficient immunoregulatory control rather than an increased generation of those cells (Comabella and Khoury 2012). After lymphocytes attack the axons, microglia is activated and takes part in inflammatory processes as well as in repairing the demyelination by removing the misfunctional myelin debris and by promoting remyelination. The key role here it is played by the oligodendrocyte precursor cells (OPC). Inflammatory cells and microglia act as signals for OPCs that are recruited to the proximity of demyelinating axons. These OPCs are the source of cells having the potential to differentiate in mature oligodendrocytes (OL) and thus, remyelinate naked or damaged axons (the lesions). With the time, reparation capacity decays, remyelination no longer being successful and thus, tissue repair does not occur, leading to chronic neurodegeneration (Compston and Coles 2008).

Patients with MS have a wide variety of clinical course (Fig.1). Approximately 85% of patients show a relapsing-remitting (RRMS) course of the disease, at least at the initial phase of the disease. This form of MS is characterized by acute relapses that last from weeks to months, where symptoms are

highly manifested (mainly: numbness, tingling, weakness, impaired balance, blurred vision, double vision, vertigo, and bladder or bowel dysfunction) and alters the neurological functioning. After that, a phase of disease remission comes where the symptoms may disappear or improve. This remission is indefinite in time. Most of RRMS patients, however, enter a progressive phase after a period of time, known as secondary-progressive (SPMS) where there is a gradual worsening of neurological disability with or without relapses. There is another type of progressive course, called primary-progressive (PPMS) that only 10% of patients show. Here, the gradual disability occurs from the beginning and although there might be slight fluctuations, there is not such a relapse or remitting phase. There is a last clinical course, related to PPMS known as progressive relapsing (PRMS). About 15% to 40% of patients with PPMS experience at least one relapse during the course of the disease but there are no periods of remission. This last one is the least common type of MS (Tullman, 2013). Patients having a single encounter with the disease but have not been yet diagnosed with MS are referred to as clinically isolated syndromes (CIS) and it is considered the first manifestation of the illness (Comabella and Khoury 2012; Stys *et al.* 2012).

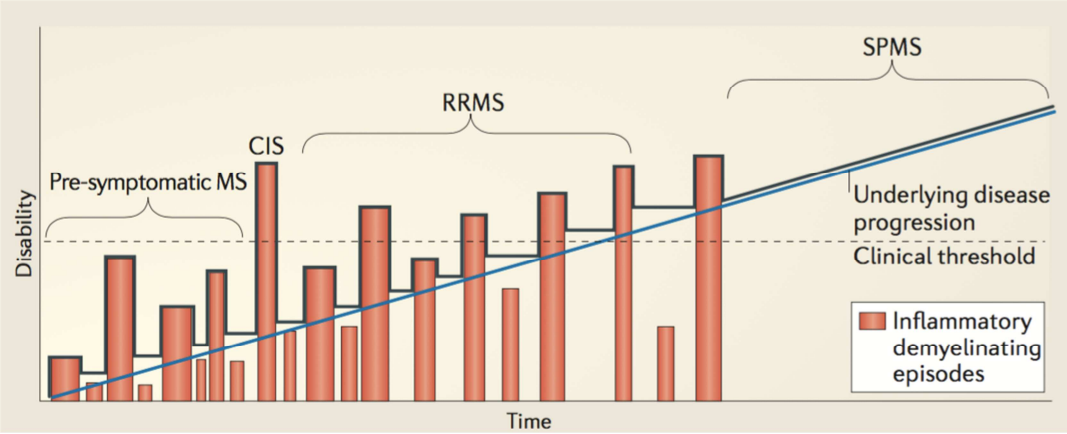


Figure 1. The graphical course of the disease visualized as time (x) and the disability grade (y) and the different type of MS, excluding PPMS. The orange bars indicate the inflammatory relapse events. The blue line wants to show that no matter what type of MS, the progression of all the disease-course types is uniform (Stys *et al.* 2012).

Therapies

During the last decades many and very different therapies or treatments have been developed but many of these have also been disapproved later on. Nowadays, the available treatments are usually expensive and do not always show effectiveness. One of the pioneers and most common of the treatments are type I interferons, first used in the 1970s. Gamma interferon promoted relapses whilst beta interferon reduced relapse frequency. At some point, the other candidate at the time, azathioprine, was proved to be toxic and insufficiently effective but another drug, glatiramer acetate, was promoted. At present, the use of beta interferons and glatiramer acetate is less common in RRMS patients. Apparently, evidence show that interferons are useful to reduce relapse frequency during the first year

of treatment only, and afterwards, no convincing efficacy has been shown, neither any effect on the accumulation of disability. Many other drugs are on the market, but two of them are especially relevant. The first one is mitoxantrone (an anthracenedione antineoplastic drug which intercalates with DNA and inhibits both DNA and RNA synthesis) but its use is very limited due to its toxic side-effects (such as cumulative cardiotoxicity or acute leukaemia in some patients). The second one is an antibody, called natalizumab (Tysabri ®), against the alfa4beta1 integrin of the surface of lymphocytes (Compston and Coles 2008; Hauser *et al.* 2013). The migration from the peripheral blood to the CNS occurs through the BBB and it is a various step process, where mostly adhesion molecules, chemokines and matrix metalloproteinases are involved. One of these adhesion molecules is alfa4beta1 integrin and it is the target molecule in the wide-used drug natalizumab (Tysabri ®), and so, its blocking with anti-alfa4beta1 reduces the migration of immune cells to the CNS and disease activity (Comabella and Khoury 2012) Natalizumab has shown greater efficacy than the rest of the drugs mentioned above. However, it has also been demonstrated that it can cause as a rare side effect, progressive multifocal leucoencephalopathy (PML) a severe demyelinating disease of the central nervous system (CNS) caused by the reactivation of a latent infection of JC virus (JCV)(Warnke *et al.* 2010; Muñoz-Culla *et al.* 2014).

The genetics of multiple sclerosis

In the 1970s the first genetic factor related to MS was discovered and it was the human leucocyte antigen (HLA) locus located within the major histocompatibility complex (MHC) of the genome. This kind of data was provided by studies of twins and sibling pairs, suggesting that genetic factors influence susceptibility to MS (Hafler *et al.* 2007). Inside the HLA, during decades the association to the disease was owed, almost exclusively, to the HLA class II region of the HLA-DR2 haplotype, for having the strongest association with MS, but also because the attributable risk of other non-HLA alleles is very small (Hoppenbrouwers and Hintzen 2010; Muñoz-Culla *et al.* 2013). At the time, linkage studies were common and they showed that the only region in the genome showing linkage with MS was the MHC region, but no other regions (Hafler *et al.* 2007). Later on, it was demonstrated that linkage studies in MS lack the statistical power to detect susceptible loci outside the HLA region, and hence, association studies were shown to have greater statistical power to detect common genetic variants. In this aspect, genome-wide association studies (GWAS) have been the best option to look for different genetic variants, such as single nucleotide polymorphisms (SNPs), in hundreds or thousands of genotypes from different samples in an objective, hypothesis-free way. The aim of a GWAS is to associate the region those SNPs lie in the genome to different diseases (Hafler *et al.* 2007). The GWAS performed by Hafler et al., concluded that the allelic variants outside the HLA region associated with MS, were not rare mutations, as it occurs in monogenic diseases, but rather polymorphic variants that even though happening in healthy controls too, each of them is more common in MS patients and each of them has a small effect on the total risk of having the illness

(Hafler *et al.* 2007). Another worth-mentioning aspect of the genome, are the DNase I hypersensitive sites (DHSs). These sites are uncondensed chromatin genome-parts that enable transcription factors to bind to their cis-regulatory elements. In 2012 Maurano *et al.* realized that from the 76.6% of all SNPs detected from GWAS in non-coding regions, the vast majority lied within a DHS or was in a linkage disequilibrium (LD) with SNPs in a nearby DHSs. Their data also showed that numerous DHSs containing GWAS SNPs were connected with promoters from distant genes, and the other way around, apparently unconnected variants associated with related diseases were connected between them for sharing common transcription factor networks (Maurano *et al.* 2012). In the same way, more GWAS and mapping studies have been of great value to understand that more than 90% of autoimmune diseases (AID)-associated SNPs are found within non-coding regions of the genome (Ricaño-Ponce et al., 2016). More than 40% of these SNPs affect the expression levels of nearby (cis-eQTLs) or far away (trans-eQTLs) located protein-coding genes and what is curious is that although different AIDs share same disease-associated loci, different genes in these loci can affect different AIDs (Ricaño-Ponce et al., 2016). All this information, makes us reconsider the information provided by GWAS, due to the fact that GWAS are based on SNPs that are representative of regions of the genome where the candidate genes selected are usually protein-coding genes and now, we should keep in mind that non protein-coding genes can also be candidate genes. All these data suggests that MS probably covers many different molecular pathologies all named as MS because of similar manifestation of symptoms that converge at the clinical level (Muñoz-Culla *et al.* 2013). From this point of view, although it is widely accepted MS is a complex disease (meaning many genes with small effect make the disease to come up), the HLA region and the other genes hitherto described explain only about the 30 % of the heritability, the rest being unknown and named as “missing heritability”. A way to explain that heterogeneity and try to find, at least, a part of the “missing heritability “, could be a Mendelian form of MS that is based in pathogenic mutations that make the disease to come out in some families (Lill 2014; Hoppenbrouwers and Hintzen 2010; Wang *et al.* 2016).

Nowadays, besides GWAS, whole genome and microarray expression studies are taking more importance, as well as, functional studies and next generation sequencing (NGS) techniques. The future challenges may focus on epigenetic alteration studies that link genetic background and environmental factors (Muñoz-Culla *et al.* 2013; Hoppenbrouwers and Hintzen 2010).

Non protein-coding genes: miRNAs and snoRNAs.

The proportion of the genome that accounts for the exons or protein-coding regions, is only the 1.5-2% (Esteller 2011). The whole rest is called the non protein-coding region of the genome and despite it has been thought to be unfunctional or of non-use (as it was called junk DNA for many decades), nowadays, it is known to be crucial for normal development and physiology of the individuals, but

also in the development of many diseases. That non-protein-coding region, when transcripted, gives room to many different types of RNAs, known as non-coding RNAs (ncRNAs). Although, it is not clear yet the division of the different ncRNA classes, it is quite common to separate them in three groups according to their size. The short ncRNAs are about 17-31 bp long and are the most studied ones, which include microRNAs (miRNAs). The mid-size ncRNAs are considered the ncRNAs that are smaller than 200 bp but bigger than 20 bp and finally, the long ncRNAs are those ones bigger that 200 bp. Long ncRNAs include amongst others, lincRNAs, and mid-size ncRNAs include, small nucleolar RNAs (snoRNAs)(Esteller 2011).

These last ones are particularly relevant in this work, and so, I am going to explain them in further detail. snoRNAs have 60 to 300 bp and they are part of small nucleolar ribonucleoproteins (snoRNPs). Indeed, snoRNAs are the sequences responsible for targeting the whole snoRNP structures to the specific target. snoRNPs, likewise, are vital proteins in the processing of ribosomal RNA (rRNA)(Esteller 2011). In eukaryotes, snoRNAs are predominantly located in introns and they usually express under the control of the promoter of the protein-coding gene they belong to, but there have also been described some snoRNAs that transcribe under the control of independent promoters (Scott and Ono 2011). After the transcription, snoRNAs within pre-messenger RNAs (pre-mRNAs), undergo the splicing process and so, become independent, but yet they need to bear debranching and trimming processes and assembly with snoRNPs in order to become mature and functional snoRNAs (Esteller 2011).

The processing of the rRNA snoRNAs make are generally post-transcriptional modifications that take place in the nucleolus and are usually 2'-O-methylation and pseudouridylation, which help rRNA to be folded and more stable (Esteller 2011). Methylation is made by a type of snoRNAs called box C/D snoRNAs (SNORDs) and pseudouridylation is made by another type of snoRNAs known as box H/ACA snoRNAs (SNORAs) and they differ mainly in the sequence length, specific motifs and the type of snoRNPs they bind to (Scott and Ono 2011). But sometimes, snoRNAs can also remain in the nucleus and take part in alternative splicing of the mRNA or several other unknown functions, yet to be discovered (Esteller 2011). For instance, a type of these nuclear snoRNAs are the so called Cajal body-specific snoRNAs (scaRNAs). These scaRNAs compile and function in the small membrane-less nuclear compartments, called Cajal bodies, and they mostly make post-transcriptional changes to the small nuclear RNAs (snRNAs)(Scott and Ono 2011). Besides, it has been demonstrated that many snoRNAs are involved in the regulation of gene expression, as they can further be processed in a Drosha-independent and Dicer-independent manner and so, they can originate new small RNAs, called snoRNA-derived RNAs (sdRNAs)(Ender *et al.* 2008). These sdRNAs are thought to be involved in the gene silencing effector complexes, as they have mostly been described as miRNA-like functioning sdRNAs. In fact, it has been estimated that around %60 of human snoRNAs are precursors

of sdRNAs with potential miRNA-like function. However, this is not the only function described in sdRNAs; a wide variety of sdRNAs spectrum has been proposed within the snoRNA and miRNA families: prototypical snoRNAs, snoRNAs that can function as miRNAs, dual function sno-miRNAs, miRNAs with snoRNA-like functionality and prototypical miRNAs (Irizar *et al.* 2015).

All these different functions, can affect normal cell functioning, and so, many diseases have been reported related to snoRNAs. To cite some, it has been seen a downregulation of snoRNAs in meningiomas, it has been described a differential expression of snoRNAs in non-small-cell lung cancer or it has been reported that a 2 bp homozygous deletion (TT) of the snoRNA U50 can produce either prostate cancer or, under some other heterozygous deletions and transcriptional downregulation, it can be associated with breast cancer development (Esteller 2011). Accordingly, MS could also be another possible illness in which these snoRNAs interfere.

The context and background

The present work was done as a part of different tasks done in IIS BioDonostia during July and August 2015. I was taught and tutored by the Multiple Sclerosis Group (Neuroscience Area) under the guidance of Dr. David Otaegui (as the director) and Dr. Maider Muñoz (as the responsible). As this work is part of a long-lasting project, for its full understanding, I feel obliged to explain the whole background and so, I am going to mention the group's main publications in our context. Two main lines should differentiate; the first one the line that investigates SNORA40 and another line that investigates the rest of the snoRNAs. I am going to explain both of them separately in further detail but in this work specifically, as far as my own duties extend, I have decided to consider both as part of the same project, as the reader will notice in ongoing pages.

First of all, in order to identify potential therapeutic targets, a co-expression network analysis and a differential expression analysis were made by Irizar *et al.*, 2015. mRNA and small non-coding RNA (sncRNA) was extracted from peripheral blood leukocyte samples from MS patients (both relapse and remitting samples) and healthy controls (HC). They identified differentially expressed sncRNAs and mRNAs in relapse vs. remission and remission vs. controls. After that a global ncRNA-mRNA co-expression network was made and status-specific networks (one for each disease condition). They identified a status-independent core network that appeared in all four networks where ten sncRNAs formed the backbone of that network. Finally, they built a disease-specific network that excluded the controls and compared the answers obtained from that network with those seen on the differential expression of sncRNAs, so as to detect the candidate sncRNAs in MS (Fig. 2). The results show that many snoRNAs may have a key role in gene silencing as revealed by the fact that many of them have a high number of connections to mRNAs and present central positions in the networks. This central positioning has been correlated to the essentiality of the snoRNA and so, Irizar *et al*. conclude giving a

list of 3 miRNAs (miR-20b, miR-331-5p and miR-1246) and a snoRNA (SNORA40) as top therapeutic target candidates for MS. Nevertheless, SNORA40 is thought to be the most prominent candidate above all the rest. Not only SNORA40 has been related to aberrant responses of the immune system (such as, asthma and MS), but also to the NK92 cells (which respond to malaria infected red blood cells) and to the antibody-secreting B plasma cells from patients with multiple myeloma (Irizar *et al.* 2015).

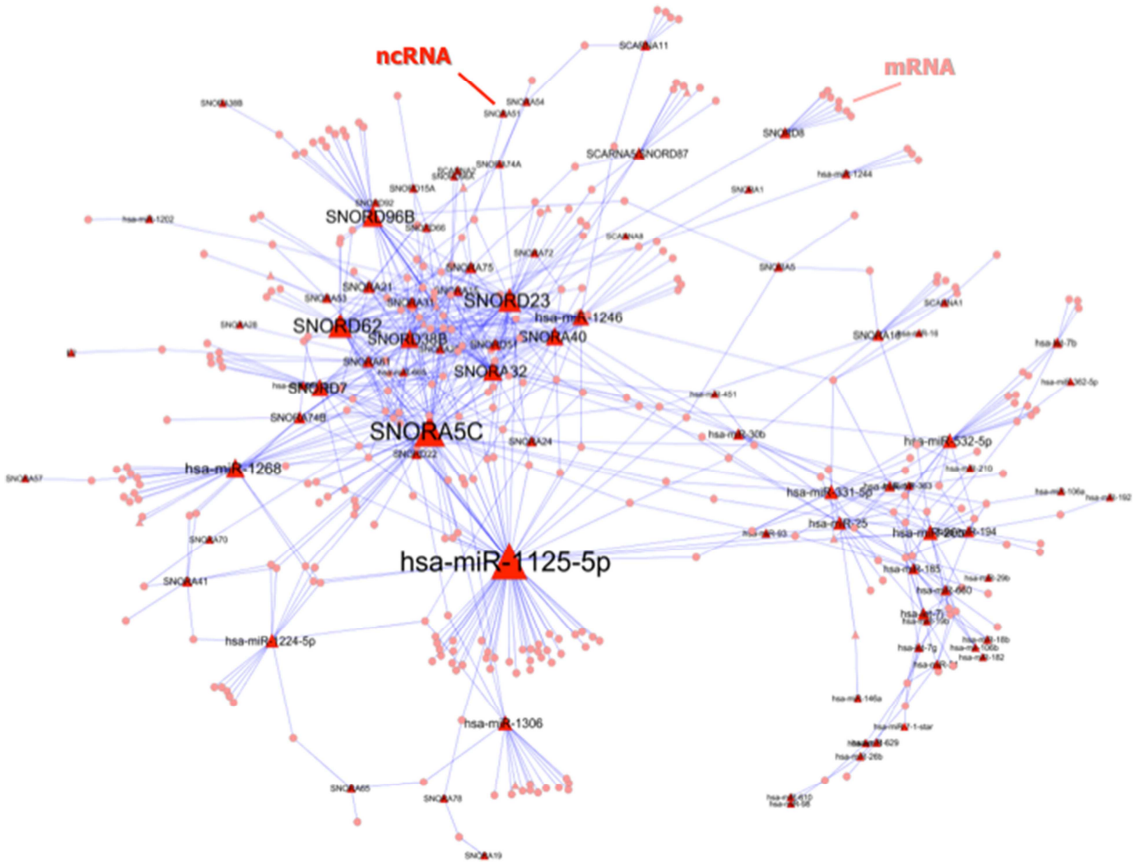


Figure 2. The resulting network of 401 nodes and 742 edges built from the 2307 edges shared by the relapse and the remission networks but not the controls'. Amongst others, SNORA40 occupies central positioning in the network.

The second work highlighting the importance of SNORA40 is the paper of Muñoz-Culla *et al*., 2016. In this study, blood samples were taken from 24 MS patients (both in relapse and remission) and 24 HC. RNA was extracted in order to obtain small RNAs and it was hybridized to an array, which covers many miRNAs and snoRNAs. The results were compared in a relapse vs. remitting and remitting vs. HC way and in the first comparison, only the patients that kept on with the treatment in relapse and in remission were taken (13 out of 24). This results were compared to those obtained from cultured peripheral blood mononuclear cells (PBMC) from a MS patient and a HC. Two subgroups of PBMC culture were made, one was induced activation by phytohemaglutinin (PHA) so as to resemble to a relapse-like state, and the other was not keeping it as a control. That way, similar conditions to

RRMS patients' cells were created and so, facilitate comparisons. Finally, miRNA-mRNA interaction networks were created, using mRNA gene expression data obtained from the same samples, as previously mentioned. The results reveal that sncRNAs express differentially in each of the states of the disease and that the set of the sncRNAs differentially expressed is not the same in males and females. In remission, the sncRNAs deregulated in females is bigger than in males (42 to 7) but the difference is even bigger in relapse, where no sncRNAs are found to be deregulated in males (and 38 in female). This seems to be due to the regulation of sexual hormones in sncRNAs and it is thought to be related to different disease susceptibility and clinical presentation and progression observed in males and females with MS but also, in other autoimmune diseases (Muñoz-Culla *et al.* 2016). If this was not enough, Muñoz-Culla *et al.* observed that the expression of the sncRNAs in remission and relapse was different from the expected. They realized that the sncRNAs that were altered in both desease statuses were deregulated in opposite directions, meaning that if a given sncRNA was overexpressed in remission it was underexpressed in relapse and the other way around (Fig. 3). They called this phenomenon "the mirror pattern". One of the sncRNAs showing the highest deregulation in females during relapse and that also showed the mirror pattern was SNORA40, overexpressed in females during relapse, with a fold-change of 9.21, in comparison to the remitting females (Muñoz-Culla *et al.* 2016). Consequently, both works, underline the importance of snoRNAs in MS.

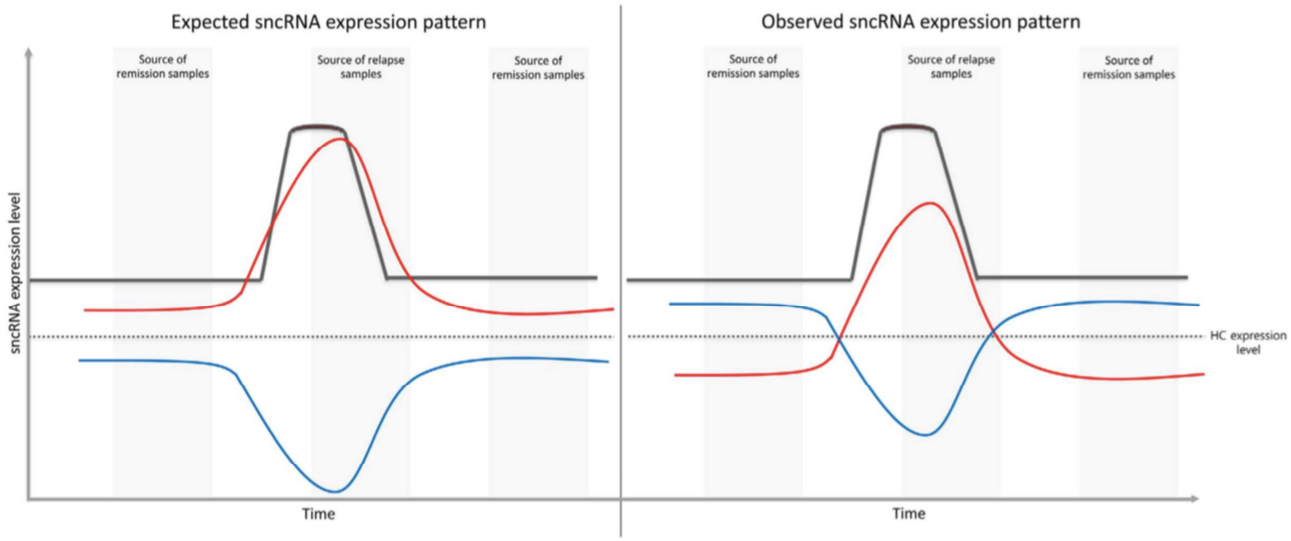


Figure 3. "The mirror pattern" effect where the expected (left) and the observed (right) patterns of regulations are shown, representing upregulated (red) and downregulated (blue) sncRNA expression compared to HC expression level.

Looking, then, for the implication of snoRNAs in MS, we turn back to genetic studies. The key paper, in this aspect, is a GWAS published in Nature in August 2011 (Sawcer *et al.* 2011). The International Multiple Sclerosis Genetics Consortium and The Wellcome Trust Case Control Consortium 2 performed this GWAS. They used 9772 MS patients and 17376 HCs; worldwide populations with

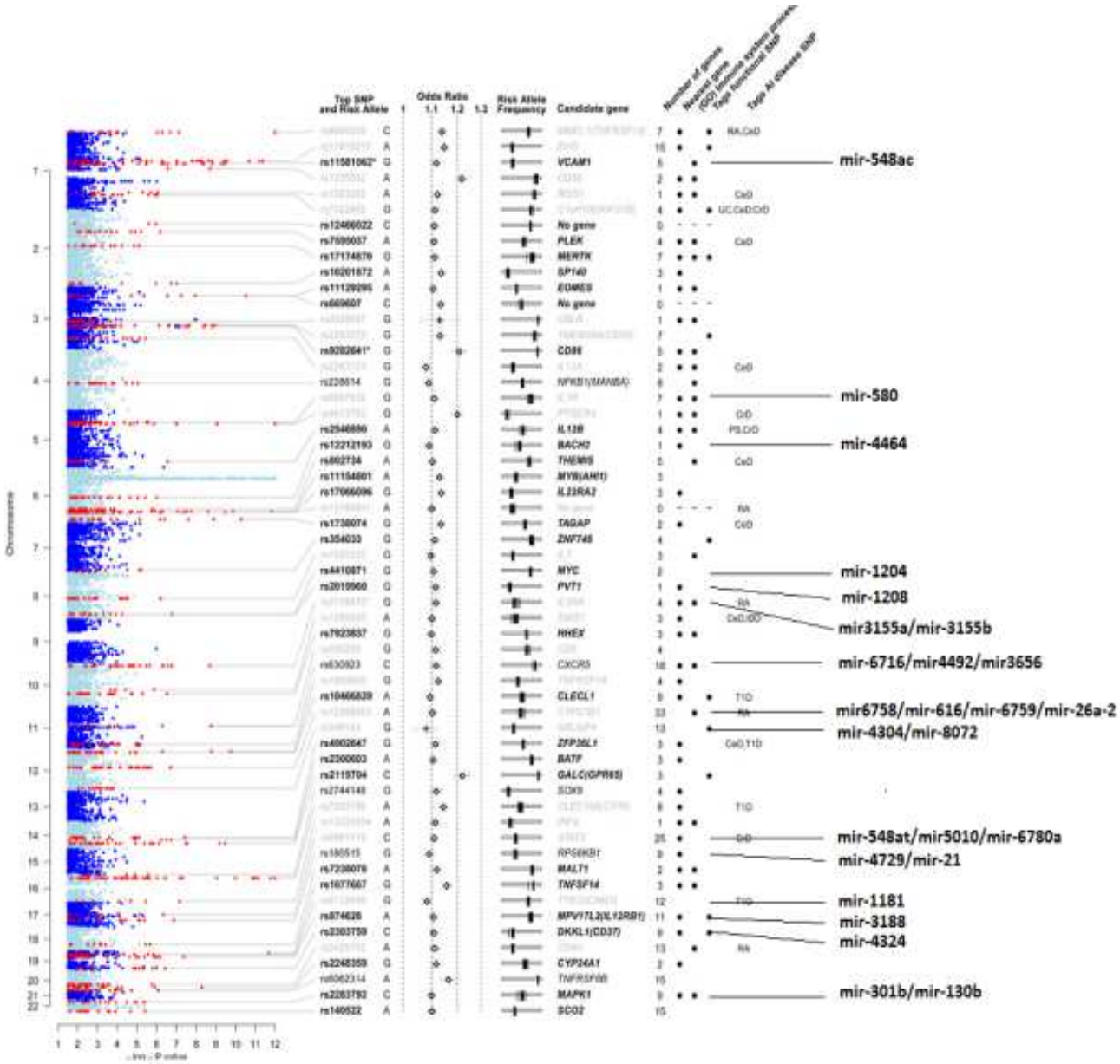
European ancestry, and they looked for 465434 autosomal SNPs. This study, identified outside the MHC region of the genome, 95 regions having at least one SNP associated with MS. Apart from the already known 26 loci, the GWAS discovered 29 novel loci associated to MS and 5 regions with strong evidence for association. Besides, over one third of the identified loci overlapped with regions already confirmed as associated with other autoimmune diseases. However, this work could not find evidence for associating genetics with clinical course, severity, month of birth or gender but could confirm the previously suggested association of the age at onset and the DRB1*15:01 allele, meaning that individual genetic susceptibility is inversely correlated with age at onset. Although the GWAS could not find which components within the nervous system are initially damaged, the overrepresentation of genes involved in the T-helper-cell differentiation pathway indicate the critical disease mechanisms involve, in the first place, immune dysregulation, as in the 30% of the association regions the nearest gene to the lead SNP is an immune system gene. Therefore, this paper provided valuable information as to regions associated to MS that could further be analyzed. It brought the idea that this genetic variants associated to genes are individually interesting, but collectively their small effects add up to contribute to genetic susceptibility for MS (Sawcer *et al.* 2011).

However, deep sequencing studies had to be performed in order to detect those exact nucleotide changes that led to the increasing of the susceptibility. In this sense, Otaegui *et al.* started by sequencing 55 candidate protein-coding genes by Next Generation Sequencing (NGS) Ion Torrent described on the GWAS 2011 in 10 Gipuzkoan families (where one relative had MS and the other was a HC). The results showed no meaningful relationship between those changes in protein-coding genes and MS, at least in the population they studied (Osorio-Querejeta *et al.* 2015).

The next step was to start looking at the non protein-coding genes, as abovementioned, they are being discovered more and more important. Hence, Otaegui *et al*. sought for miRNAs in the 57 regions of the genome related to the MS, according to the GWAS 2011. That study concluded with 26 miRNA genes (Fig. 4) that were further analyzed in 7 Gipuzkoan families with some siblings of them affected by MS. The results obtained, pointed out two regions in three of the miRNAs (miR3155a, miR3155b and miR3656) that had numerous single nucleotide polymorphisms (SNPs) that could possibly be related to the disease (Osorio-Querejeta *et al.* 2015).

Another important work to mention here is Maialen de la Cuesta's final work of the Bachelor's degree (a.k.a TFG/GrAL). Maialen's job was to demonstrate whether these SNPs are susceptible to MS by being significantly overexpressed in MS patients. To do so, she checked in a population of 285 MS patients and 285 HCs. She amplified each of the three miRNAs by PCR and also made a restriction fragment length polymorphism (RFLP) so that she could genotype the miRNAs. After doing statistical analysis (Chi-square and Fisher's exact test), she determined that the differences between the

genotypic frequencies of the MS patients and HCs, were not statistically significant and so, she concluded in saying that the SNPs found in the three miRNAs could not be associated to having MS



and therefore, rejected her hypothesis (De la Cuesta 2015).

Figure 4. Modified diagram from the GWAS 2011 showing the 26 miRNAs that are present in the regions to be associated to MS (Osorio-Querejeta *et al.* 2015).

After not having found anything statistically meaningful as to gene-coding proteins and miRNAs in the regions selected by the GWAS from Nature 2011, to complete the picture of ncRNA in MS, the group decided to keep on investigating the snoRNAs.

HYPOTHESIS AND OBJECTIVES

This work has two main hypotheses:

1. A genetic variant or some genetic variants in the DNA sequence of SNORA40 make it to overexpress in MS patients (and especially in females) and not in HCs.
2. Genetic variants in the snoRNAs located in previously MS-associated genome loci, may be related to the disease in a Mendelian MS-form way.

To test those hypothesis two main scientific objectives are proposed:

1. Analyze SNORA40 in the 47 subjects where RNA expression was studied (Muñoz-Culla *et al.* 2016) in order to determine genetic variants that could explain that change in the RNA expression between disease statuses.
2. Characterize SNPs within different snoRNAs related to the association regions of the GWAS 2011 and see if some of these polymorphisms are more linked to the disease in 7 Gipuzkoan families' samples.

Besides the previously stated scientific objectives, this work in all its extends, wants, amongst other, to promote the critical thinking much needed in science, understanding and familiarization with different molecular biology techniques, such as, PCR, sequencing or culturing of cells and acquire the necessary skills to work in a molecular and biomedical laboratory as a researcher.

MATERIAL AND METHODS

Selection of samples

For the SNORA40 study, we used DNA blood samples from the 47 previously analyzed samples in the paper (Muñoz-Culla *et al.* 2016) and for the rest of the snoRNAs, we obtained DNA blood samples from 14 family duos (consisting of a MS sibling and a HC sibling) belonging to 7 Gipuzkoan (Basque) families. All the samples were acquired from the Basque Biobank (www.biobancovasco.org).

Selection of snoRNAs

For the reasons abovementioned, we selected SNORA40 in order to be further investigated and the rest of snoRNAs were chosen based in the GWAS of 2011. We downloaded a list of all human snoRNAs and their location in each chromosome based on Affymetrix GeneChip miRNA array 4.0 annotation. Overlapping that to the 57 association regions defined in the 2011 GWAS with GenomicRanges package in R 3.2.2 using RStudio v0.99.486, a list of snoRNAs and their Ensembl code (www.ensembl.org) was obtained.

Primers' design

Knowing the Enseml code of each snoRNA we looked up for the common names of each of the snoRNAs and we entered them in UCSC (www.genome.ucsc.edu) in order to obtain the sequence. After that, we designed the primers using Primer3 software online (www.bioinfo.ut.ee) and verify back in UCSC (using an *in silico* PCR tool and the assembly to the Feb. 2009 version (GRCh37/hg19)) whether the primers were valid and there was not any SNP described in the hybridation zone.

Once having the primers designed, we purchased them from Integrated DNA technologies (IDT oligos) and make the appropriate dilutions for each primer in order to have each one of them in a concentration of 100 μM.

PCR tuning

In order to find the optimum annealing temperature for each PCR reaction, different PCR tunings were made in Veriti 96-well Thermal Cyclers (Applied Biosystems). For the tuning, two DNA samples (in 100 ng/μL concentration) were taken from healthy individuals and a negative control (H_2O).

Taking as reference the melting temperature (Tm) of the primers given by Primer3 and IDT oligos, we proposed different programs for each snoRNA and several tries were made before obtaining the adequate annealing temperature. The most used programs were different Touch Down programs (Td) that fall down the temperature 0,5 ºC in each cycle during 10 or 20 cycles, depending on the program; allowing to guess the annealing temperature range. However, in some cases, fixed programs were also used, especially in the cases where we had obtained faint bands and we wanted to sharpen the quality of the amplification. The PCR conditions varied between snoRNAs and between different programs within the same snoRNA, and in some cases, dimethyl sulfoxide (DMSO) was added (an organic solvent that minimizes the self-complementarity structures of the DNA, known as unspecific bands). The PCR conditions of those PCR tunings that were afterwards used for the subjects' PCRs are summarized in the "results and discussion" section (Table 4).

After doing the PCR reactions, a 35 minutes electrophoresis was made in 2% agarose gel in order to migrate de samples according to their size and charge. To do so, an electric current of 200 mV was applied and the molecules would migrate to the positive pole. Yet previously, when making the gel a DNA intercalant agent was used (Ethidium bromide) so that the different bands created in the electrophoresis were visible after the migration in the ultraviolet (UV) picture. To load the molecules easily in the gel, a dying (blue bromophenol) was employed to make the molecules visible. The gel

was inserted in a buffer tray composed of mainly water (H_2O) and Tris/Borate/EDTA (TBE) in order the electrical conductance to be high. The bands obtained were compared to the bands from an already known molecular size marker, called ϕX174 RF DNA/Hae III Fragments.

PCR reactions

For the snoRNAs that we managed to tune a PCR program, we took the experimental DNA samples wanting to be amplified in each case and we made real PCR reactions.

In one hand, samples for all the snoRNAs excluding SNORA40 were 14 DNA samples from 7 Gipuzkoan families (two samples from each family, one from a healthy brother and one from a MS brother, the proband)(Table 1) and two controls (one of the DNAs used in the tuning of the PCRs, as a positive control, and H_2O as a negative control). All of them were in a concentration of 100 ng/μL.

Table 1. General information of the 14 subjects from 7 Gipuzkoan families whose DNA was obtained for the snoRNAs amplification, visualized as the subject's reference code, sex (male: M or female: F), the disease status (healthy control: HC or multiple sclerosis: MS) and the family number each of the siblings belongs to.

Code	Sex	Disease status	Family
1	M	MS	1
2	M	HC	1
3	F	MS	2
4	F	HC	2
5	F	MS	4
6	F	HC	4
7	F	MS	3
8	F	HC	3
9	F	MS	8
10	F	HC	8
11	F	MS	5
12	F	HC	5
13	F	MS	7
14	F	HC	7

On the other hand, samples for SNORA40 were DNA samples from 47 subjects (10 MS male, 14 MS female, 7 HC male and 15 HC female) and two controls (one of the DNAs used in the tuning of the PCRs, as a positive control and H_2O as a negative control). All of them were in a concentration of 100 ng/μL.

The programs used for amplifying each of the snoRNAs (Table 3) and the PCR conditions for different snoRNAs (Table 4) are summarized in in the "results and discussion" section.

After doing the PCR reactions in Veriti 96-well Thermal Cyclers (Applied Biosystems), a 35 minutes electrophoresis was made in 2% agarose gel with 200 mV with previously dying the samples in blue bromophenol, as abovementioned.

Purification of the PCR product

Once obtaining the results of the PCR reactions, those ones wanting to be sequenced had to be purified with ExoSAP-IT. ExoSAP-IT utilizes two hydrolytic enzymes, Exonuclease I and Shrimp Alkaline Phosphatase, to remove the unwanted dNTPs and primers remaining in the PCR product mixture that unless removing them would interfere with the sequencing reaction. The Exonuclease I degrades residual single-stranded primers and any extraneous single-stranded DNA produced by the PCR. The Shrimp Alkaline Phosphatase hydrolyzes remaining dNTPs from the PCR mixture.

We added 2 μL of ExoSAP-IT for each 5 μL PCR product and incubated at 37ºC for 15 minutes in order to degrade the unwanted primers and dNTPs. Then, we did another incubation at 80ºC for 15 minutes for inactivating the ExoSAP-IT.

Sequencing

After the last incubation time, the samples ready to be sequenced, we sent them to sequencing in an ABI3130 automatic sequencer (Applied Biosystems) using Bigdye v2.1. The further analysis of the results was carried out with SeqScape 2.5.0 and genotyping results were extracted.

Characterization of SNPs

For assuring whether the genetic variants found in the sequences had already been described or were novel genetic variants, we downloaded from UCSC a table with all the SNPs in our PCR product region based on NCBI's dbSNP build 146 (ftp.ncbi.nih.gov/snp) and compare it to our own genetic variants. This dbSNP database contains information of SNPs and small insertions and deletions (indels).

For each SNP two subgroups were made dividing MS and HC and we counted for each group the number of individuals having each of the different possible genotypes. Then, a percentage for each subgroup's possible genotypes was calculated.

RESULTS AND DISCUSSION

Technical aspects prior to sequencing

The overlapping of the miRNA and snoRNA genes in the microarray to the regions selected on the GWAS 2011, detected 10 snoRNAs that were seek in the databases in order to obtain the necessary information to amplify them. The names, location and the most relevant information obtained prior to

the PCR tunings, can be summarized in table 2 and 3. Those tables also include an eleventh snoRNA, which is the already mentioned SNORA40. It is necessary to mention that in the case of U68 the primers chosen on the first place did not work out, thus, needing to redesign new primers after some tries. For that reason, both pair of primers can be seen on table 2.

Table 2. Each of the snoRNAs and their primers' melting temperature (Tm), guanine and cytosine proportion (%GC), the sequence (5'→3') and the length in base pairs (bp) of each primer, and the amplicon's location in the genome (chromosome number and position in base pairs) using UCSC's Feb. 2009 version (GRCh37/hg19).

snoRNA	Tm (F/R)	GC% (F/R)	Primers: Sequence (5'→3') & length (bp) (F/R)	Genome location
U8	54,6	50,0/50,0	CCTGGCCTGATAGTTTCCAA (20) CAACCAGCCCTATCAGGAAA (20)	chr18:56485839-56486438
ACA51	56,6	50,0/47,8	GCAAGACCCTGCCAAAACAA (20) TCCAGATATGAGGAGGGAGTTCA (23)	chr1:93311539-93312127
U23	57,8/57,2	60,0/54,5	CACCACCTCGTCCTCCAAAG (20) GCAAGGATAGTTACTGACCGGG (22)	chr2:232320163-232320761
SNORA14	56,7/57,9	52,3/60,0	CCACAACTCCTGGGCTATCTT (21) GACCCAGGGCAGCATCTATG (20)	chr10:6058821-6059300
SNORD112	55,8/55,4	45,4/43,3	TGTGCCAATGTACCCTAGAACT (22) TCAAAGGGCAAGCTTCTTCTTAC (23)	chr3:121964449-121964847
SCARNA16	56,6/56,1	47,6/50,0	ACGGGGAGTTGTTTAATGGGT (21) GCAAAAGGAGCCTGACCAAA (20)	chr1:101598574-101599042
U66	54,6/56,0	40,9/50,0	ACCAAGTACTGTTTGCTTTCCT (22) GGTGCAAAGAATGGGCAGAT (20)	chr1:93306077-93306752
U68	56,7/57,4	55,0/57,1	TCAAGAGTTCGAGACCAGCC (20) AGCTCTCTCTAGGATCCCACC (21)	chr5:158656739-158657588
	59,9/59,9	45,0/52,6	TCAAAGGCCCCCTTAAAAGT (20) GGAAAGTGCCTTGTGAGCA (19)	chr5:158657014-158657513
U21	55,3/53,9	50,0/45,0	GTGGTGGAAAGCCTTGGTAA (20) AACCAGGGAATCGTTTGGTA (20)	chr1:93302765-93303032
SNORA63	57,7/57,1	55,0/52,3	CCAGAGCCAGGAAAGCACAT (20) TCCAGTTCTCAAGGGGAATGC (21)	chr5:40654816-40655371
SNORA40	58,5/59,3	50,0/47,6	ATGGCACCTGGCCTAATACT (20) CCCCAGTACCCAGCTAATTTT (21)	chr11:93468080-93468579

PCR tunings were held during almost two months, and different tries had to be made in order to find the most accurate annealing temperatures (Table 3). The most common causes for rejecting a program and trying different temperatures were either that the bands in the electrophoresis would not come out or that not a unique band was obtained (non-specific bands).

Table 3. Each of the snoRNAs, the different programs tried before obtaining the desired annealing temperature and the final tuning temperature (the one with the most accurate annealing temperature).

snoRNA	Tries	Tuning temperature
U8	Td 57-52 Td 60-50	Td 60-55
ACA51	Td 57-52 Td 60-55	Td 60-50
U23	Td 60-55 Td 65-55	Td 61-56
SNORA14	Td 60-55 Td 61-56	Td 65-55
SNORD112	Td 57-52 Td 60-50	Td 60-55
SCARNA16	Td 57-52 Td 60-55 Td 60-50 58	60
U66	Td 60-55 Td 60-50 Td 57-52	55
U68	Td 57-52 Td 60-55 Td 60-50 60 59	Td 60-50
U21	Td 57-52 Td 60-55 Td 60-50 61	62
SNORA63	Td 57-52 Td 60-55 Td 60-50 Td 65-55 Td 65-60 60 59	Td 65-60
SNORA40	–	Td 60-55

As to the PCR mixes, several changes in the mix were held. We started the PCR tunings with a mix of a total volume of 25 μL and used it with SNORA40, ACA51, SNORD112 and U8 (being successful in all except in SNORD112 and despite being successful in ACA51 we preferred to refine it with fixed temperature programs). However, we decided that as the process of trying optimal PCRs seemed quite long and therefore it could turn out to be expensive, it was better to reduce the final volume of the mix to 20 and so, we recalculated the mix volumes out of a final volume of 20 μL. With this new PCR mix, we did the tuning of U23, ACA51, SNORD112, U21, SNORA63 and U68 (being successful in all except in SNORA63, SNORD112 and U68). At this point, as we had several snoRNAs already in

tune, we started amplifying them with the subjects' DNA. After that, we detected a mistake in the PCR mix we had been using; the final volume was 21,6 μL, instead of the expected 20 μL (it was diluted), as the primers' volume added was wrong. Nonetheless, we continued in tuning the rest of the snoRNAs' PCRs, thus, we recalculated the mix with the adequate volume of the primers. Later on, we changed another time the mix; not the final volume, but we added DMSO (1 μL) to avoid unspecific bands in the electrophoresis, accordingly needing to reduce the water volume (1 μL). We used this DMSO-mix with U68, however, as not being successful, afterwards another PCR mix was used for U68. The rest of the snoRNAs, was tried out with different PCR mixes until achieving the ideal annealing temperatures. The final PCR mixes used in the end, at the time of amplifying the snoRNAs with the subject's DNA, are listed in table 4.

Table 4. Volumes (V) in μL added for a single PCR reaction for each of the snoRNAs from the total volume indicated in each case, out of the stock concentrations mentioned in the reactives (column 1).

Reactives	**SNORA40:** V for a total of 25 μL	**U8, ACA51, U23, SNORA14 & SNORD112:** V for a total of 21,6 μL	**U68:** V for a total of 20 μL	**SCARNA16:** V for a total of 20 μL	**U66:** V for a total of 20 μL	**SNORA63 & U21:** V for a total of 20 μL
NH_4 Reaction Buffer (Bioline) (10X)	2,5	2	2	2	2	2
$MgCl_2$ (50 mM)	1	0,8	0,8	1	1	0,8
dNTPs (5 mM)	1	0,8	0,8	2	2	2
Primers F/R (5 mM)	2/2	1,6/1,6	1,6/1,6	1,6/1,6	1/1	1,6/1,6
Enzyme (BioTaq)(Bioline) (5 U/μL)	0,3	0,24	0,24	0,2	0,3	0,2
DNA (100 ng/μL)	1	1	1	1	1	1
H_2O	15,2	13,56	11,96	10,6	10,7	9,8
DMSO (Sigma-Aldrich)	-	-	-	-	1	1

After doing the PCRs with the subjects' DNA, in the posterior electrophoresis, we wanted to obtain same size bands in all the wells except in the well of the negative control (the water)(Fig. 5). Pictures similar to figure 5 were accepted and in the cases where some of the wells did not have strong bands,

we repeated them. Take into account that these amplified PCR products are required for the further sequencing of the snoRNAs.

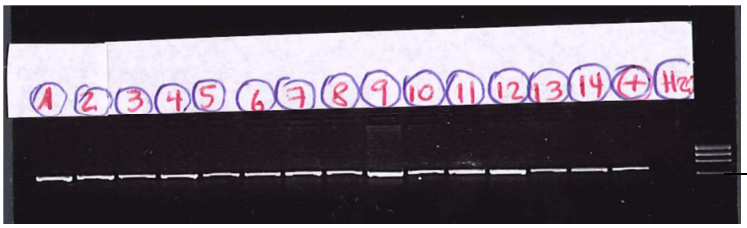


Figure 5. Picture of the gel electrophoresis of the PCR product of snoRNA U66 amplified with the set of primers indicated in table 2 and with the DNA of the Gipuzkoan families in a fixed temperature program of 55 (table 3). The numbers account for the codes listed in table 1. Plus sign indicates the positive control (previously amplified DNA) and H_2O the negative control, water. The last column is the marker: ϕX174 RF DNA/Hae III Fragments.

Sequencing of selected snoRNAs

The sequencing results of the snoRNAs showed in most of the cases good quality sequences that could be analyzed. We checked if the nucleotides that the software selected as being different to the reference sequence were really different to it or were sequences with a high background noise. Many sequences were not as clean as expected and the reported variations in the nucleotide sequence by the software were really not SNPs. However, some of them had unquestionably different nucleotides to the reference sequence, and were either heterozygous or homozygous (Fig. 6).

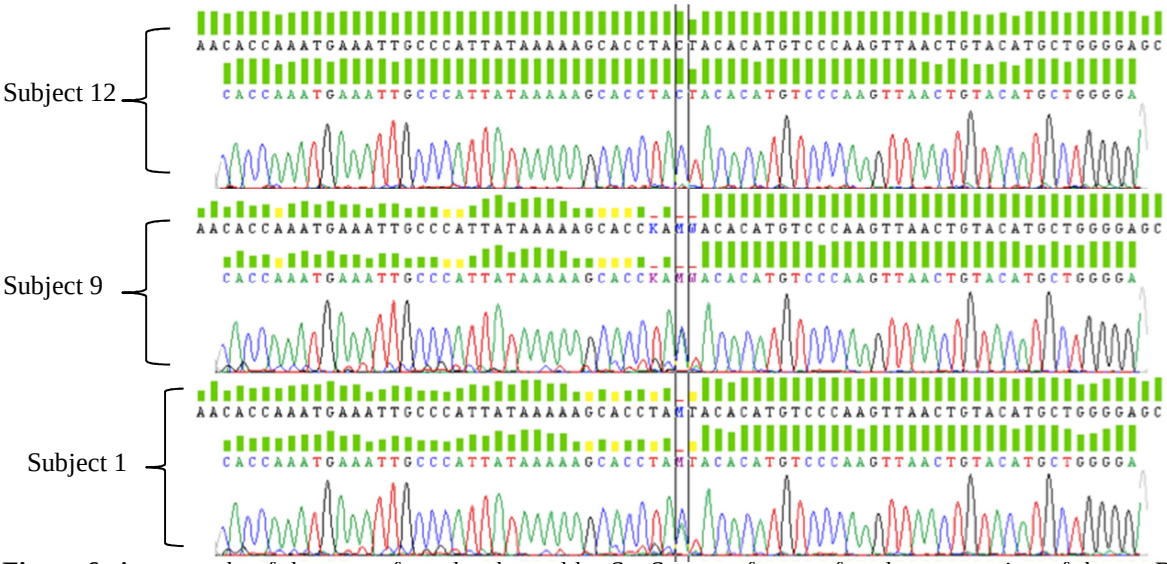


Figure 6. An example of the type of results showed by SeqScape software after the sequencing of the snoRNA U23. The 3 electropherograms represent the DNA sequence of the subjects with the code number 12, 9 and 1 and the two strands for each subject indicate the forward sequence (the first being a resume sequence). Colors on the electropherogram indicate each of the different nucleotides and colors on the strands indicate the intensity of each peak. The column in the middle of the image indicates the position 215 bp. We can distinguish that for the selected nucleotide number 12 is homozygous for the C allele (CC) and number 9 and 1 are heterozygous (M=AC based on the IUPAC codes).

Characterization of SNPs

In those cases where the sequences had background noise, we resequenced the snoRNAs with the reverse primer (and so, the other strand's sequence was obtained). We checked whether the genetic variants in our sequences were already described in the databases. The majority of the sequences did not have any SNP different to the consensus sequence. That is the case of SNORA40 where there cannot be seen any of the expected genetic variation, meaning the DNA configuration is not the responsible of the described RNA expression change in our patients. However, some of them had SNPs, and in that case, some were already described and some were not (table 5).

Table 5. The number of SNPs characterized in each of the snoRNAs' PCR product based on NCBI's dbSNP database. Column 2 indicated the number of SNPs described in dbSNP's "all SNPs" mode. Columns 3 and 4 indicate the SNPs detected in the sequences from the 47 subjects (for SNORA40) and 14 subjects listed in table 1 (for the rest of snoRNAs). Column 3 indicates the SNPs from our subjects that match with the already described SNPs in the database and the column 4 indicates the SNPs that are not described in the database but can be found in our subjects.

snoRNA's region PCR product	**SNPs in dbSNP**	**SNPs found in our samples**	
		Described SNPs	**New SNPs**
U8	27	-	-
ACA51	33	3 rs12750269 rs11164825 rs11164826	3
U23	119	1 rs13019380	1
SNORA14	14	-	-
SNORD112	22	1 rs77254509	2
SCARNA16	27	1 rs780500159	-
U66	82	1 rs10874744	-
U68	36	-	-
U21	33	-	-
SNORA63	41	-	-
SNORA40	75	-	-

The SNPs already described in dbSNP are listed with the percentage of each possible genotype indicated for MS subjects and HCs (Table 6). As previously mentioned, the variants detected in the dbSNP database can be either SNPs or small indels. In our case, all of them are SNPs except for a variant from SCARNA16 (rs780500159) that is a small indel with the reference allele being the

sequence AAAC and in our case all the samples having the deletion. Subject 12 in the snoRNA U23 was not included when calculating the percentage because it could not be sequenced till the end due to the sequence being truncated. Therefore, the percentage was calculated out of 6 individuals.

Table 6. The already described SNPs table, indicating for each SNP already described in dbSNP database detected in the snoRNAs ACA51, U23, SNORD112, SCARNA16 and U66 (written with their rs... name) and for MS and HC subjects, the percentage of each genotype. All the percentages where calculated out of 7 (being that the number of subjects in each MS HC group) except for the group of HCs of U23 where one subject was not sequenced, and so the percentage was calculated out of 6.

Genotypes	ACA51						U23		SNORD112		SCARNA16		U66	
	rs12750269		rs11164825		rs11164826		rs13019380		rs77254509		rs780500159		rs10874744	
	MS	HC	MS	HC	MS	HC	MS	HC	MS	HC	MS	HC	MS	HC
TT	28.57	28.57	14.28	28.57			71,42	50	14,28	-				
CC	14.28	28.57	28.57	28.57										
CT	57.14	42,85	57.14	42.85										
GG					57,14	42.85	-	16.66	42,85	42.85			57,14	42.85
AG					42,85	42.85							14,28	28.57
AA					-	14.28							28,57	28.57
GT							28,57	33.33	42,85	57.14				
AAAC														
-AAAC											100	100		

The genetic variants that were not found at dbSNP database are listed (Table 7) taken as the name for the SNP the position from the beginning of the amplified PCR product (PCR product region in table 2). For each SNP there can be seen percentage of the possible genotypes divided in MS subjects and HCs.

Table 7. New SNPs table, indicating the new SNPs detected in the snoRNAs ACA51, U23 and SNORD112 (the SNP position is written in base pairs (bp) from the beginning of the PCR product region). The numbers inside the table mean the percentage of each genotype for each SNP differentiating MS and HC.

Genotypes	ACA51						U23		SNORD112			
	SNP200		SNP201		SNP204		SNP215		SNP248		SNP266	
	MS	HC	MS	HC	MS	HC	MS	HC	MS	HC	MS	HC
AA	57.14	57.14			42.85	57.14						
AG	42.85	42.85			57.14	42.85			100	100	100	100
GG			57.14	57.14								
GT			42.85	42.85								
CC							-	71.42				
AC							100	28.57				

These SNPs show different alleles that appear with different allele frequencies compared to the general population stated in the database information (dbSNP), probably, because of the small sample size but also, due to the Basque genetic background (showing a high degree of genetic isolation)(Massó *et al.* 2015). In no case was detected a clear relationship between a determinate allele and having the illness or not. The only case where there could be seen a relationship between having a certain variant and having the disease, was in the SNP 215 bp of U23 where except for the families 8 and 3, all the rest are heterozygous (AC) for the disease and homozygous (CC) for the controls (Table

8), or in other words, 100 % of MS are AC and 72,42 % of HC are CC (Table 7). This new SNP, as it seems a prominent candidate SNP, it should be further analyzed in other MS-affected families or other non-Basque population in order to relate it to MS. The SNPs included for the SNORD112 are doubtful, as they are different to the reference sequence for all the subjects (heterozygous for all the subjects and the reference sequence is homozygous)(Table 8). This sequences were included here because they did not seem to have much background noise, but being heterozygous for all the subjects probably means the sequence was not as clean as it should be.

Table 8. U23 SNP 215 table, indicating each subject's code (more details in table 1), disease status, family that belongs to (each pair of siblings marked with same color) and the new SNP 215 (the SNP position is written in base pairs (bp) from the beginning of the PCR product region) detected in U23. Ref. means the genotype found in the reference sequence. Except for the families 3 and 8 all the rest are heterozygous for MS (AC) and homozygous for HC (CC).

Code	**Disease status**	**Family**	**U23**	**SNORD112**	
			SNP 215	SNP 248	SNP 266
Ref.			CC	GG	GG
1	MS	1	AC	AG	AG
2	HC	1	CC	AG	AG
3	MS	2	AC	AG	AG
4	HC	2	CC	AG	AG
5	MS	4	AC	AG	AG
6	HC	4	CC	AG	AG
7	MS	3	AC	AG	AG
8	HC	3	AC	AG	AG
9	MS	8	AC	AG	AG
10	HC	8	AC	AG	AG
11	MS	5	AC	AG	AG
12	HC	5	CC	AG	AG
13	MS	7	AC	AG	AG
14	HC	7	CC	AG	AG

Polymorphisms and MS relationship

We start this project with the hypothesis of a Mendelian inheritance in our families. Such a Mendelian inheritance has been described in other neurodegenerative diseases, such as Parkinson's disease, where few cases (around 10 %) can be familial and thus, can present a Mendelian pattern (i.e a mutation that segregates with the cases)(Gasser 2009; Hernandez *et al.* 2016). Although no Mendelian inheritance is expected in MS as disease, the hypothesis of such a Mendelian form of MS that would be based in pathogenic mutations, has been a way to explain the heterogeneity of the heritability of the disease, therefore, trying to solve the "missing heritability" that is estimated to be around the 70%

(Lill 2014; Hoppenbrouwers and Hintzen 2010; Wang *et al.* 2016). That "missing heritability" is thought to be, amongst other causes, because multitude of genetic variants may never exceed the genome-wide significance threshold in association studies due to small effect sizes but also, owing to the presence of rare variants (with minor allele frequency, MAF, <0.5%). These rare variants, although being infrequent in the overall population and thus not detecting by GWAS, they wield a much larger effect than common variants, especially in multiplex MS families (Lill 2014; Hoppenbrouwers and Hintzen 2010). However, some authors also highlight the little evidence present regarding this kind of a Mendelian form of MS (Hauser and Oksenberg 2006; The International Multiple Sclerosis Genetics Consortium (IMSGC) 2010). That being said, we should not reject this type of heredity, at least in some families, as there is more and more evidence supporting such a heritability way. For instance, a novel pathogenic mutation (NR1H3 p.Arg415Gln missense substitution) that encodes liver X receptor alpha (LXRA) was identified recently in two unrelated MS-affected families. This subjects' clinical phenotype was an unusual RRMS that soon after the onset became PPMS (Wang *et al.* 2016). Such discoveries, remind us that all the multipopulational analyses made so far, do not explain all the genetics of the illness, and that from time to time, this sort of smaller analysis can discover otherwise undetectable genes that work in a mendelian way.

With this in mind, we checked the sequence of candidate genes (GWAS 2011) related with MS in a proband and his healthy brother in 7 MS families. Due to the negative results (always under a Mendelian hypothesis where mutation should be present in proband and not in the healthy brother) we checked the miRNA genes inside the associated regions with no results (after validation in another dataset). These results drove us to the study of snoRNAs under the same methodology and the same hypothesis. No changes in these snoRNAs have been found in the proband and his healthy brother. Our results, then, show that no relationship can be established between the disease and snoRNA in our families. We also demonstrate that the overexpression of SNORA40 found in PBMCs of some of the 47 subjects is not due to changes in the DNA configuration as no SNPs are detected in the sequences. Nevertheless, further research should be made in snoRNAs and other type of ncRNAs, like long ncRNAs; now that the role of ncRNAs is so prominent, pursuing the non-coding investigation line of this group.

We must highlight that although our sample sizes could seem insufficient, we are studying in one hand the relationship between expression and polymorphism (therefore we study SNORA40 in the 47 samples in which the RNA expression were studied), and on the other hand, we are checking, under a Mendelian hypothesis the snoRNAs in the proband and the healthy brother in 7 families (n=14). These sample sizes could be sufficient to find a promising change to be validated in more samples after the segregation studies in the own family, as done by Wang *et al.* 2016 on a similar approach.

In any case, different points of view should be tackled and each of the distinct research methods for investigating MS shed light for creating the final image of the disease. Linkage studies, GWAS, expression networks, epigenetic analysis and sequencing methods, amongst others, provide diverse information that putting all together, will definitely help us understand better the entrails of multiple sclerosis.

CONCLUSIONS

1) Despite SNORA40 having 75 already described SNPs throughout the entire amplified region, we did not detect any SNP in our samples, meaning that the change in the RNA expression found in some of the 47 subjects would not be due to changes in the DNA sequence, and suggesting other extra-DNA changes, such as, epigenetics or environmental factors, that should be investigated.

2) The snoRNAs located in the association regions selected by the GWAS 2011, do not seem to be responsible for causing MS in a Mendelian way, at least in the 7 Gipuzkoan families studied. This cannot rule out their role as adding probability of having MS, in a typical polygenic way, which we could not detect in this study. Besides, this study does not either discard that this snoRNAs or some of them might act in a mendelian way in other families.

REFERENCES

Comabella M., Khoury S. J. (2012) Immunopathogenesis of multiple sclerosis. *Clin. Immunol.* **142**, 2–8.

Compston A., Coles A. (2008) Multiple sclerosis. *Lancet* **372**, 1502–1517.

Ender C., Krek A., Friedländer M. R., Beitzinger M., Weinmann L., Chen W., Pfeffer S., Rajewsky N., Meister G. (2008) A Human snoRNA with MicroRNA-Like Functions. *Mol. Cell* **32**, 519–528.

Esteller M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.* **12**, 861–74.

Gasser T. (2009) Mendelian forms of Parkinson's disease. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1792**, 587–596.

Hafler D. A., Compston A., Sawcer S., Lander E. S., Daly M. J., Jager P. L. De, Bakker P. I. W. de, et al. (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.* **357**, 851–62.

Hauser S. L., Chan J. R., Oksenberg J. R. (2013) Multiple sclerosis: Prospects and promise. *Ann. Neurol.* **74**, 317–327.

Hauser S. L., Oksenberg J. R. (2006) The Neurobiology of Multiple Sclerosis: Genes, Inflammation, and Neurodegeneration. *Neuron* **52**, 61–76.

Hernandez D. G., Reed X., Singleton A. B. (2016) Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. *J. Neurochem.* **41**, 1–16.

Hoppenbrouwers I. A., Hintzen R. Q. (2010) Genetics of multiple sclerosis. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1812**, 194–201.

Irizar H., Muñoz-Culla M., Sáenz-Cuesta M., Osorio-Querejeta I., Sepúlveda L., Castillo-Triviño T., Prada A., Lopez de Munain A., Olascoaga J., Otaegui D. (2015) Identification of ncRNAs as potential therapeutic targets in multiple sclerosis through differential ncRNA - mRNA network analysis. *BMC Genomics* **16**, 250.

Lill C. M. (2014) Recent advances and future challenges in the genetics of multiple sclerosis. *Front. Neurol.* **5 JUL**, 1–4.

Massó J. F. M., Zarranz J. J., Otaegui D., Munain A. L. De (2015) Neurogenetic Disorders in the Basque Population. *Ann. Hum. Genet.* **79**, 57–75.

Maurano M. T., Humbert R., Rynes E., Thurman R. E., Haugen E., Wang H., Reynolds A. P., et al. (2012) Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science (80-.).* **337**, 1190–1195.

Muñoz-Culla M., Irizar H., Castillo-Triviño T., Sáenz-Cuesta M., Sepúlveda L., Lopetegi I., Munain a L. de, Olascoaga J., Baranzini S., Otaegui D. (2014) Blood miRNA expression pattern is a possible risk marker for natalizumab-associated progressive multifocal leukoencephalopathy in multiple sclerosis patients. *Mult. Scler.*, 4–13.

Muñoz-Culla M., Irizar H., Otaegui D. (2013) The genetics of multiple sclerosis: review of current and emerging candidates. *Appl. Clin. Genet.* **6**, 63–73.

Muñoz-Culla M., Irizar H., Sáenz-Cuesta M., Castillo-Triviño T., Osorio-Querejeta I., Sepúlveda L., López de Munain A., Olascoaga J., Otaegui D. (2016) SncRNA (microRNA &snoRNA) opposite expression pattern found in multiple sclerosis relapse and remission is sex dependent. *Sci. Rep.* **6**, 20126.

Osorio-Querejeta I., Muñoz-Culla M., Alberich M., la Cuesta M. De, López de Munain A., Castillo-Triviño T., Olascoaga J., Otaegui D. (2015) Resequencing of the genes proposed to be associated to multiple sclerosis in the GWAS 2011 (Nature) in Basque- MS families and the microRNAs inside these loci. *Mult. Scler. J.* **21**, 654–779.

Ricaño-Ponce I., Zhernakova D. V., Deelen P., Luo O., Li X., Isaacs A., Karjalainen J., et al. (2016) Refined mapping of autoimmune disease associated genetic variants with gene expression suggests an important role for non-coding RNAs. *J. Autoimmun.* **68**, 62–74.

Sawcer S., Hellenthal G., Pirinen M., Spencer C. C., Patsopoulos N. A., Moutsianas L., Dilthey A., et al. (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219.

Scott M. S., Ono M. (2011) From snoRNA to miRNA: Dual function regulatory non-coding RNAs. *Biochimie* **93**, 1987–1992.

Stys P. K., Zamponi G. W., Minnen J. van, Geurts J. J. G. (2012) Will the real multiple sclerosis please stand up? *Nat. Rev. Neurosci.* **13**, 507–514.

The International Multiple Sclerosis Genetics Consortium (IMSGC) (2010) Evidence for Polygenic Susceptibility to Multiple Sclerosis-The Shape of Things to Come. *Am. J. Hum. Genet.* **86**, 621–625.

Tullman M. J. (2013) Overview of the Epidemiology, Diagnosis, and Disease Progression Associated With Multiple Sclerosis. *Am. J. Manag. Care.* **19**, 15–20.

Wang Z., Sadovnick A. D., Traboulsee A. L., Ross J. P., Bernales C. Q., Encarnacion M., Yee I. M., et al. (2016) Nuclear Receptor NR1H3 in Familial Multiple Sclerosis. *Neuron* **90**, 948–954.

Warnke C., Menge T., Hartung H.-P., Racke M. K., Cravens P. D., Bennett J. L., Frohman E. M., et al. (2010) Natalizumab and progressive multifocal leukoencephalopathy: what are the causal factors and can it be avoided? *Arch. Neurol.* **67**, 923–30.