UNIVERSITÀ DEGLI STUDI DI MILANO

Corso di Dottorato: in Biologia Molecolare e Cellulare

Dipartimento di Bioscienze

Modulation of LINE-1 retrotransposition by Aicardi-Goutières syndrome-related genes.

BMC

Candidato:
Valentina Menetti

Tutor:
Federico Lazzaro

Coordinatore del dottorato:
Martin Kater

Anno accademico 2019

Index

Abstract

La sindrome di Aicardi-Goutières (AGS) è una rara malattia infantile di origine genetica, i pazienti AGS sono caratterizzati da: atrofia cerebrale, calcificazioni intracraniche ed elevati livelli di interferone-alfa nel liquido cerebro-spinale. Questa inappropriata produzione di IFNα può diventare deleteria, promuovendo l'innesco di una risposta autoinfiammatoria nel paziente. AGS è causata da mutazioni in alcuni geni che codificano per enzimi che metabolizzano o rilevano la presenza di acidi nucleici: TREX1, ognuna delle tre subunità (RNASIH2A, RNASIH2B, e RNASIH2C) del complesso enzimatico RNasi H2, SAMHD1, ADAR1 e MDA5. Molteplici evidenze sperimentali suggeriscono che gli acidi nucleici siano responsabili dell'attivazione di sensori dell'immunità innata. Pertanto, l'ipotesi più accreditata è che un accumulo di acidi nucleici endogeni non processati possa essere la principale causa della patologia, inducendo gli elevati livelli di IFNα nel liquido cerebro-spinale. Tuttavia, quale sia la natura e il meccanismo di accumulo di questi acidi nucleici resta ancora ignoto. Tra tutti i geni responsabili della sindrome di Aicardi-Goutières, è stato dimostrato che TREX1, ADAR1 e SAMHD1 possono interferire nel metabolismo dei retroelementi LINE1, sequenze di DNA che possono muoversi nel genoma tramite un intermedio a RNA. Un'ipotesi plausibile è quindi che sia proprio l'accumulo di alcuni intermedi di retrotrasposizione una delle cause scatenanti della iperattivazione dell'immunità innata. Caratterizzare un possibile coinvolgimento del metabolismo dei retroelementi e in particolare di LINE1, nella sindrome di Aicardi-Goutières è l'obbiettivo a lungo termine di questo progetto.

Abstract

Aicardi-Goutières syndrome (AGS) is a rare genetically heterogeneous disease that typically affects newborns and infants and AGS patients are characterized by cerebral atrophy, intracranial calcifications and elevated levels of IFNα in the cerebrospinal fluid (CSF). Such inappropriate activation of type I IFN can be detrimental to the host by promoting autoinflammatory responses. AGS is caused by mutations in several genes encoding nucleic acids sensor or metabolizing enzymes: the 3′ exonuclease 1 (TREX1), any of the three subunits (RNASEH2A, RNASEH2B, and RNASEH2C) of the ribonuclease H2 (RNase H2) enzyme complex, the triphosphohydrolase encoded by SAMHD1, the adenosine deaminase acting on RNA 1 (ADAR1), or the RNA sensor melanoma differentiation associated protein 5 (MDA5). Strong evidence suggest that the accumulation of nucleic acids is responsible for triggering sensors of the innate immunity. Therefore, it is generally believed that an accumulation of unprocessed, endogenous nucleic acids could be the main pathogenic trigger of AGS, increased CSF IFNα levels. Among all the AGS-causing genes, TREX1, ADAR1 and SAMHD1 could affect the correct metabolism of LINE1 retroelements, DNA sequences able to move in the genome through an RNA intermediate. Then, a conceivable cause of innate immunity hyperactivation could be retrotransposition intermediates accumulation. To characterize a possible involvement of retroelements metabolism in the AGS pathogenesis is the long term goal of this project.

Aim of the thesis

RNase H2 is a hetero-trimeric enzyme and represents the primary source of ribonuclease H activity in mammalian cells, the non-sequence-specific endonuclease cleavage of RNA in an RNA:DNA substrate. Mutations in the three RNase H2 subunits account for over 50% of Aicardi–Goutières syndrome (AGS) patients and the molecular mechanisms linking RNase H2 mutations to AGS pathology have not been elucidated yet. Identify the cellular roles of RNase H2 that can be related to the AGS pathogenesis, and their characterization is the main objective of this project.

Among all the AGS-causing genes, TREX1, ADAR1, and SAMHD1 could affect the correct metabolism of LINE1 retroelements, the most active autonomous transposable elements in humans. Considering RNase H2 enzymatic activities, it can potentially regulate the mobility of endogenous L1 retroelements by degrading RNA:DNA hybrids formed during their replication. On the basis of this hypothesis, I started to investigate the effects of RNase H2 on L1 retrotransposition using a genetic assay in human cells silenced for one of the three RNase H2 subunits. I set up these assays to verify if RNase H2 inhibits the mobility of endogenous retroelements.

During the first two years of my PhD, other researcher groups published contrasting results concerning RNase H2 and L1 mobility. Therefore, my work assumes most importance in order to clarify controversial RNase H2 roles responsible for AGS pathogenesis. In particular, Choi and colleagues obtained similar results respect to mine in terms of RNase H2 impact on L1 retrotransposition, and they concluded that RNase H2 inhibits L1 mobility[29]. On the other hand, in Benitez-Guijarro et al. has been demonstrated that in the absence of RNase H2, L1 mobility decreases. Based on their results, they concluded that RNase H2 is necessary for endogenous L1 mobility[30].

Since all these data are obtained using a genetic assay, which can be influenced by many factors, we decided to continue this characterization, starting to look at endogenous L1 regulation. It would be really interesting to repeat these kinds of L1 endogenous evaluations considering the other AGS-causing genes, and I began to extend this analysis to two known L1 retrotransposition inhibitors, ADAR1 and TREX1. To demonstrate that the dysregulation of L1 mobility is a common role of AGS related nucleic acid metabolizing enzymes, it became the long term objective of my PhD thesis. Furthermore, we are also working on the demonstration that L1 intermediates accumulation in AGS patient cells is responsible for the overproduction of IFNα and at the end for the onset of AGS.

Introduction

Aicardi-Goutières syndrome

Aicardi–Goutières syndrome (AGS) was first described in 1984 by two French pediatric neurologists Jean Aicardi and Françoise Goutières. They get into a case of a child of consanguineous parents, suffering from calcification of the brain basal ganglia and severe encephalopathy. That infant also had a chronic cerebrospinal fluid (CSF) lymphocytosis, and for that reason, the case was initially misdiagnosed as intrauterine viral infection, despite the serological assays failed to detect any trace of viruses in the patient[1].

However, when a female sibling and then a brother of the child was affected by the same condition shortly after their birth, Aicardi and Goutières discarded the infection diagnosis and proposed an unidentified genetic disorder instead[1,2]. This drove Aicardi and Goutières to expand the set by searching for other patients who presented congenital viral infection-like symptoms, and in a short space of time, they moved from the recognition of the first AGS-causing gene (AGS1) in 2007, to the discovery of four genes responsible for the pathogenesis. AGS1 was the DNA exonuclease TREX1 and AGS2-4 were genes encoding ribonuclease H2 subunits, RNaseH2B (AGS2), RNaseH2C (AGS3) and RNaseH2A (AGS4). A deeper analysis of AGS cases showed that in over 80% of families, there were biallelic mutations in one of these four genes. In contrast, for the remaining 20%, no mutations were found, so at least one further AGS-causing gene must be discovered.

Type I IFNs play central roles in innate immunity, but overproduction of IFNα can lead to immunopathologies. IFNα in the CSF during gestation or soon after birth has been considered to be the most informative marker for the diagnosis of AGS[2]. Prenatal diagnosis of AGS, measuring IFNα of fetal blood or cerebrospinal fluid, has not been accomplished, considering that this procedure would be risky and speculative. Indeed, IFNα levels in cerebrospinal fluid of AGS patients were high at birth and declined during the following years[3]. More recently, an increased level of expression of IFN-stimulated genes (ISGs) in peripheral blood, a so called IFN signature, has been reported to be present at any age in almost all AGS patients[1]. Patients with RNaseH2B mutations are an exception because approximately 30% of them had no such upregulation of ISGs[1].

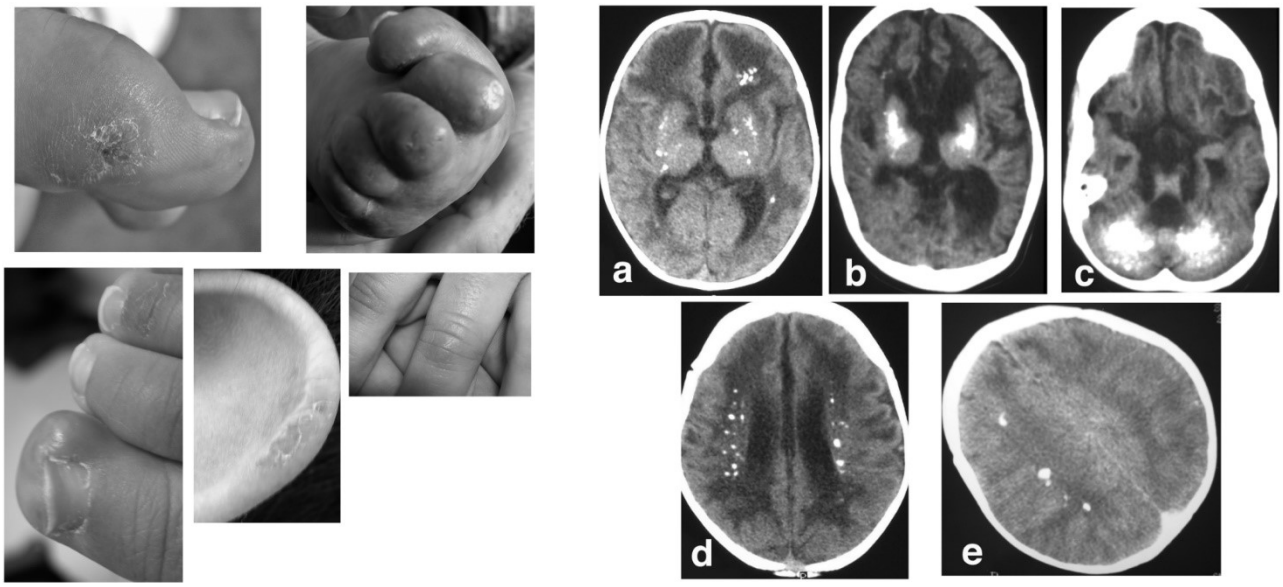


Figure 1: Chilblain lesions and intracranial calcifications of AGS patients[1].

The incidence of AGS is very low, with approximately 500 affected families known worldwide, onset occurs before 3 and 7 months of age in most patients, and the death rate is estimated to be around 30%[4]. Although AGS is rare, its clinical importance is exaggerated by the high chance that its symptoms and signs may be mistaken for a non-genetic congenital infection[3].

Radiologically, the disease was characterized by basal ganglia calcification, white matter abnormalities, and brain atrophy. These features, together with the later development of chilblain skin lesions, constitute 'classical' AGS phenotypes[1,5](Figure1).

There are two main forms of AGS, the neonatal and the later onset ones. The neonatal form is typically due to TREX1 mutations, and neurological abnormalities are manifest at birth or in the first following days. These patients present features that can suggest congenital viral infection but with negative serologic assay results. In the later onset form, instead, a normal development period precedes a sub-acute regression with extreme irritability and sterile pyrexias, followed by loss of skills and slowing circumference growth. This last described form of AGS is associated with a prolonged course, low mortality, and the gene most implicated is RNaseH2B[2].

Since 2007 when Yanick Crow and his colleagues have identified the first AGS causing gene TREX1, due to the advent of the new sequencing technologies, they characterize mutation in seven different AGS-related genes: TREX1, RNaseH2A, RNaseH2B, RNaseH2C, SAMHD1, ADAR and IFIH1(MDA5)[2,4] (Table 1). These seven genes encode proteins that are involved in nucleic acid metabolism/signalling. This observation, together with the increased interferon levels in CSF, defines AGS as an autoimmune disorder associated with the induction of a type I interferon response driven by improper endogenous nucleic acid accumulation[6].

While there is significant phenotypic overlap in patients with these seven mutations, inflammation in certain tissues appears to be associated with mutations in some AGS genes but not others, and the molecular mechanisms behind each mutation and symptom remain a matter of study[4].

Gene name (alternative name)	Inheritance	Human phenotypes	Protein function
TREX1	Autosomal recessive or autosomal dominant	AGS, FCL, SLE and RVCL	3′–5′ DNA exonuclease
RNASEH2A	Autosomal recessive	AGS	A catalytic component of the RNase H2 complex that acts on the RNA portion of RNA–DNA hybrids and removes ribonucleotides embedded in DNA
RNASEH2B	Autosomal recessive	AGS and spastic paraparesis	A non-catalytic component of the RNase H2 complex
RNASEH2C	Autosomal recessive	AGS	A non-catalytic component of the RNase H2 complex
SAMHD1	Autosomal recessive	AGS, FCL and CLL	dNTP triphosphohydrolase triphosphatase and ribonuclease activity
ADAR (DRADA)	Autosomal recessive or autosomal dominant	AGS, DSH, BSN and spastic paraparesis, as well as CNP	Hydrolytic deamination of adenosine to inosine in dsRNA
IFIH1 (MDA5)	Autosomal dominant	Various neuroimmunological and non-neurological phenotypes, including AGS, spastic paraparesis, CNP and SMS	Cytosolic sensor of dsRNA

Table 1: Summary of AGS causing genes and their functions[2].

Intracellular nucleic acids detection and type I IFNs response

A fundamental function of the immune system is to distinguish self from non-self and initiate a specific response against only the latter. During the past decade, there has been rapid progress in understanding how the innate immune system accomplishes this self and non-self discrimination. The discovery of germline-encoded pattern-recognition receptors (PRRs) in the 1990s showed how innate immunity recognizes danger signals emanating from pathogen-associated molecular patterns (PAMPs). Importantly, these PAMPs must be essential to the microbes so they cannot readily mutate them to avoid PRRs detection.

Whereas bacteria and fungi possess microbe-specific structures for immune recognition that are absent in the host, viruses present a unique challenge because every component of a virion is made from the infected host cell. Most antiviral responses are initiated by nucleic acids innate immune receptors. On the one hand, viruses cannot replicate without their DNA or RNA genome, but on the other hand, DNA and RNA are among the most abundant macromolecules in all of our cells. Therefore, it has become increasingly evident that these same sensors that protect us from a viral infection can also drive human autoimmune diseases when endogenous nucleic acids inappropriately activate them, and Aicardi-Goutières syndrome is an example of that. Indeed, AGS is an autoimmune disorder associated with the induction of type I IFNs response driven by improper endogenous nucleic acid accumulation in the cytoplasm of patient cells[4].

While some PRR transmembrane receptors, named Toll-like receptors, detect nucleic acids within endosomes of specialized cells such as dendritic cells or B cells (TLR3, TLR7, TLR8, and TLR9), a distinct and complementary set of nucleic acid-sensing receptors reside in the cytosol of virtually all cells. This last category of cytosolic nucleic acid sensors plays a central role in the initiation of antiviral immunity, and also in AGS pathogenesis.

The cytosolic RNA sensors retinoic acid–inducible gene 1 (RIG-I), and melanoma differentiation-associated protein 5 (MDA5) helicases, known as RIG-I-like receptors (RLRs) detect structural features of viral RNA, that are scarce within host RNAs, and activate an antiviral response that includes the inducible production of type I IFNs. In particular, RIG-I binds to short double-stranded RNAs that contain 5' tri or di-phosphates and long, double-stranded RNAs activate MDA5. The binding of virus-derived RNAs to RIG-I or MDA5 leads to a conformational change that facilitates the exposure of their active N-terminal domains and the consequent association with the mitochondrial adaptor protein MAVS[7].

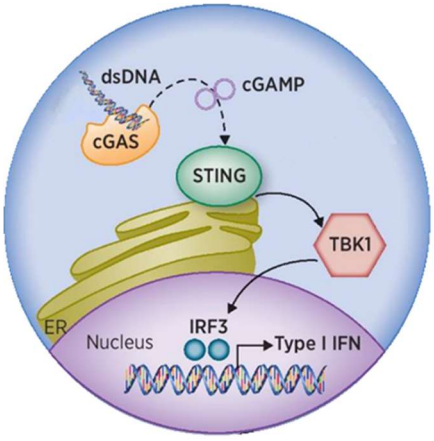


Figure 2: cGAS-STING pathway activation.

The principal intracellular DNA sensor is the nucleotidyl transferase cGAS, which upon ligand binding, catalyzes the synthesis of the cyclic GMP-AMP dinucleotide (cGAMP) using ATP and GTP as substrates. This second messenger molecule cGAMP binds to and activates an endoplasmic reticulum protein, the adapter molecule stimulator of interferon genes (STING). The activation of the adaptor protein MAVS, for what concerns RNA sensors, or STING, in case of DNA sensing, leads to their re-localization and association with TANK binding kinase 1 (TBK1) and causes the recruitment of the interferon regulatory factor 3 (IRF3) and nuclear factor-kB (NF-kB). After translocation into the nucleus, IRF3 induces the transcription of type I IFNs and ISGs, while NF-kB induces the expression of pro-inflammatory cytokines[7,8] (Figure 2). Innate immune cells such as macrophages and dendritic cells, produce IFNα, whereas non-immune cells, such as fibroblasts or epithelial cells, predominantly produce IFNβ.

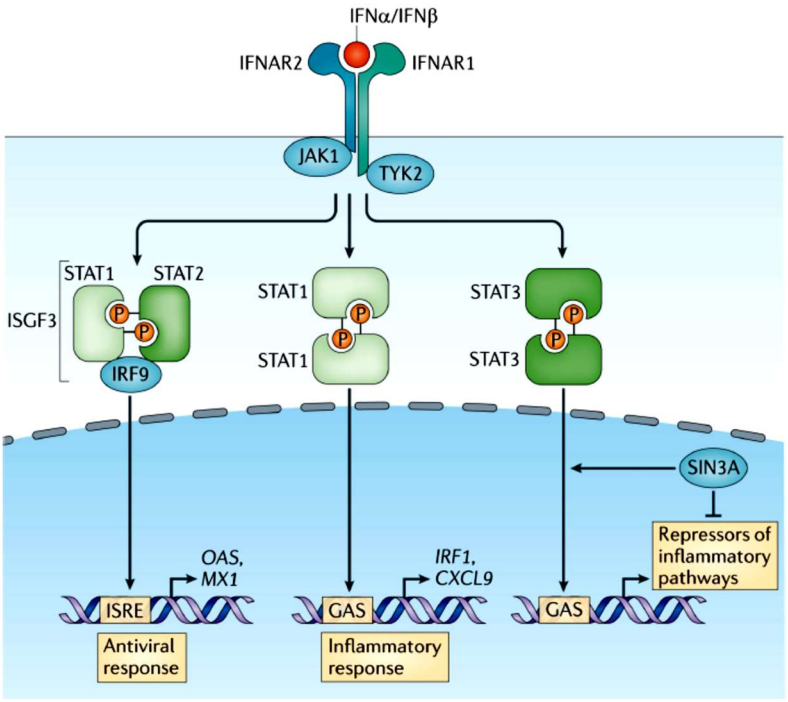


Figure 3: The canonical type I IFN signalling pathway[9].

Secreted IFNα and IFNβ bind the transmembrane IFNα receptors (IFNARs), composed of two subunits IFNAR1 and IFNAR2, present on near cells. In the canonical type I IFN-induced signalling, IFNAR activates and is phosphorylated by Janus kinase 1(JAK1) and tyrosine kinase 2 (TYK2). Phosphorylation of IFNAR by these kinases results in the recruitment and activation of signal transducer and activator of transcription 1 and 2 (STAT1 and STAT2). Phosphorylated STAT1 and STAT2 dimerize and translocate to the nucleus where they can control different gene expression programmes.

In particular, there are three main STAT complexes formed in response to type I IFNs:

- IFN-stimulated gene factor 3(ISGF3) complex, built by STAT1, STAT2 and IFN-regulatory factor 9 (IRF9) that binds to IFN-stimulated response elements (ISREs; consensus sequence TTTCNNTTTC) to activate classical antiviral genes;
- STAT1 homodimers bind to gamma-activated sequences (GASs; consensus sequence TTCNNNGAA) to induce pro-inflammatory genes expression;
- STAT3 homodimers bind to GAS sequences and indirectly suppress pro-inflammatory genes expression by the induction of unknown transcriptional repressors (Figure 3).

Innate immune cells respond to type I IFNs by enhancing antigen presentation and the production of cytokine and chemokines. Type I IFNs also activate the adaptive immune system, promoting maturation and proliferation of lymphocytes[9]. Conversely, in case of an inappropriate antiviral immunity activation, a persistent type I IFNs exposure results in dendritic cell activation and loss of B and T cell tolerance leading to autoantibody production. The targets of these autoantibodies are ubiquitous self-antigens and antigen-antibody complexes deposition in the capillary bed, followed by local leucocyte activation that causes destructive tissue inflammation. This pathologic condition, characterized by a constitutive production of IFNs and an inflammatory response against self-tissues, is known as interferonopathy[9,10] (Figure 4).

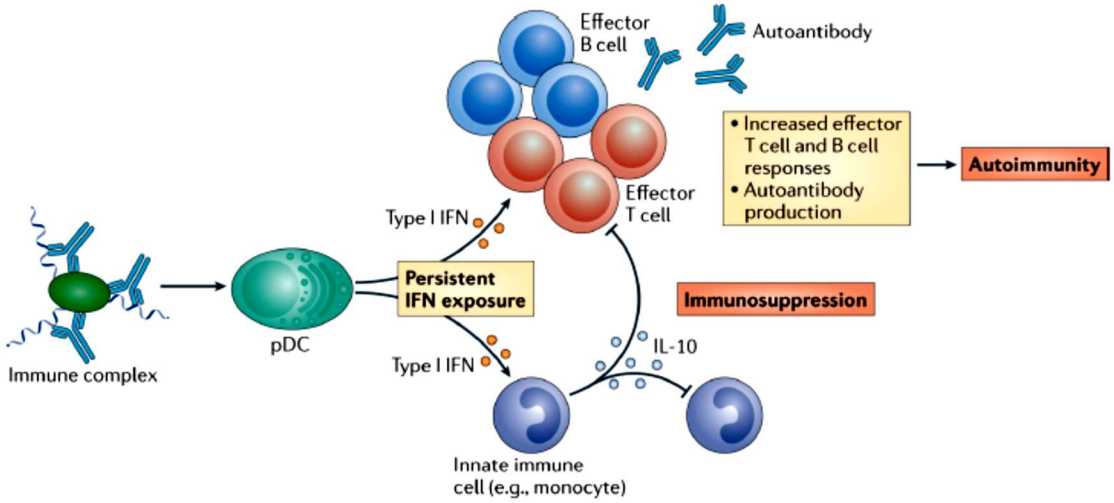


Figure 4: Chronic IFN production in autoimmune diseases[9].

The type I interferonopathies can be categorized into four main classes. The first class, represented by AGS, is caused by mutations in genes encoding enzymes that regulate intracellular nucleic acid accumulation that activate either cGAS-STING or RLR-MAVS pathways. The second class is induced by mutations that provoke an enhanced sensitivity or ligand-independent activation of cGAS-STING or RLR-MAVS pathways. The third is represented by systemic lupus erythematosus (SLE)-like diseases caused by the mutation in extracellular nucleases that activate TLRs. Finally, the last class is due to defects in pathways that modulate type I IFN responses independent of nucleic acid sensing[4,7]. Type I IFN activation induced by immune recognition of self nucleic acids is a key event in the pathogenesis of type I interferonopathies, excluding the fourth class. For that reason, the identification of the endogenous nucleic acid species involved in the cGAS-STING or RLR-MAVS activation, in the case of AGS, is of primary relevance. To characterize the nucleic acid species involved in AGS pathogenesis, the starting point is the study of the heterogeneous genetic basis of this syndrome.

The genetic basis of AGS

As mentioned before, AGS is a genetically heterogeneous disease resulting from mutations in any one of the genes encoding: the 3′ exonuclease 1 (TREX1), any of the three subunits (RNASEH2A, RNASEH2B, and RNASEH2C) of the ribonuclease H2 (RNase H2) enzyme complex, the tri-phosphohydrolase encoded by SAMHD1, the adenosine deaminase acting on RNA 1 (ADAR1), or the RNA sensor melanoma differentiation associated protein 5 (MDA5; encoded by IFIH1)[11] (Figure 5,6). All these genes are deeply characterized, but the link with the unset of pathologies remains unclear.

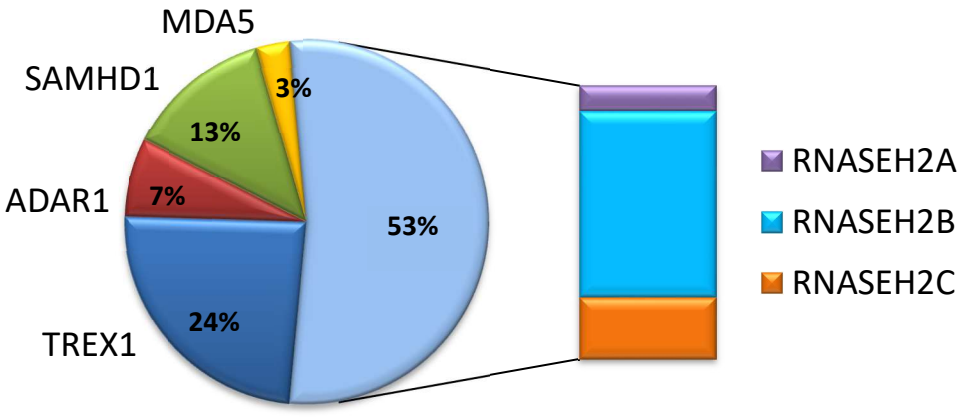


Figure 5: AGS-causing genes percentages.

TREX1

The DNA nucleases are involved in DNA replication, repair, and recombination and are essential to maintain genome stability. These enzymes can have endo or exonuclease activities, a selective affinity for single or double-stranded DNA (ssDNA or dsDNA) and hydrolyse in a 5'-3' or 3'-5' direction producing 5' mono or di-nucleotides and 3' mononucleotides. The 3' repair exonuclease I (TREX1) is the most abundant DNA 3'-5' exonuclease in mammalian cells[12]. TREX1 is a widely expressed homo-dimeric protein with no orthologues in lower eukaryotes, that preferentially degrades ssDNA. TREX1 shares homology to proof-reading DNA exonucleases in bacteria, but unlike the proof-reading exonucleases, TREX1 is anchored by its C terminus to the cytosolic face of the ER membrane. Loss-of-function mutations of TREX1 in humans cause autoimmune diseases, including AGS, familial chilblain lupus (FCL), systemic lupus erythematosus (SLE) and retinal vasculopathy with cerebral leukodystrophy (RVCL)[13]. Similar to AGS, TREX1 ko mice have an elevated type I ISG and develop high levels of inflammation in multiple tissues, leading to significant mortality. Interestingly, TREX1 ko mice develop inflammatory myocarditis and lack detectable inflammation in the brain. Whereas, in humans, AGS affects the brain, but only rarely the heart and that difference in target tissues between TREX1 deficiency in humans and mice remain unclear[4].

To investigate the mechanism underlying the autoimmune diseases caused by TREX1 deficiency, Stetson and colleagues generated TREX1 ko mice lacking IRF3, IFNAR, or RAG2, a DNA recombinase required for the generation of functional lymphocytes[14]. In each case, the mortality and cardiac inflammation observed for just TREX1 ko mice were nearly completely rescued. These results provide evidence that the activation of the IRF3 pathway is responsible for the autoimmune diseases caused by TREX1 loss of function mutations, at least in mice[15,14]. In the last years, they farther demonstrated that the accumulated endogenous DNAs in TREX1 null mice activate the cGAS-STING pathway[16].

Given that TREX1 is a DNA nuclease, a reasonable hypothesis is that cells lacking this enzyme accumulate DNA that activates the cGAS-STING pathway. The studies of TREX1 ko mice showed an accumulation of ssDNA in the endoplasmic reticulum in MEFs and of cytosolic DNA in mice hearts respect to wild type once. Although the nucleic acid sensors involved in TREX1 autoimmune diseases have been identified, it remains to clarify the source of these accumulated endogenous DNA molecules[15,14]. There are at least three main sources that are not mutually exclusive.

The first possibility is those TREX1 substrates derived from retroelements. Supporting this possibility, in mice, overexpression of TREX1 inhibited the retrotranscription and integration of two synthetic retroelements, LTR and LINE1[4,15,14]. Moreover, it is observed abundant extrachromosomal DNA in human pluripotent stem cells lacking TREX1 respect to control cells, of which endogenous

LINE1 was a major source[17,18]. However, retroelement-derived DNA may not be the only substrate metabolized by TREX1; a second possibility is that this enzyme eliminates DNA by-products of DNA replication or repair. Indeed, Yang et al. found that TREX1 deficient MEFs had cell-cycle defects resulting from chronic activation of a DNA-damage checkpoint[4,19]. A third possibility is that TREX1 regulates inflammation independent of its DNA exonuclease activity. Supporting this hypothesis, in Peng Li et al., it is reported a nuclease-independent involvement of TREX1 in preventing LINE1 retrotransposition. In particular, TREX1 interacted with LINE1 ORF1 protein triggering its depletion and causing at the end, according to their model, reduced LINE1-mediated nicking of genomic DNA[4,20].

RNase H2

Ribonuclease H2 (RNase H2) belongs to the family of RNase H enzymes, that process the RNA moiety of RNA:DNA hybrid molecules. These hybrids are physiological intermediates produced during multiple cellular processes, such as retroviral infection, retroelement mobilization, during the synthesis of Okazaki fragments, and when a replication fork collides with the transcriptional machinery. Besides cleaving the RNA moiety of a plain RNA:DNA hybrid, RNase H enzymes participate in the removal of ribonucleotides embedded in a DNA duplex. Two classes of RNases H, with partially overlapping substrate specificity, have been characterized, RNase H1 and RNase H2[21]. Mammalian RNase H1 has two isoforms: a nuclear isoform of undefined function and a mitochondrial one that, probably due to its ability to hydrolyse RNA:DNA hybrids, is essential for mitochondrial DNA replication and maintenance[22].

RNase H2 is a trimeric enzyme conserved in all eukaryotes, and it's the major source of ribonuclease H activity in mammalian cells. This protein complex is composed of three subunits designated as A, B, and C. RNase H2A contains the catalytic center and forms a complex with RNase H2B and H2C subunits, which presumably fulfill accessory functions by serving as docking platforms for intracellular binding partners. Mutations in RNase H2 account for over 50% of AGS patients, and the molecular mechanisms linking RNase H2 mutations to AGS phenotypes have not been elucidated yet. RNase H2, besides being able to process long RNA:DNA hybrids, has the unique property of cleaving single ribononucleotide triphosphates (rNMPs) embedded in genomic DNA, as opposed to RNase H1 that requires at least four consecutive ribonucleotides embedded in a dsDNA sequence to cleave[21,22]. DNA polymerases can distinguish between rNTPs and dNTPs and select the latter during DNA replication. However, the fidelity of DNA polymerases is challenged by the high ratio of rNTPs respect to dNTPs that ranges from 10- to 100-fold in mammalian cells, and rNTPs are mis-incorporated into genomic DNA with high frequency during normal replication.

Stable incorporation of rNTPs in DNA it is usually avoided because it makes DNA prone to strand breakage and mutagenesis[22,23]. Recent studies from our and other groups have attributed a critical role to RNase H2 in preserving genome integrity by controlling genomic rNTPs incorporation during DNA replication[21,24,25,26]. Complete RNase H2 deficiency in mice is embryonic lethal at the gastrulation stage of development. This lethality is caused by defective removal of rNTPs from replicating genomic DNA, which results in massive genome instability and p53-dependent DNA damage response leading to apoptosis. The evidence that RNase H2 deficiency in mammals is not compatible with life, in contrast to what happens in a unicellular organism like yeast and bacteria, is consistent with the fact that only hypomorphic RNase H2 mutations have been reported in AGS patients[4,23,26].

Probably due to the early lethality, RNase H2 ko mice do not show a systemic type I interferon signature. Instead, two recently generated knock-in mice that express RNase H2A and H2B AGS mutations have an interferon signature, and this type I IFN production is cGAS and STING dependent. The nature of the immunostimulatory nucleic acids accumulated in RNase H2-deficient cells triggering cGAS activation, remains undefined. Some evidence suggests that cGAS-STING activation is related to the accumulation of DNA repair byproducts caused by genomic rNTPs incorporation[4,27,28]. Another possible source of immunostimulatory nucleic acids in RNase H2 deficiency condition can be endogenous retroelements intermediates. RNase H2 may regulate the mobility of endogenous retroelements by degrading RNA:DNA hybrids formed during their replication. Choi et al. and my data indicate that RNase H2 knockdown in different human cell lines seems to cause an increase in LINE1 retrotransposition efficiency[29]. Unexpectedly, another researcher group published opposite results by setting up retrotransposition assays on RNase H2A KO HeLa cellular clones[30]. Further work is needed to establish the exact RNase H2 involvement in the LINE1 retrotransposition process, and in general, to characterize the nature of accumulated nucleic acids that trigger cGAS-STING activation leading to IFNα production.

SAMHD1

Both prokaryotes and eukaryotes have developed defence mechanisms to protect their cells from viral invasions. In addition to innate and adaptative immune responses, an intrinsic antiviral system based on constitutively expressed intracellular proteins, known as restriction factors, are emerging. These restriction factors act during the first steps of virus-host interactions, and viral proteins often counteract them. Among known intrinsic restrictor factors, there are: APOBEC3 class of cytidine deaminases, components of nuclear domain structures, DNA repair proteins and viral capsid inhibitors.

Recently, the uncharacterized human sterile alpha motif (SAM) domain and histidine-aspartate (HD) domain-containing protein 1 (SAMHD1) has been shown to be a novel restriction factor that inhibits replication of HIV-1 genome in myeloid cells[31]. SAMHD1 is a deoxynucleoside triphosphate (dNTPs) hydrolase that depletes intracellular dNTP pools in non-cycling cells. It has been proposed that SAMHD1 inhibits HIV-1 infection by depleting the dNTP pools required for reverse transcription of the viral RNA genome[32,33,34].

17% of AGS cases is caused by biallelic loss-of-function mutations of SAMHD1. SAMHD1 ko mice show a chronic induction of type 1 IFNs in a cGAS and STING dependent manner. Like other AGS genes, the identity of the immunostimulatory nucleic acids accumulated in SAMHD1 depleted cells remains completely undefined[35,36].

Zhao et al. demonstrated that several domains of SAMHD1 are critical for its inhibition of LINE1 retrotransposition in dividing cells[37]. SAMHD1 ability to block LINE1 retrotransposition suggests that retroelement cDNA can be a source of cGAS ligands. Interestingly, it seems that SAMHD1 regulates retroviruses and retrotransposons through different mechanisms. In fact, SAMHD1 inhibits HIV infection in just non-dividing cells and inhibits LINE1 mobility in dividing cells. In G1, SAMHD1 is dephosphorylated by the PP2A-B55α phosphatase, it forms a tetramer and acts as a dNTP hydrolase to inhibit HIV infections. Upon entry in S phase, phosphorylation by Cyclin A-CDKs leads to the dissociation of SAMHD1 tetramers compromising its enzymatic activity and antiviral potency without interfering with LINE1 inhibition[38]. Furthermore, phosphorylated SAMHD1 is recruited to DNA repair foci in response to DNA damage, where it binds with high affinity RPA, fork structures, and MRE11, promoting an MRE11-dependent resection at stalled replication forks. ssDNA accumulated at stalled forks recruits the checkpoint kinase ATR, which in turn activates CHK1 to promote fork restart. In human SAMHD1 depleted cells, ssDNA fragments are displaced from stalled forks by RECQ1 DNA helicase, cleaved by MRE11, and accumulated in the cytosol. These replication stress byproducts can activate the cGAS-STING pathway to induce innate immunity activation[39,40].

ADAR1

Adenosine deaminases that act on RNA (ADARs) are RNA editing enzymes that target double-stranded regions of nuclear-encoded and viral RNAs and catalyse the deamination of adenosine to produce inosine. Hypoxanthine, the base of the nucleotide inosine generated by the editing process, is recognised as guanine by the translational and transcriptional machinery, so ADAR proteins change the primary sequence information in an RNA potentially leading to functional alterations of the affected genes. In addition, the I-U base pair is unstable compared to the parental A-U base pair, and

consequent changes in the secondary structure of RNA can result in its degradation. In general, ADARs have been shown to affect the splicing rate, the translational efficacy, or the stability of the edited mRNAs[41].

Among ADAR proteins, in 2012, AGS mutations in ADAR1 have been identified[42]. ADAR1 is widely expressed during embryonic and postnatal development in its constitutive and predominantly nuclear p110 isoform. An additional IFN inducible ADAR1 isoform, the p150 one, is found in both nucleus and cytoplasm of mammalian cells[43]. ADAR1 deficiency in mice results in hematopoietic failure and massive type I IFN signature before embryonic lethality. Several mechanisms have been proposed to explain these severe phenotypes of ADAR1 ko mice[44]. Pestal et al. demonstrate that ADAR1 is an essential negative regulator of the MDA5-MAVS RNA sensing pathway. In particular, they show that ADAR1 p150 isoform, the only one mutated in AGS patients, regulates the MDA5 pathway, whereas both the p150 and p110 isoforms contribute to development[45].

It remains to identify the RNA ligands of MDA5 that are edited by ADAR1. Approximately half of the mammalian genome is composed of retrotransposons, which typically form dsRNAs that are subjected to extensive A-to-I editing. Retrotransposons located in 3'UTR regions can be accumulated in the cytoplasm triggering MDA5 pathway activation. Based on these considerations, it is proposed that ADAR1 editing of retrotransposons dsRNA prevents MDA5 activation. Conversely, in ADAR1 deficiency conditions, retrotransposons can form long dsRNA stem-loops that can be recognised by MDA5[46].

Alu retroelements that are the most abundant and active retrotransposons, with more than 1 million copies in the human genome, have been proposed to be the primary targets of ADAR1[46].

In Orecchini et al., it is reported a possible ADAR1 LINE1 inhibitory function that is independent of its editing activity[47]. LINE1s can retrotranspose not only a copy of their RNA but also other RNAs, such as Alu elements[48]. Therefore, in AGS patients carrying ADAR1 mutations, non-edited Alu elements RNA widespread by LINE1 could trigger an MDA5 dependent IFN overproduction.

IFIH1/MDA5

The interferon-induced with helicase C domain 1 (IFIH1), also known as melanoma differentiation-associated protein 5 (MDA5), is a cytoplasmic viral RNA RIG-I-like receptor (RLR). Either RIG-I, and MDA5 RLR members are RNA helicases that bind to dsRNA; however, RIG-I seems to 'prefer' short dsRNA, whereas MDA5 can specifically bind long dsRNA. When bound to dsRNA, MDA5 forms a closed ring structure around the dsRNA stem leading to a conformational change that facilitates the exposure of its N-terminal domains that activate the adaptor protein MAVS located on the mitochondrial outer membrane.

Oligomerization of MAVS induces TBK1 activation, IRF3 phosphorylation, and initiation of signalling cascades that induce the expression of cytokines, including type I IFNs[7].

The previously described studies about TREX1, RNase H2 complex, SAMHD1, and ADAR1 AGS-related proteins suggest that an inappropriate accumulation of self-derived nucleic acids can induce type I IFNs production. The finding of AGS mutations in the IFIH1 gene implicates the aberrant sensing of nucleic acids as a cause of hyperactivation of innate immunity. AGS-causing mutations in the IFIH1 gene are the last identified and are the only known gain-of-function mutations in this disease; these mutants bind dsRNAs more avidly and tightly than wild-type. The precise nature of dsRNA species that can stimulate the mutant, but not wild-type MDA5 remains unknown[4,49]. Although the nature of such nucleic acid ligands is unknown, one may speculate that they could be derived from retroelements.

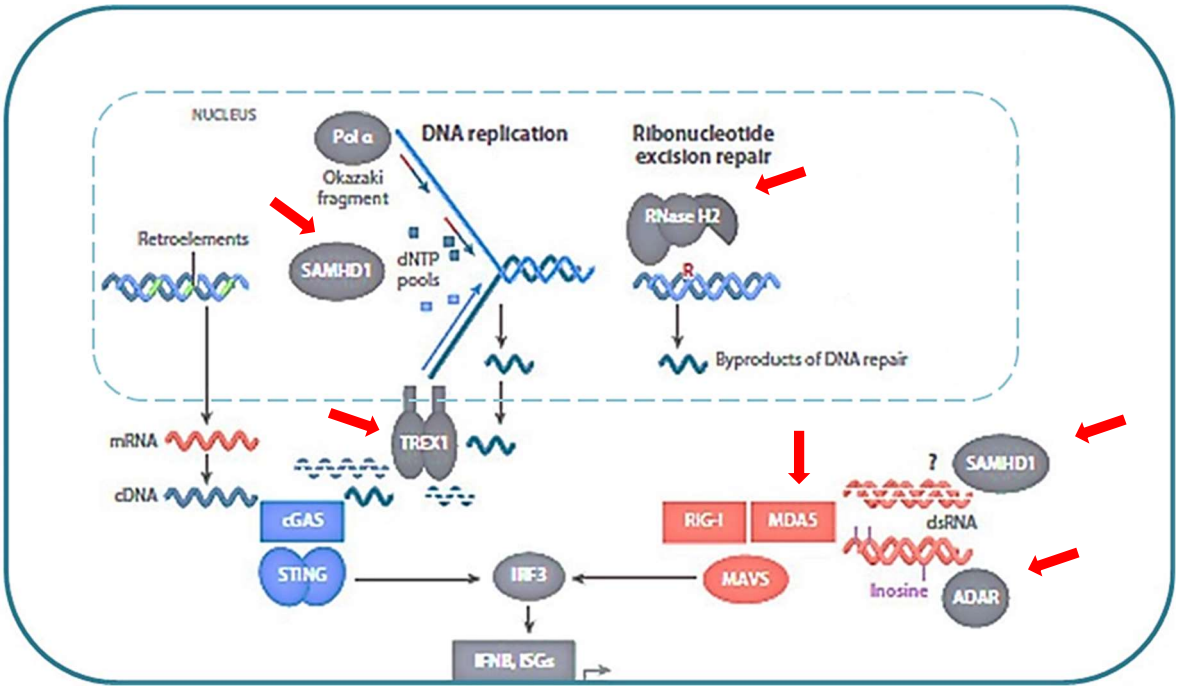


Figure 5: Cytosolic nucleic acids triggers IFNs production through cGAS-STING or RIG1-MDA5 pathways. Red arrows indicate AGS-causing genes. (Adapted from Yanick J. Crow and Nicolas Manel, 2015)

RNase T2, a putative AGS gene

The RNase T2 is the only member of the Rh/T2/S glycoprotein family in humans. This protein family of acid hydrolases is widely distributed in organisms from viruses to humans. RNase T2 is an ssRNA extracellular ribonuclease with a preferential cleavage of poly-A and poly-U of homo-polyribonucleotides. RNase T2 is involved in the pathogenesis of several human neoplasias such as ovarian cancer, melanoma, and non-Hodgkin lymphoma. In ovarian cancer models, the onco-suppressive role of this protein is associated with the recruitment of macrophages into the tumour mass[50].

Several RNase T2 loss-of-function mutations have been reported in infants affected by cystic leukoencephalopathy, an autosomal recessive disorder. Although the pathogenic trigger of this RNase T2-deficient disease is still unclear, it has emerged that brain magnetic resonance images of affected patients are very similar to those of children suffering from intrauterine cytomegalovirus (CMV) infection. It is known that to counteract cellular antiviral mechanisms, CMV inactivates the endoribonuclease RNase L, which stops the infection by viral mRNA and rRNA degradation. RNase T2 might play a similar role in cellular immune response processes[51,52].

Recently, a group of patients affected by a genetic encephalopathy characterised by pediatric onset and cerebral calcification, clinically diagnosed as AGS, carry a biallelic loss of function mutations in the RNase T2 gene (Table 2). It is still unclear whether these patients belong to AGS cases or not and which can be the molecular role of RNase T2 in that pathology[53].

1	Early onset encephalopathy with psychomotor delay, spasticity, extrapyramidal signs and microcephaly, the latter appearing in the course of the first year of life.
2	Calcifications particularly visible at basal ganglia level (putamen, pallidus and thalamus), but also extending to the periventricular white matter.
3	Cerebral white matter abnormalities.
4	Cerebral atrophy.
5	Exclusion of pre–/perinatal infections, in particular the TORCH complex (toxoplasmosis, rubella, cytomegalovirus, herpes simplex virus).
6	Chronic lymphocytosis (> 5 cells/mm3) on CSF examination, not accompanied by any other sign of an infectious process.
7	Raised INF-alpha in the CSF (> 2 IU/ml).
8	Elevated neopterins and biopterins in CSF, sometimes associated with decreased folates.
9	Important systemic symptoms in the early stages of the disease include irritability, feeding and sleeping difficulties, unexplained fevers and the appearance of chilblain-like skin lesions on the fingers, toes and ears.
10	Genetic screening for mutations in the seven genes known to cause AGS allows definitive confirmation of the diagnosis in the majority (95%) of cases.

Criteria 1–5 plus criteria 6 or 7 were considered necessary to estabilish the clinical diagnosis of AGS. Criteria 8–10 were considered supportive criteria

Table 2: AGS diagnostic criteria[53].

Retroelements and DNA damage in AGS

Strong evidence suggests that for the AGS pathogenesis, the accumulation of nucleic acids is responsible for triggering the activation of immunity sensors, inducing the increase of CSF IFNα levels. As mentioned before, the nucleic acid species accumulated in the cytoplasm of AGS patients are yet to be defined. Currently, there are two main hypotheses: these accumulated nucleic acids can be derived by retroelements mobility, as LINE 1, or by DNA damage repair products. Indeed, both L1 intermediates and DDR can activate the innate immunity[1,4,6,7]. Furthermore, these two different pathways are interconnected to each other, L1 can induce double-strand breaks (DSBs) on DNA target by L1 endonuclease activity and there is evidence that some Nucleotide Excision Repair (NER), and DSB repair proteins regulate L1 mobility[54].

Transposable elements

Mobile DNAs, also known as transposons or "jumping genes", are repetitive mobile sequences widespread in nature and comprise 45% of the human genome. Transposable elements (TEs) are divided into two general classes based on their transposition intermediate: DNA for transposons and RNA for retrotransposons[55] (Figure 6).

DNA transposons

DNA transposons are mobile DNA sequences that move in the genome utilizing a single or double-stranded DNA as intermediate. Eukaryotic DNA transposons can be divided into three major subclasses: those that use a dsDNA as intermediate, the classic "cut and paste" transposons; those that use a related rolling-circle replication mechanism, the Helitrons; and finally, Mavericks transposons that replicate themselves through a self-encoded DNA polymerase. Both Helitrons and Mavericks mobility implicate the displacement and replication of an ssDNA intermediate, respectively, through a replicative, copy and paste process. All cut-and-paste transposons are characterized by a transposase encoded by autonomous copies and by the presence of terminal inverted repeats (TIRs). Retrotransposons dominate the human TE landscape, nonetheless, DNA transposons constitute about 3% of the human genome. Generally, there are no known active transposons in mammals[56].

Retrotransposons

Retrotransposons are subdivided into those sequences that contain Long Terminal Repeats (LTR) and those that do not (non-LTR). Both move by a "copy and paste" mechanism: the original transposon is maintained in situ, where it is transcribed; its RNA transcript is then reverse transcribed into DNA, which integrates into a new genomic location[55].

LTR Retrotransposons

LTR retrotransposons are relics of ancient retroviruses that once integrated into the germline and then lost their ability to infect other cells. Their genomic organization and replication cycle resemble that of infectious retroviruses like HIV, with reverse transcription of the viral mRNA primed by a specific cellular tRNA. LTR retrotransposons make up about 8% of the human genome, but no active LTR retrotransposon has been isolated. However, some intracisternal A-particle (IAP) and Etn/MusD family of LTR elements remain active in mice. Ty1/3 elements in yeast and the copia retrotransposons of Drosophila melanogaster are examples of active LTR retrotransposons in other species. A retrovirus that loses the function of its envelope (env) gene can become an LTR element. Conversely, an LTR retrotransposon that acquires an env gene could become a retrovirus. In humans, the closest elements to LTR retrotransposons are endogenous retroviruses (HERVs), which represent 7% of the genome. Most HERVs have accumulated several nonsense mutations, and no replication-competent HERVs are known. The HERV family that is most likely to be functional is the HERVK. However, many HERVs and their LTRs, through non-homologous recombination, can be expressed and may act as transcriptional regulatory elements for genes[55,57,58].

Non-LTR Retrotransposons

Non-long terminal repeat (LTR) retroposons are widespread in eukaryotic genomes. Non-LTR retrotransposons comprise a significant fraction, 15-17%, of human genomic DNA. Long interspersed elements (LINEs or L1s) and short interspersed elements (SINEs) comprise most of this group in mammals.

L1 elements are the most active autonomous transposable elements in humans, and a substantial fraction of the genome, more than 30%, is derived directly or indirectly from its retrotransposition activity. Despite the presence of more than 500,000 copies in the human genome, most L1s, due to point mutations, rearrangements, or truncations, no longer "jump", and only an estimated 80-100 elements remain currently active in any individual diploid human genome. L1 elements preferentially retrotranspose the RNA from which they were translated, a phenomenon termed cis preference. Although, this propensity for L1 to mobilize itself, other RNAs, such as Alu retroelements, SVAs, and other cellular transcripts are commonly retrotransposed by L1.

Alu elements are about 300 base pairs long and are therefore classified as short interspersed nuclear elements (SINEs). Alus are the most abundant and currently most active retrotransposon in humans with more than 1 million copies. They are 300bp Pol III RNAs that contain: an internal A and B box, an A-rich region that separates the left monomer from the right monomer, and a polyA tail. Alu elements, and SINEs in general, are non-autonomous retrotransposons that rely on LINE1-encoded proteins to catalyse their retrotransposition[55,59].

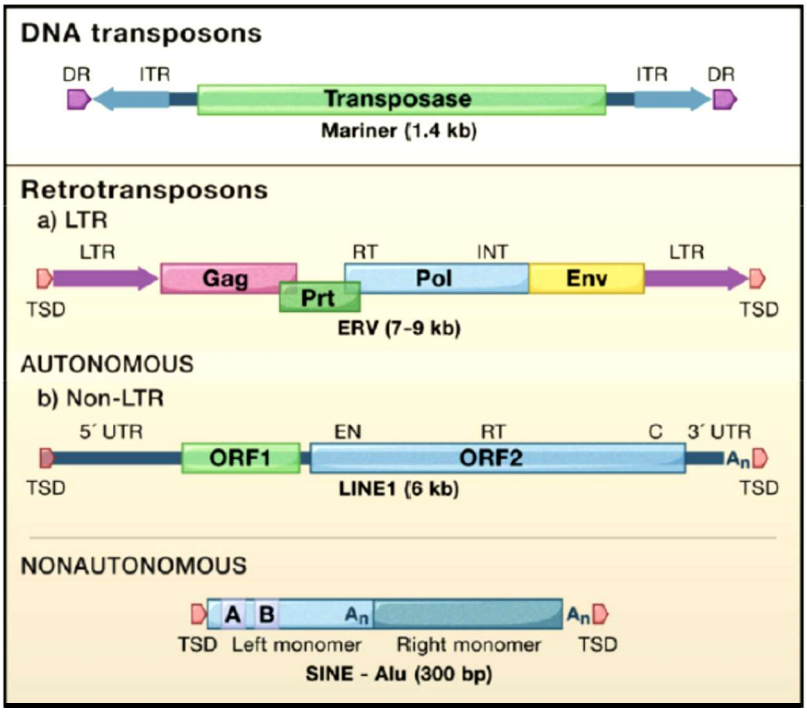


Figure 6: Types of transposable elements in Mammals. (Adapted from John L. Goodier, 2016)

Retrotransposition mechanism of L1 elements

The 6 kilobases bicistronic human L1 has a 5' untranslated region (UTR) that functions as an internal promoter, a 3' UTR that ends in a poly (A) tail, and two open reading frames (ORF1 and ORF2) on the sense strand. A weak L1 promoter on the antisense strand of the 5'UTR lies upstream of a recently identified 216-nt ORF0. ORF1 is a 40 kDa RNA-binding protein that forms trimers. ORF1p possesses chaperone activity in vitro, and although it is essential for L1 retrotransposition, its precise function remains unclear. ORF2 is a 149 kDa protein with endonuclease (EN) and reverse transcriptase (RT) activities. The L1 replication cycle starts with the synthesis of a bicistronic mRNA coding the ORF1 and ORF2 L1 proteins. ORF1p and ORF2p bind the L1 RNA and form a stable ribonucleoprotein particle (RNP) (Figure 7).

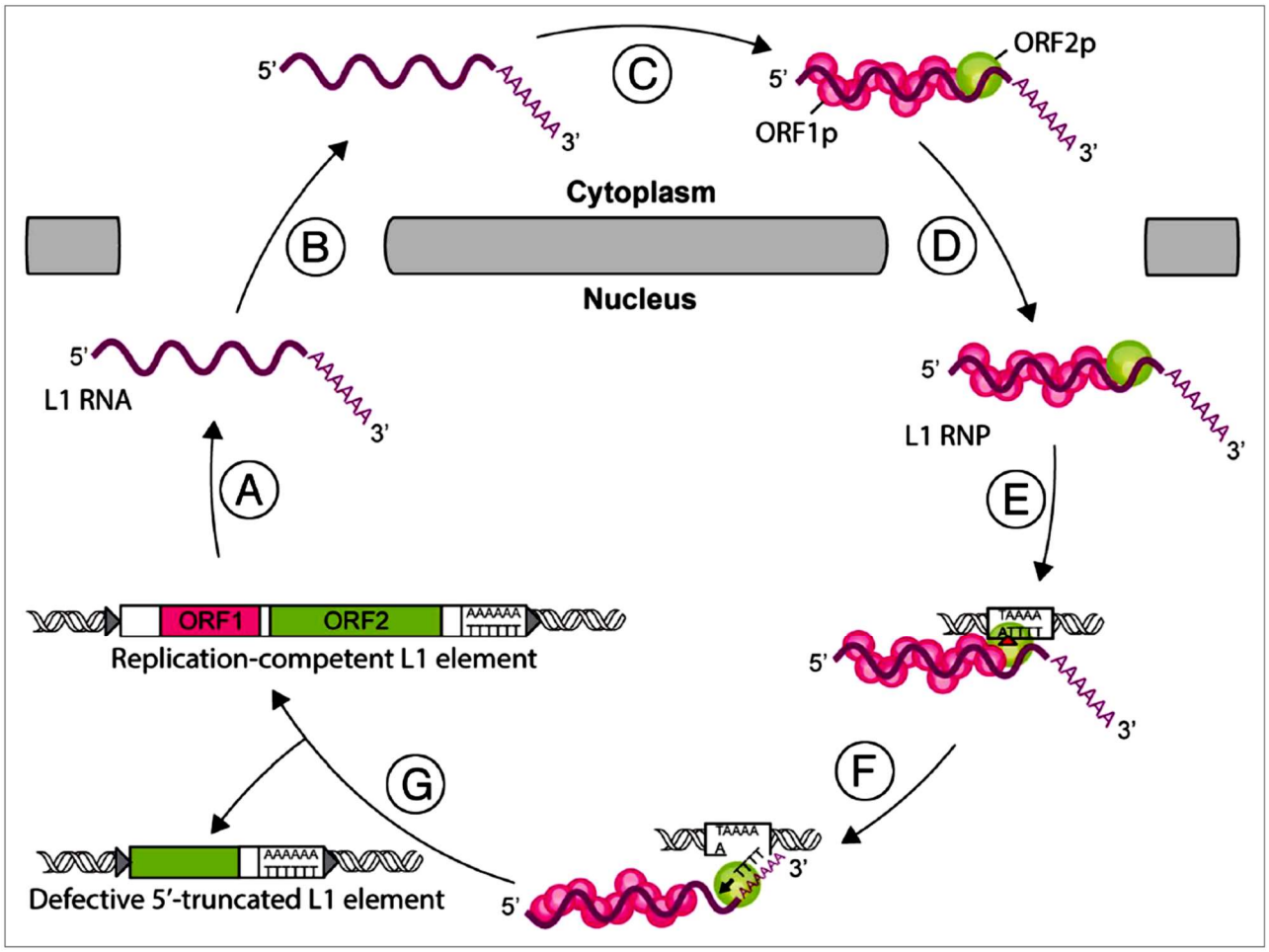


Figure 7: L1 life cycle[59].

This L1 RNP can mediate two different integration processes. In the canonical pathway, called target-primed reverse transcription (TPRT), ORF2p, with its endonuclease activity, nicks the chromosomal DNA at a 5′-TTTTA-3′ target site and then extends this liberated 3′-OH group, using the L1 RNA as a template. In an alternative pathway, the endonuclease-independent (ENi) retrotransposition or non-classical L1 insertion (NCLI) one, reverse transcription starts from a pre-existing DNA lesion, without the need of an endonuclease nick. An example of this second mechanism of L1 retrotransposition occurs at telomeres, the natural extremity of chromosomes[55,60] (Figure 8).

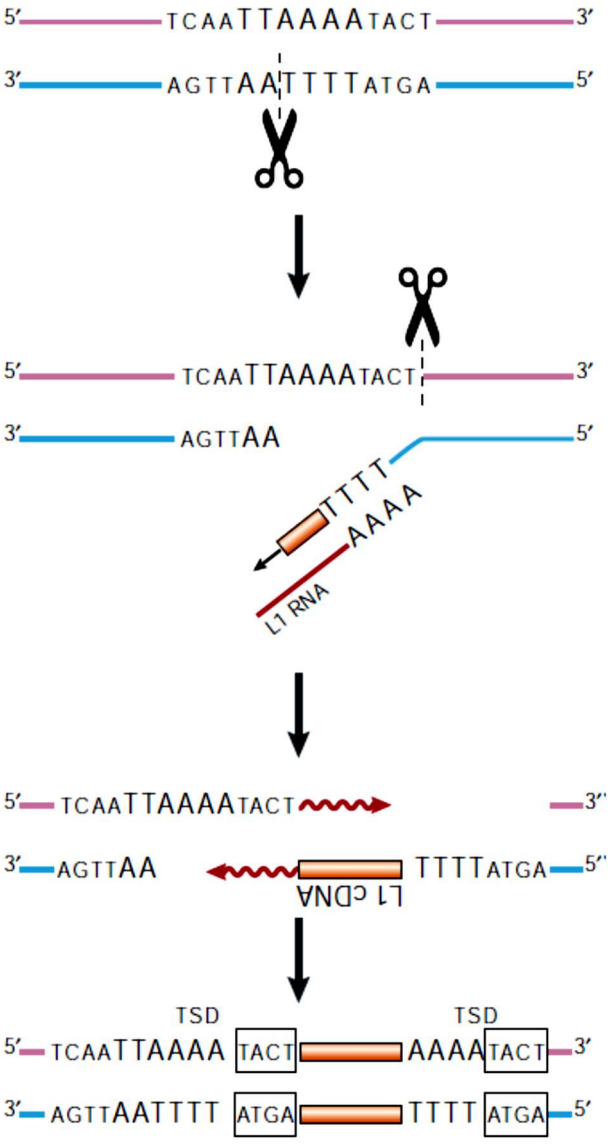


Figure 8: Target-primed reverse transcription (TPRT)[57].

L1 impact on the human genome

The first reported identification of an L1 insertion was in 1988, when Haig Kazazian and colleagues, studying a patient with hemophilia A, observed a new exonic L1 insertion in the X-linked gene factor VIII, where the next generation can inherit this genetic information. Since then, examples of human genetic disorders caused by de novo L1 insertions continue to accumulate, and more than 100 cases have been shown to trigger heritable diseases, such as: haemophilia, β-thalassaemia, Duchenne muscular dystrophy, cystic fibrosis, Apert syndrome, neurofibromatosis and cancer.

However, recent research has determined that L1 retrotransposition is not only limited to the germline. L1 expression and activity have been observed in several human tumours and in some tissues of the brain, suggesting that somatic retrotransposition is more frequent than previously

expected. Intriguingly, a new L1 insertion can not only be mutagenic by disrupting a coding sequence. Indeed, L1 insertions can impact the expression of nearby genes by generating: new splice sites, alternative promoters, adenylation signals, and transcription factor-binding sites. L1s can also contribute to genetic instability by generating target site deletions, insertions of flanking DNA (target site duplication sequences, TSD), recombination with other retrotransposons, and the possible generation of chromosomal inversions and interchromosomal translocations [61,62] (Figure 9).

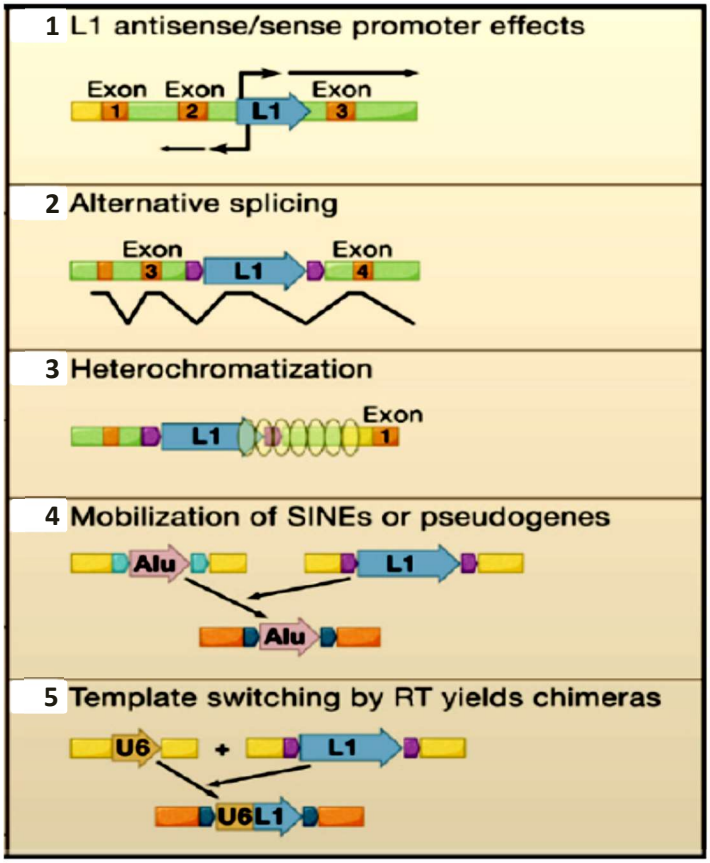


Figure 9: How L1 retrotransposons can affect the cell. (1) The L1 sense/antisense promoters in the 5'UTR can generate new transcription start sites of flanking genes. (2) Splice sites of a L1 copy integrated in an intronic region can cause alternative splicing, including a new exon, of a gene transcript. (3) L1 can alter the chromatin state, altering gene expression. (4) L1 reverse transcriptase can mobilize SINE retroelements and other mRNAs, leading to further genome expansion. (5) L1 reverse transcriptase switching from L1 RNA to other sequences, such as U6 or Alu RNAs, create chimeric insertions in the genome (Adapted from Goodier, J. L. & Kazazian, H. H, 2008).

Host defense mechanisms against LINE1 retrotransposition

Because of the high potential impact on cellular processes of L1, cells have developed several repressive mechanisms or checkpoints to inhibit L1 expression and retrotransposition. These repressive mechanisms differ depending on whether the cell is a germ or somatic one.

L1 is normally silenced in somatic cells in different ways. L1 epigenetic repression involves DNA hypermethylation, H3K9me3 and H3K27me3, and recruitment of NuRD repressive complex leading to the formation of heterochromatin at the L1 5' UTR promoter. When L1 is transcribed, its mRNA is deaminated and so targeted for degradation by the APOBEC family. L1 mRNA can also be targeted for degradation by piRNAs and miRNAs. Then L1 mRNA and proteins can be sequestered into stress granules to prevent L1 RNP entering into the nucleus. If L1 RNP crosses the nuclear envelope, there are a series of host inhibitory mechanisms to prevent L1 integration. During L1 reverse transcription, ERCC1/XPF complex, involved in NER process, can recognizes and remove L1 cDNA to restore the original target DNA sequence. Moreover, TREX1 binds and prevents the accumulation of L1 ssDNA transcribed by ORF2. Finally, phosphorylation of L1 ORF1 is required for retrotransposition, but the L1 life cycle step affected is still not clear.

L1 is predominately expressed in the germline, but in a controlled manner to maintain genomic integrity through each generation. Germline regulation uses L1's RNAs produced by the bidirectional promoter. Sense and antisense L1 RNAs can bind to each other to form dsRNAs, which can be cut by dicer into small fragments known as endo-siRNAs. Endo-siRNAs then degrade L1 mRNA through RNA interference. Another kind of L1 inhibition in germ cells involves the noncoding PIWI-interacting RNAs (piRNAs). These piRNAs inhibit L1 in germ cells through four mechanisms: RNA degradation, translational inhibition, DNA methylation, and histone modification.

These inhibitory mechanisms that protect healthy somatic and germ cells from the damaging effects of L1 are not all present in every cell and vary by cell type and embryonic origin[63].

Results

Over 50% of AGS cases carry a mutation in one of the three subunits of the RNase H2 complex. This heterotrimeric enzyme is conserved in all eukaryotes representing the major source of ribonuclease H activity in mammalian cells. RNase H2 plays several roles in mammalian cells: it removes ribonucleotides embedded in chromosomal DNA, it processes Okazaki fragments during DNA replication, it disrupts R-loops formed during replication and transcription collisions, and it removes RNA:DNA hybrids at telomeres [21,23,24,25,26]. The molecular mechanisms linking RNase H2 mutations with the AGS phenotype have not been elucidated yet.

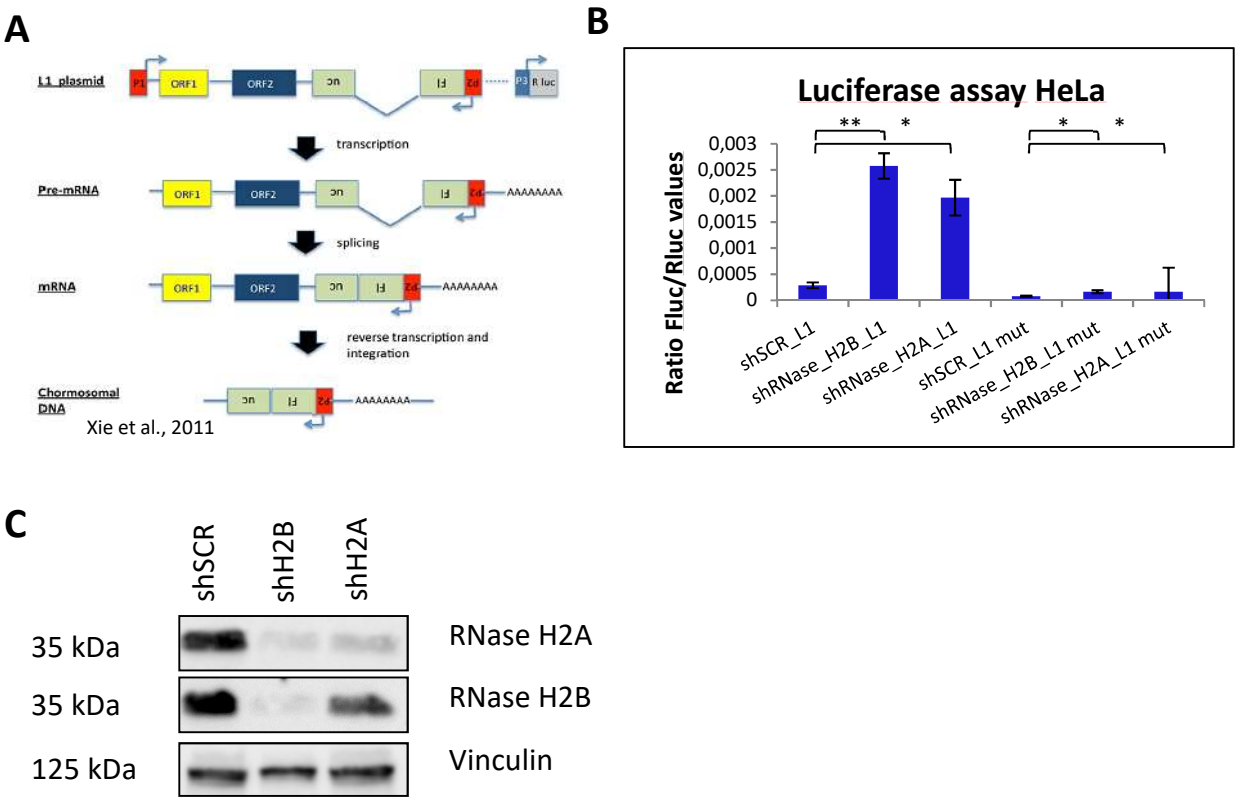


Figure 1: RNase H2 inhibition of L1 mobility shown in L1 luciferase retrotransposition assay. (A) A schematic drawing the L1 construct and of the Dual-Luciferase retrotransposition assay. A firefly luciferase (Fluc) gene is disrupted by an intronic sequence and inserted in the 3'UTR of L1 in anti-sense orientation respect to the L1 gene. A Renilla luciferase (Rluc) cassette is inserted in the backbone to allow the normalization. (B) L1 retrotransposition assay using L1-Fluc cassette is performed in HeLa cells transduced with RNase H2 shRNAs. Data are reported as the ratio between Fluc and Rluc signals and presented as the mean ± SD of three independent experiments. NS, not significant; *P ≤ 0.05, **P ≤ 0.01. (C) The efficiency of RNase H2 silencing was monitored by immunoblotting. The silencing of an RNase H2 subunit affects the protein level of other subunits.

Among all the AGS-causing genes, TREX1, ADAR1, and SAMHD1 have been shown to affect the correct metabolism of LINE1 retroelements, DNA sequences able to move in the genome through an RNA intermediate[18,20,37,47]. These findings provide a possible link between the dysregulation of retroelemets and the pathogenesis of AGS. Indeed, the nucleic acid species accumulated in the cytoplasm as a consequence of a mutation in AGS-causing genes are yet to be defined, and one of the major candidates is retrotransposition intermediates.

To test the possible involvement of endogenous retrotransposons in causing AGS, I used L1 retrotransposition assays to determine the effects of down-regulation of endogenous RNase H2 expression on the activity of engineered L1 retrotransposons[64,65,66] (Figure 1).

To knockdown RNase H2 expression, I transfected lentiviral constructs expressing specific short hairpin RNAs (shRNAs) into HEK293T cells, where lentiviruses were generated. I employed constructs specific for the B and C subunits of RNase H2. Then, I transduced the viruses collected from HEK293T cells into either HeLa or fibroblast MRC5 cells. Cells infected with RNase H2 shRNAs showed a reduction of the levels of RNase H2 subunits as compared to levels in cells infected with a plasmid expressing a scrambled shRNA lentiviral vector (negative control) by Western Blot (Figure 1C).

To test the effect of RNase H2 depletion on L1 retrotransposition, I employed a dual-luciferase retrotransposition assay using the pYX014 plasmid[66]. This plasmid carries an engineered L1 element that harbours the Firefly luciferase (Fluc) gene as a retrotransposition indicator. The Fluc reporter gene was cloned in the antisense orientation (relative to L1) in the 3'UTR of an L1 element, and an intron disrupted its coding sequence an L1 element and its coding sequence in the sense orientation (Figure 1A). Therefore, Fluc expression occurs in the transfected cells only after one round of retrotransposition. In particular, when L1 is transcribed, the intron is removed by splicing, and reverse transcription and integration are followed by the transcription of the intact Fluc gene. The resulting Firefly luciferase measurement serves as a read-out of L1 retrotransposition efficiency. A Renilla Luciferase expression cassette cloned in the backbone of the pYX014 plasmid has been used to normalize transfection efficiency.

I infected HeLa or MRC5 cells with anti-RNase H2 shRNA lentiviral constructs, and after 72 hours, I further transfected them with the pYX014 plasmid. Four days after pYX014 transfection, cells were lysed for luminescence analysis and L1 activity was measured by Fluc/Rluc ratio (L1 retrotransposition events/transfection efficiency). Our results showed that, in the absence of RNase H2, mobility of L1 is about 4-fold increase compared to the control (scramble shRNA). To confirm that the Fluc signal was originated from L1 retrotransposition, I performed these experiments by using

in parallel, as a negative control, the pYX015 vector: a construct that presents a missense mutation in the L1 ORF1 coding sequence that dramatically impairs L1 retrotransposition[66] (Figure 1B).

To further validate these preliminary data, I also employed a different pJM101/L1.3 retrotransposition cassette[65]. This cassette is similar to pYX014 described above, but the retrotransposition indicator Fluc gene is substituted by the neomycin phosphotransferase gene (neo^r) (Figure 2A). I measured the retrotransposition efficiency by counting the number of G418 (neomycin) resistant cells as a result of new L1 insertions.

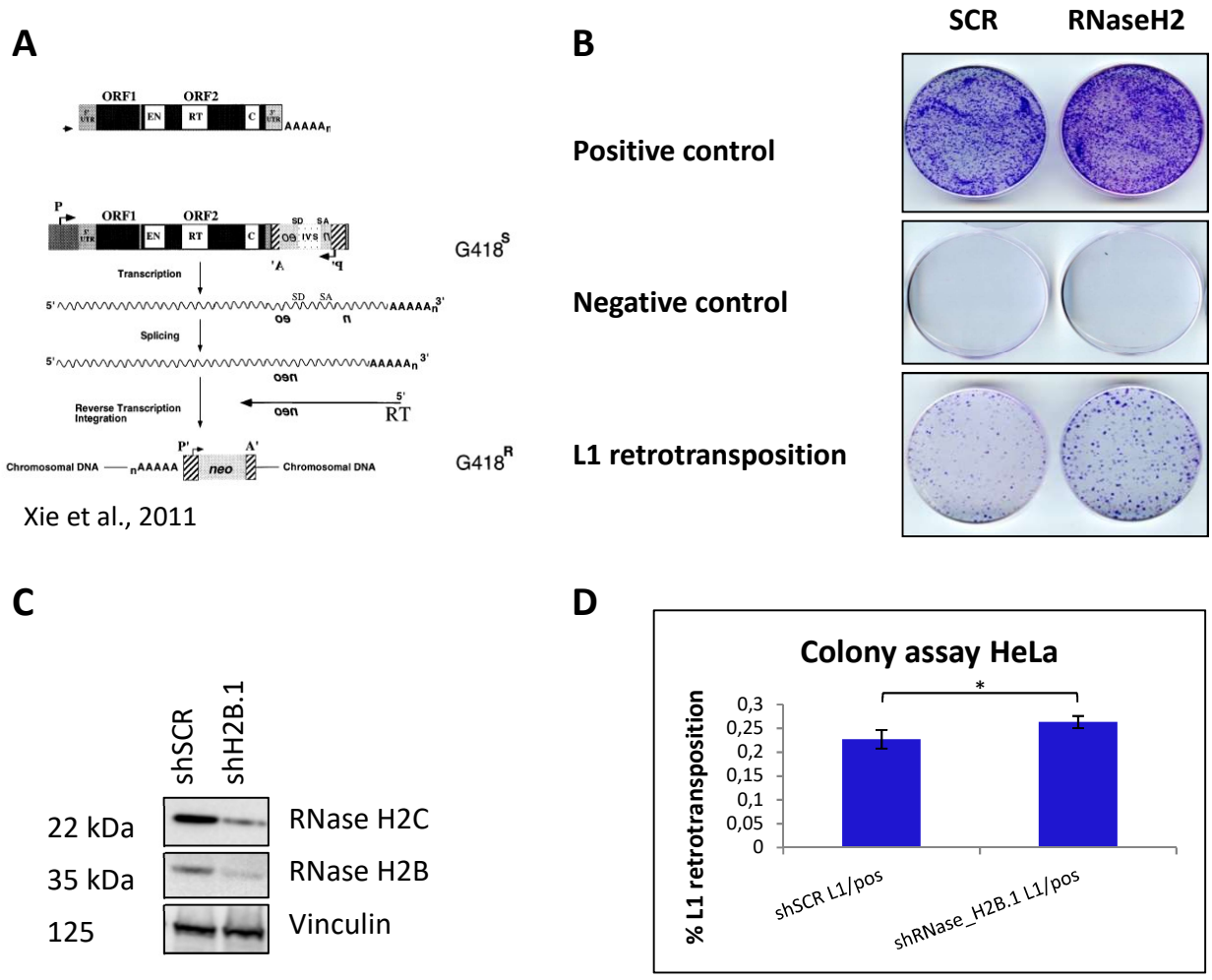


Figure 2: Suppression of L1 mobility by RNase H2 shown in the L1 colony retrotransposition assay. (A) A schematic drawing of the L1 construct and the colony retrotransposition assay. A neomycin phosphotransferase gene (neo^r), instead of Fluc one, is disrupted by an intronic sequence and inserted in the 3' UTR of L1 in an anti-sense orientation. Retrotransposition of this L1 element confers G418 (neomycin) resistance to host cells. (B) L1 assays were carried out transfecting HeLa RNase H2 silenced cells or not with this L1 retrotransposition construct and selecting for up of 14 days only G418 resistant cells. Representative culture dishes for each condition are shown. (C) The efficiency of RNase H2 silencing was tested in a Western Blot analysis. (D) L1 retrotransposition assays performed in RNase H2 depleted HeLa cells. In this graph, the colonies indicating L1 retrotransposition events, in RNase H2 silenced cells or not, are normalised on the positive control colonies, cells transfected with a plasmid that harbour a G418 resistant cassette. Data are shown as the mean ± SD of a single experiment with three replicates. NS, not significant; *P ≤ 0.05, **P ≤ 0.01.

In particular, by using an experimental protocol similar to the one described above, I infected cells first with the anti-RNase H2 shRNA plasmids or control plasmid, and, after 72 hours, I transfected them with the pJM101/L1.3 vector. Seventy-two hours later, cells were grown in media supplemented with G418 and after about 10 days in G418 selection, the remaining cells were fixed and stained with crystal violet to facilitate the visualization and allow the counting of the colonies formed, indicative of individual retrotransposition events (Figure 2B).

To confirm that the results obtained were due to the reduction of RNase H2 expression acting specifically on L1 retrotransposition, I carried out these experiments by using the following controls. A pcDNA3 plasmid that confers resistance to neomycin to the transfected cells without the need of retrotransposition was employed; the resistant colonies were thus indicative of the transfection efficiency. Moreover, I employed, as a negative control, the pJM105/L1.3 cassette, identical to pJM101/L1.3, but containing a missense mutation in the RT domain of the ORF2 gene that dramatically reduces L1 retrotransposition efficiency. Even in this assay, the knockdown of RNase H2 expression in HeLa or MRC5 cells induced increased L1 retrotransposition efficiency, as compared to scramble shRNA infected cells (Figure 2D).

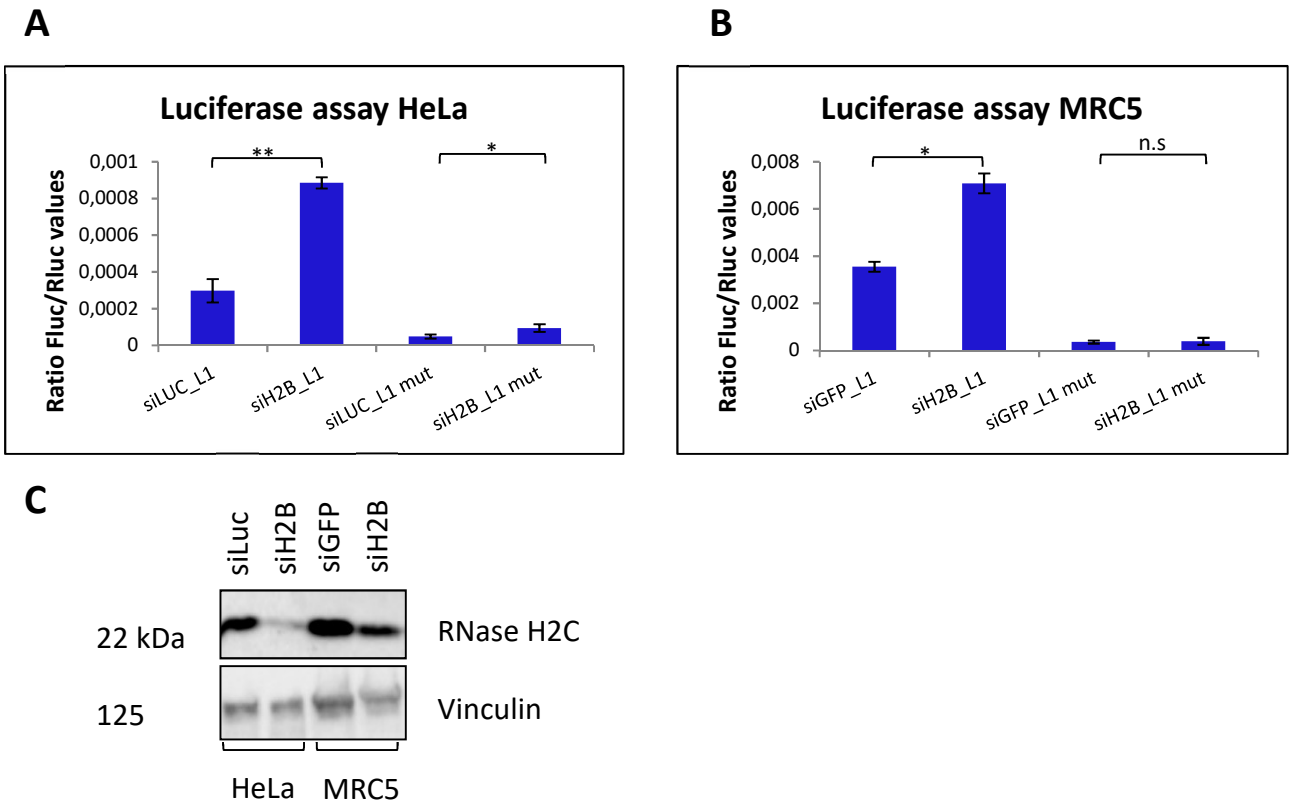


Figure 3: RNase H2 inhibits L1 mobility shown in an infection-independent manner. (A), (B) Dual-luciferase L1 retrotransposition assays performed in HeLa and MRC5 cells respectively, using siRNA against RNase H2 instead of lentiviral construct carrying shRNAs. Data are reported as the ratio between Fluc and Rluc signals and presented as the mean ± SD of three independent experiments. NS, not significant; *P ≤ 0.05, **P ≤ 0.01. (C) Western Blot analysis of RNase H2 silencing in HeLa and MRC5 cells.

To exclude any infection-dependent effect on L1 mobility levels in the previous assays, I performed luciferase retrotransposition assays in HeLa and MRC5 cells transfected with siRNAs against RNase H2. As shown in Figure 3, also in this case, in the absence of RNase H2, L1 mobility increases. Since similar results were obtained from retrotransposition assays in RNase H2 depleted cells through infection or transfection, we can exclude an infection-dependent impact on L1 retrotransposition, confirming the involvement of RNase H2 in L1 regulation.

To confirm the specificity of the RNase H2 knockdown experiments, I complemented silenced cells with a corresponding shRNA-resistant subunit in the context of retrotransposition assays. In particular, cells infected with a shRNA against RNase H2B subunit were transfected with H2A, H2C, and a shRNA-resistant H2B subunit (H2Br) (Figure 4A). As shown in figure 4B, the resistant form of RNase H2B was successfully expressed in RNase H2 silenced cells and partially suppressed the increased retrotransposition observed when RNase H2 is knocked down (Figure 4C). These results confirm the involvement of RNase H2 in modulating L1 retrotransposition.

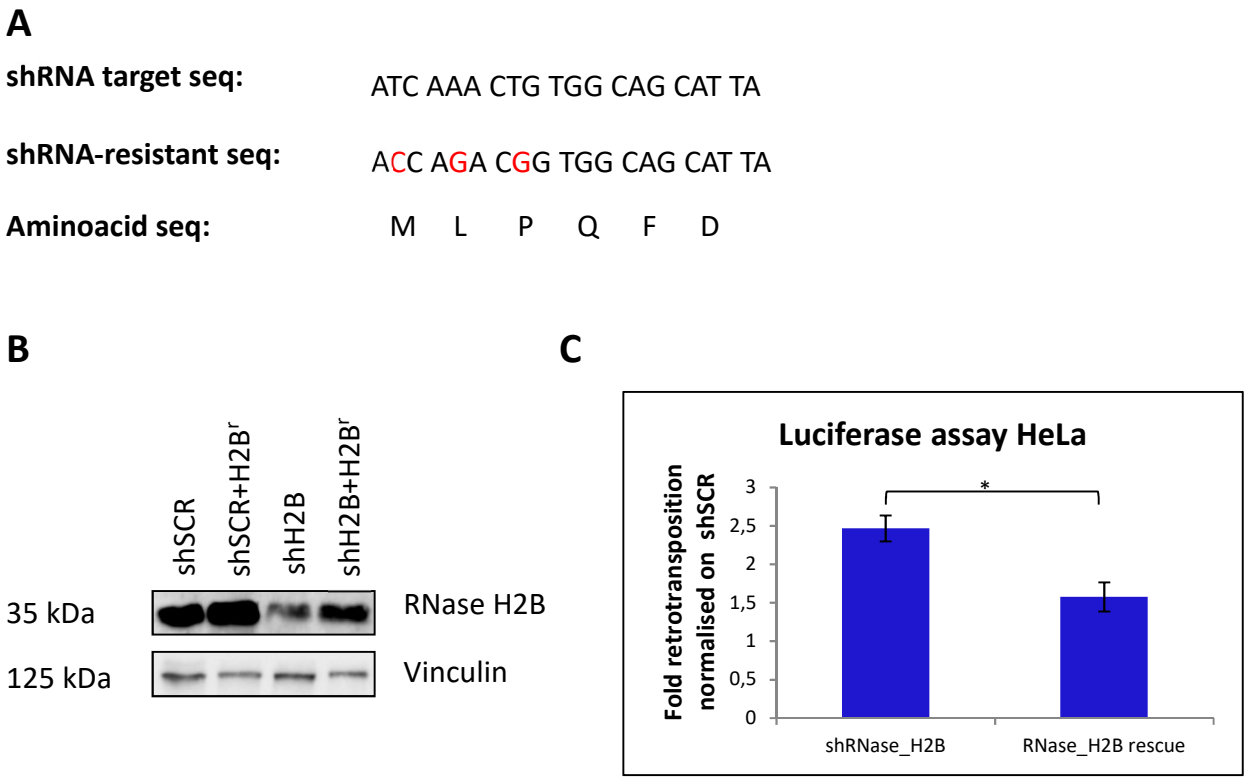


Figure 4: RNase H2 complementation in L1 retrotransposition assays. (A) Schematic representation of RNase H2 shRNA-resistant rescue strategy. (B) The efficiency of RNase H2 silencing and complementation was monitored by Western Blot analysis. (C) L1 retrotransposition assays performed in RNase H2 depleted or complemented HeLa cells. In this graph, L1 retrotransposition efficiency is normalised on shSCR cells one. Data are reported as the ratio between Fluc and Rluc signals, normalised on shSCR, and presented as the mean ± SD of three independent experiments. NS, not significant; *P ≤ 0.05, **P ≤ 0.01.

During the first two years of my PhD, other groups published conflicting results concerning RNase H2 and L1 mobility. In particular, Choi et al., using an shRNA strategy, concluded that RNase H2 inhibits L1 mobility[29], similarly to what we observe. On the other hand, Benitez-Guijarro et al., studying RNase H2 KO clones, reported that in the absence of RNase H2, L1 mobility decreases[30]. To try to address these contrasting observations, we decided to change our approach and start to evaluate RNase H2 involvement in endogenous L1 regulation. First of all, we wanted to verify whether RNase H2 has an impact on the number of L1 copies integrated into the genome. To measure these little differences of highly repetitive genomic L1s, I used a qPCR approach[67]. Briefly, I used in the same qPCR reaction a Taqman probe designed against L1 ORF2, conjugated with the FAM fluorophore, and as control, one conjugated with the VIC fluorophore and designed against α satellite sequences, repetitive non-coding DNA of the centromeric heterochromatin. In this way, I analysed L1 copy number variations (L1 CNVs), in genomic preparation extracted from RNase H2-silenced fibroblasts or control ones (Figure 5B). The same approach has been used with RNase H2B mutated primary fibroblasts and control fibroblasts (Figure 5C). This analysis revealed that RNase H2 silenced/mutated cells, host a lower number of copies of L1 elements compared to controls (Figure 5B, C).

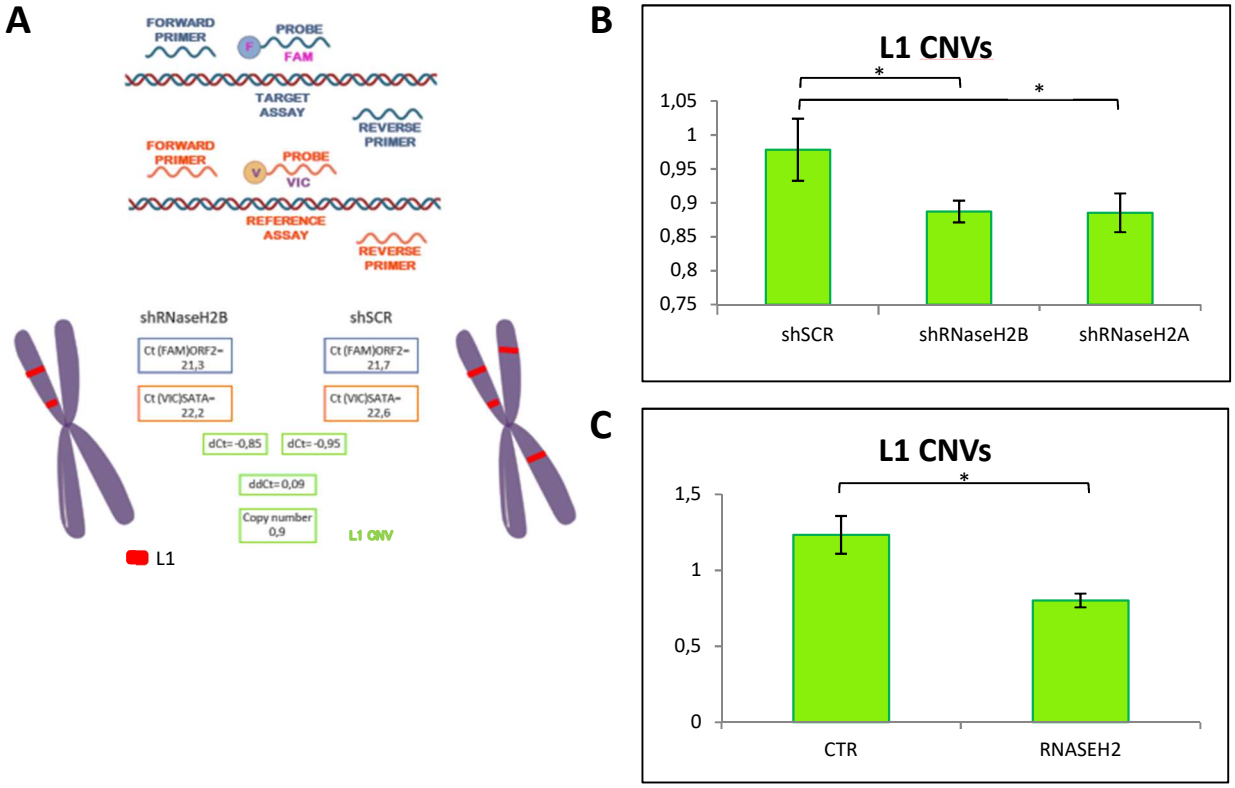


Figure 5: L1 number of copies decreases in RNase H2 silenced or mutated fibroblasts. (A) A Schematic draw of the L1 CNVs analysis through qRT PCR using probes against L1 ORF2 and SATA repetitive sequences to normalise. (B), (C) L1 CNVs analysis of RNase H2 silenced MRC5 or primary mutated fibroblasts, respectively. Each experimental condition is represented by six biological replicates and analysed through at least six technical replicates, and the data in the graphs are shown as the mean ± SD. NS, not significant; *P ≤ 0.05, **P ≤ 0.01.

Then, we evaluated how RNase H2 affects endogenous L1 expression levels by qRT PCR. To this aim, we employed MRC5 fibroblasts where RNase H2 was silenced with shRNAs, and primary fibroblasts from AGS patients carrying a mutation in RNaseH2. As shown in Figure 6A, when I silenced RNase H2, I observed an increase in both L1 5'UTR and ORF2 (L1 N-terminal portion) expression levels compared to non-silenced cells, indicating a full-length L1 mRNA increase in a depleted condition. Furthermore, in RNase H2-silenced MRC5 fibroblasts, I observed increased levels of INFα that can activate type I IFN-induced signalling leading to the main pathogenic trigger of AGS syndrome, the immunity hyperactivation (Figure 6A). I obtained similar expression level variations in primary fibroblasts derived from an AGS patient carrying RNase H2B mutation compared to normal primary fibroblasts: in mutated cells, L1 and IFNα levels are higher (Figure 6B). These results seem to suggest the existence of a negative correlation between RNase H2 levels and those of L1 and IFNα.

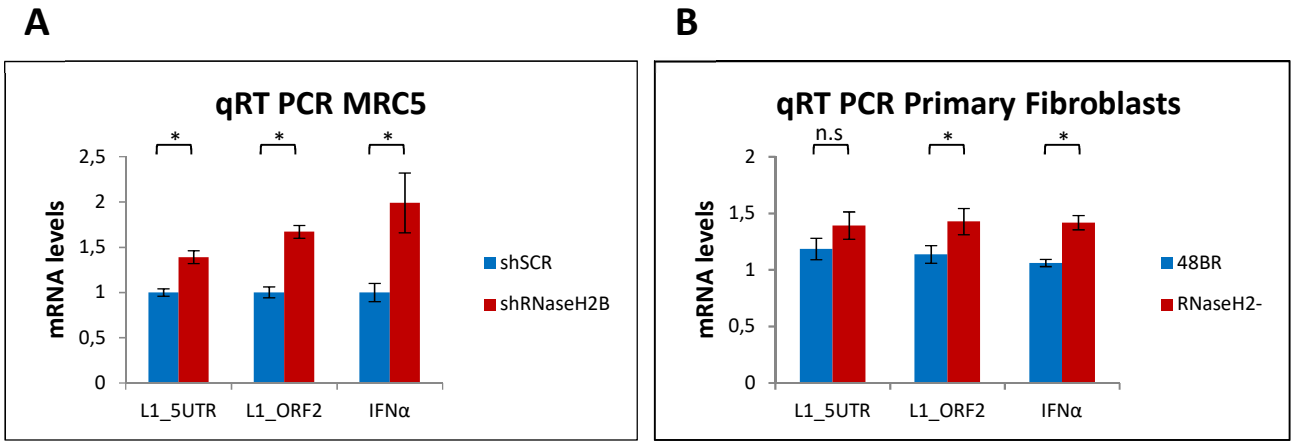


Figure 6: L1 and IFNα increased expression levels in RNase H2 silenced or mutated fibroblasts. (A) Expression levels analysis of L1 5'UTR, L1 ORF2, and IFNα by qRT PCR in RNase H2 depleted MRC5 fibroblasts. (B) L1 and IFNα expression levels analysis by qRT PCR in primary fibroblasts derived from an AGS patient carrying a point mutation in RNase H2B subunit (A177T) compared to control primary fibroblasts. (A), (B) Data in those graphs are presented as the mean ± SD of three biological replicates analysed through technical triplicates. NS, not significant; *P ≤ 0.05, **P ≤ 0.01.

Considering that we report increased levels of L1 RNA, but fewer copies of L1 integrated into the genome, it is reasonable to speculate that, in cells carrying AGS mutated genes, there must be an accumulation of L1 intermediates in the cytoplasm.

I then started to study a possible correlation between loss of RNase H2 genes functions, L1 RNA levels, and cytoplasmatic accumulation of nucleic acids.

In order to modulate L1 mRNA levels, I tested different technical approaches to silence L1, and finally, I opted for a combination of siRNAs (Figure 7A). As shown in the graph of Figure 7A, in RNase H2-silenced MRC5, L1 and IFNα levels increase compared to control cells (shSCR), but when

I lowered L1 levels, though L1-specific shRNAs, also IFNα levels decreased, suggesting a link between these two phenotypes. To look for cytoplasmic nucleic acids, I started to analyse the presence of ssDNA by immunofluorescence. The data reported in Figure 7B, C, show that in RNase H2-depleted cells, there is a higher content of ssDNA signal compared to control cells, which is suppressed by silencing also L1 (Figures 7B, C).

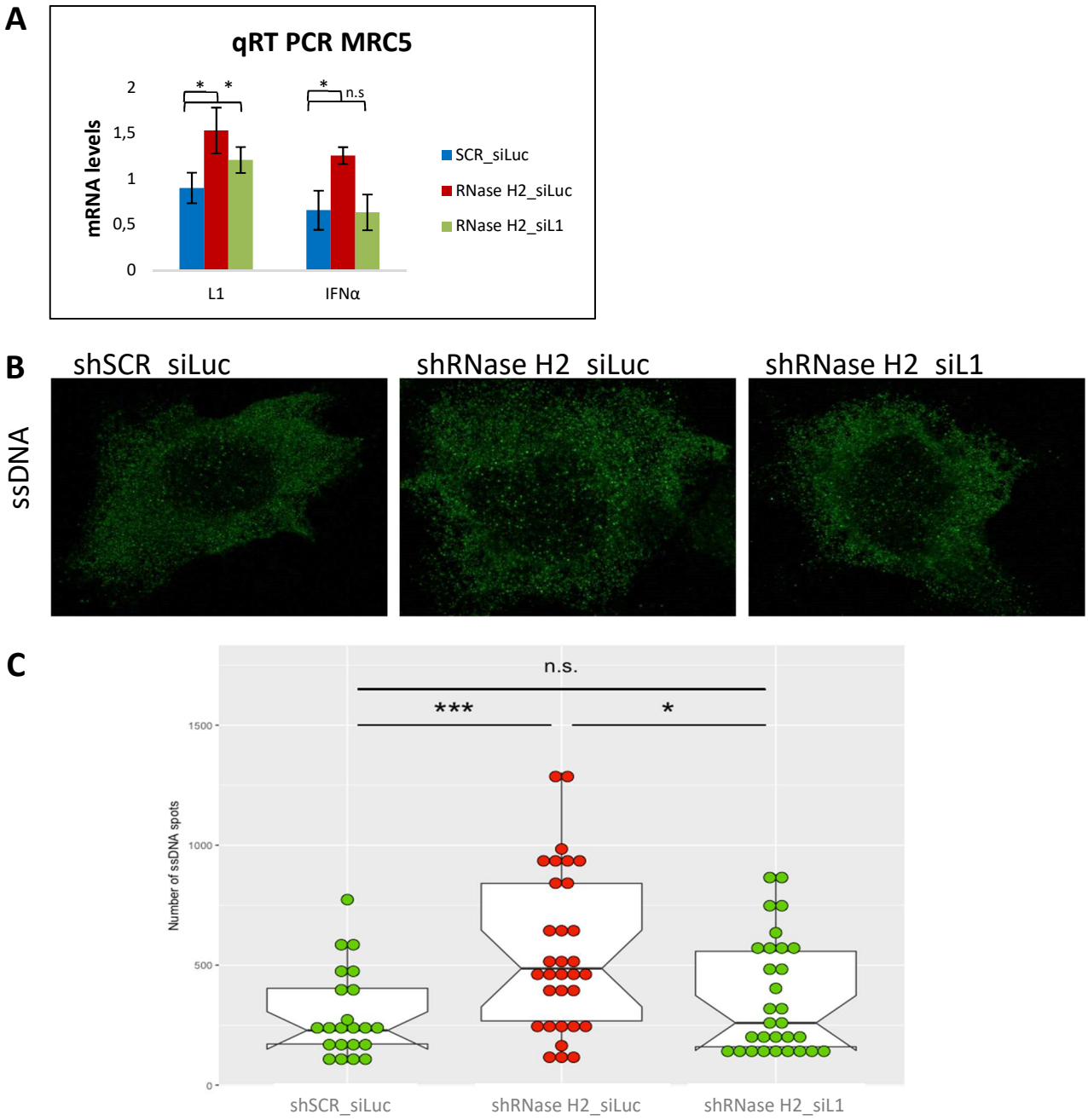


Figure 7: L1 silencing effects on IFNα levels and ssDNA accumulation in RNase H2 depleted cells. (A) L1 and IFNα expression levels analysis by qRT PCR in RNase H2 +/- L1 siRNAs depleted MRC5 compared to shSCR_siLuc cells. (B), (C) Representative images of ssDNA immunofluorescence and relative quantification (automated analysis with ImageJ software) of mainly cytoplasmatic ssDNA puncta in single MRC5 cells. Statistical significance was determined by two-tailed Student's t-test with the p-values indicated in the graph.

We then extended the studies on endogenous L1 to other AGS-causing genes to investigate whether the phenotypes observed in RNase H2-depleted cells are general characteristics of AGS patients. In other words, we wanted to verify if loss of the nucleic acid metabolizing enzymes in AGS patients leads to the dysregulation of L1 mobility, which may be, at least in part, responsible for the pathogenesis. ADAR1 and TREX1 have been reported to inhibit the mobility of an exogenous L1 construct[18,20,47].

First of all, I analysed L1 copy number variations in the genome of ADAR1-silenced MRC5 cells and TREX1 mutated primary fibroblasts. As reported in Figure 8, the number of integrated L1 copies is lower in ADAR1-silenced cells or TREX1 mutated fibroblasts derived from an AGS patient. It is not clear why the results reported in the literature, using retrotransposition assays via an L1 artificial exogenous construct, are contrasting with the results obtained with these endogenous LINE1 analyses. It can be possible that L1 retrotranspostition assays are influenced by an upstream transcriptional regulation.

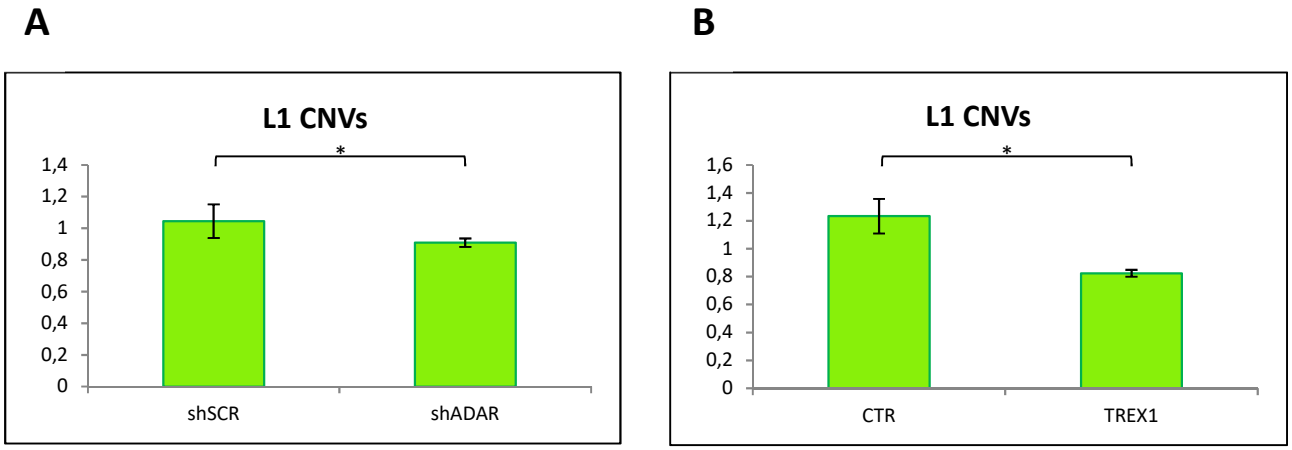


Figure 8: L1 number of copies decreases in ADAR1 silenced MRC5 or TREX1 mutated primary fibroblasts. (A) L1 CNVs analysis of ADAR1 silenced MRC5 (with shRNA). (B) L1 CNVs analysis of TREX1 primary mutated fibroblasts. (A), (B) L1 CNVs analysis through qRT PCR performed using probes against L1 ORF2 and SATA repetitive sequences to normalise. Each experimental condition is represented by six biological replicates and analysed through at least six technical replicates, and the data in the graphs are shown as the mean ± SD. NS, not significant; *P ≤ 0.05, **P ≤ 0.01.

I then estimated L1 expression levels in depleted/mutated ADAR1 and TREX1 cells compared to controls. As shown in Figure 9, in ADAR1 or TREX1 silenced/mutated cells, L1 expression levels are higher compared to control cells. Furthermore, I observed a positive correlation between L1 expression and IFNα expression levels in depleted conditions again.
Interestingly, the effects on endogenous L1 are similar for all of these three AGS-related genes, suggesting a common thread in AGS pathogenesis.

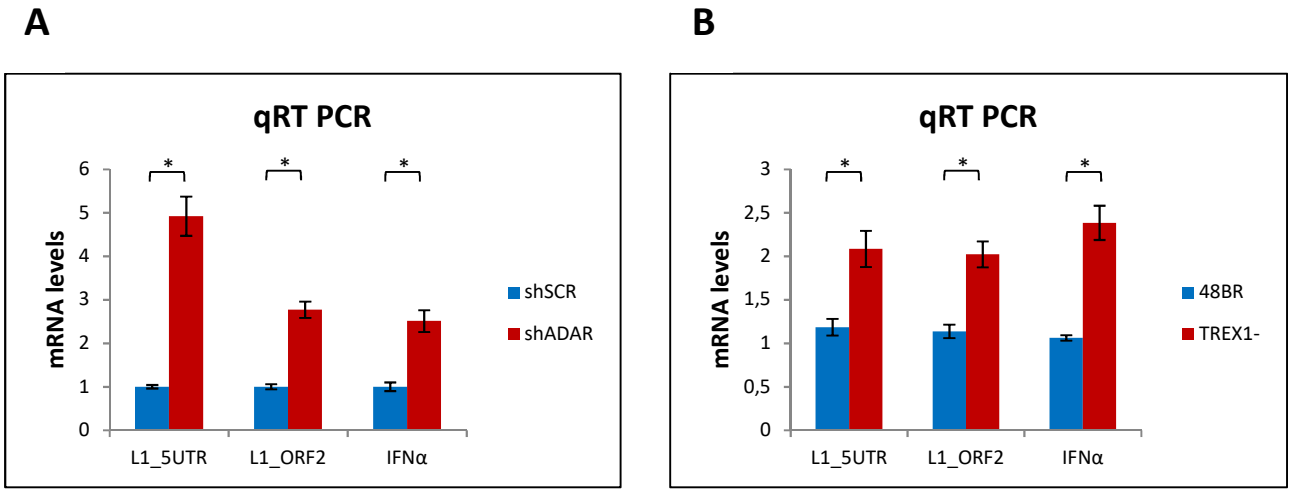
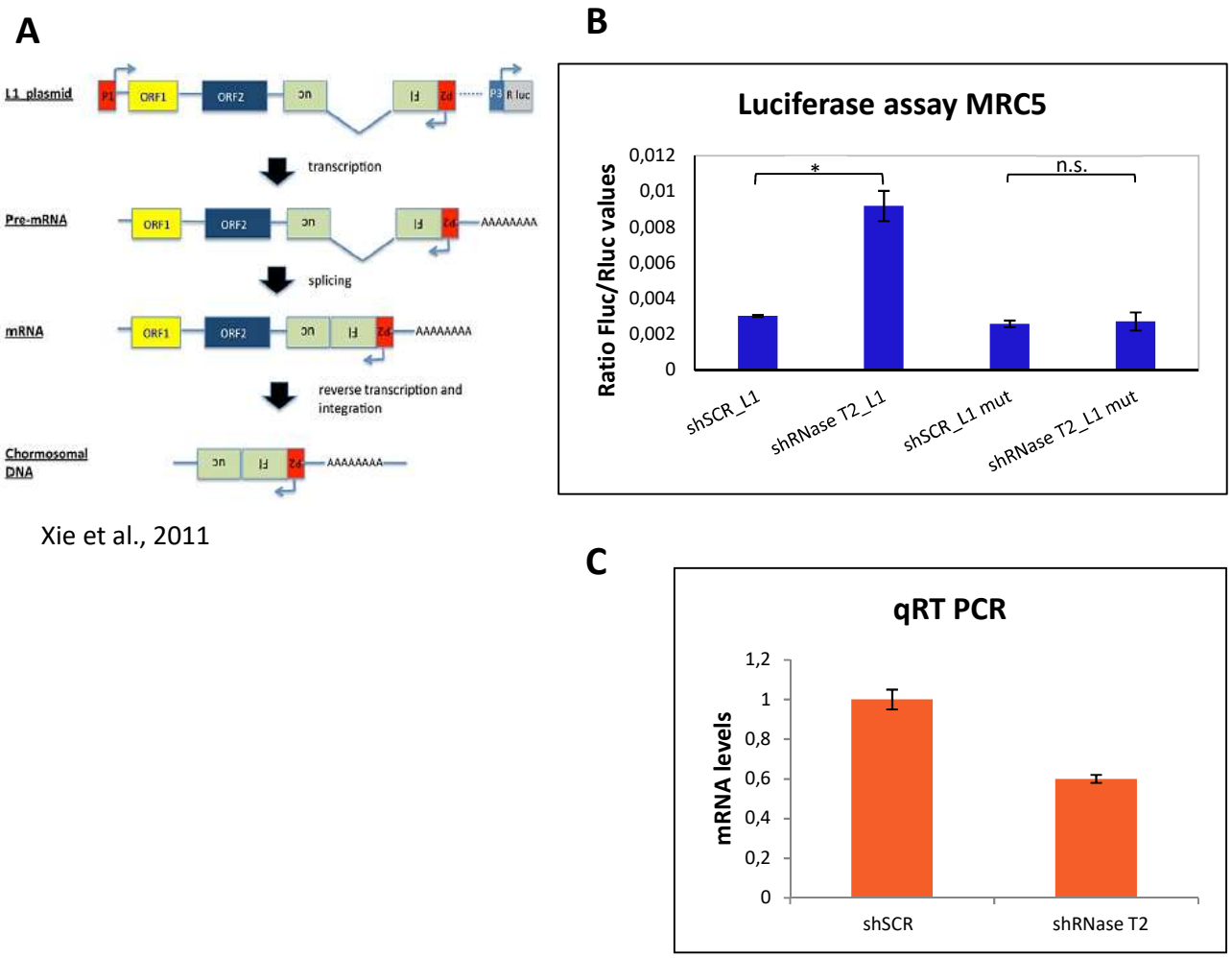


Figure 9: L1 and IFNα increased expression levels in ADAR1 or TREX1 depleted cells. (A) Expression levels analysis of L1 5'UTR, L1 ORF2, and IFNα by qRT PCR in MRC5 infected with shRNA against ADAR1. (B) L1 and IFNα expression levels analysis by qRT PCR in primary fibroblasts derived from an AGS patient carrying a point mutation in the TREX1 gene and control primary fibroblasts. (A), (B) Data in the graphs are presented as the mean ± SD of three biological replicates analysed through technical triplicates. NS, not significant; *P ≤ 0.05, **P ≤ 0.01.

The RNase T2 is the only member of the Rh/T2/S glycoprotein family in humans. Recently, a group of infants affected by a genetic encephalopathy with cerebral calcification was reported to carry a biallelic loss of function mutations in the RNase T2 gene[53]. It is still unclear whether these patients could be classified as AGS cases or not and what is the molecular role of RNase T2 in that syndrome[53]. To explore the possibility that RNase T2 mutations may be linked to the pathogenesis of AGS, I verified, through retrotransposition assays, if this enzyme is be involved in L1 mobility regulation, like other AGS related enzymes (Figure 10). Using L1 luciferase retrotransposition assays, I show that in RNase T2-silenced cells, L1 mobility increases (Figure 10A), suggesting that this enzyme is an L1 retrotransposition inhibitor.



Xie et al., 2011

Figure 10: RNase T2 inhibition of L1 mobility shown in L1 luciferase retrotransposition assay. (A) A schematic drawing of the L1 construct and the Dual-Luciferase retrotransposition assay. (B) L1 retrotransposition assay using L1-Luc cassette is performed in MRC5 cells transduced with RNase T2 shRNA. Data are reported as the ratio between Fluc and Rluc signals and presented as the mean ± SD of three independent experiments. NS, not significant; *P ≤ 0.05, **P ≤ 0.01. (C) The efficiency of RNase T2 silencing was monitored by qRT PCR.

Furthermore, I analysed by qRT PCR whether RNase T2 affects endogenous L1 expression levels. As it is shown in figure 11A, in primary fibroblasts carrying an RNase T2 gene mutation, I observed an increase of L1 expression levels compared to normal fibroblasts; interestingly, also INFα expression levels are elevated in mutated cells, similarly to what observed in the AGS conditions described above. Therefore, in these RNase T2 mutated fibroblasts, I can observe a set of phenotypes that are likely related to AGS syndrome.

Finally, we wanted to verify the RNase T2 impact on the number of L1 copies integrated into the genome. I analysed genomic preparation extracted from RNase T2 mutated primary fibroblasts or control fibroblasts by qRT PCR, using the previously described probes against L1 and the α satellite repetitive sequences as normalizer (Figure 11B).

In this set of experiments, I did not detect significant differences in L1 copy number between mutated fibroblasts and control cells (Figure 11B). These data suggest that RNase T2 may be involved in L1 regulation in a different way compared to the other AGS-causing genes that we tested.

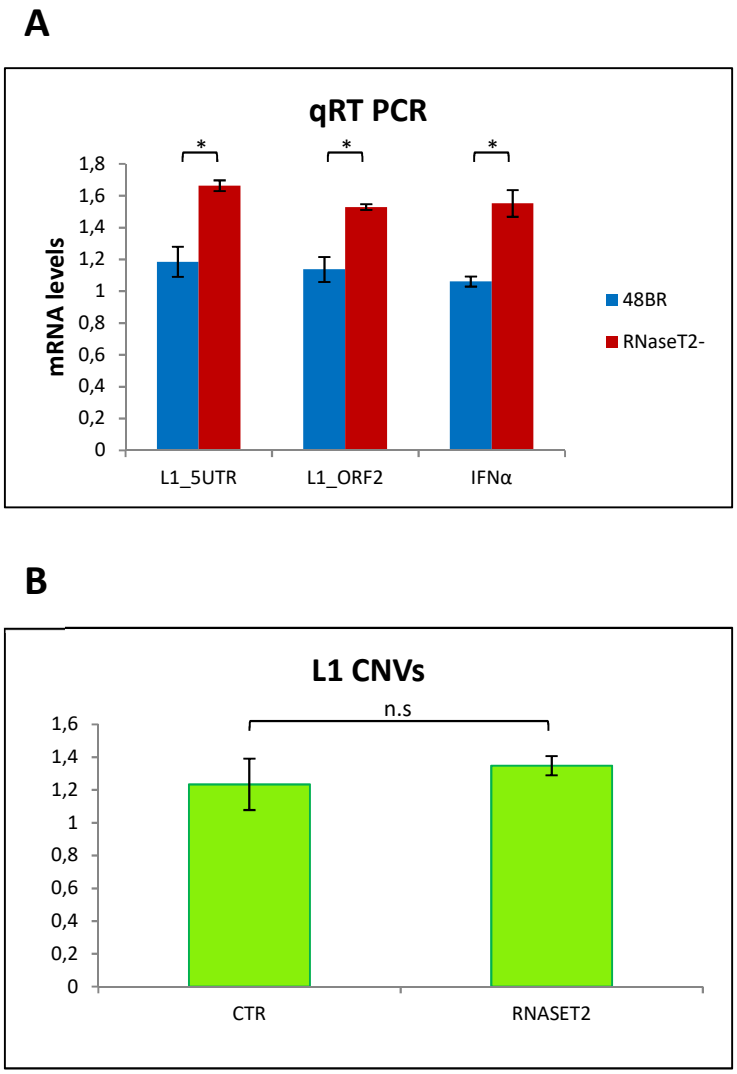


Figure 11: L1 endogenous expression levels and genomic CNVs in RNase T2 mutated fibroblasts. (A) Expression levels analysis of L1 5'UTR, L1 ORF2, and IFNα by qRT PCR in RNase T2 mutated fibroblasts. (B) qRT PCR L1 CNVs analysis using probes against L1 ORF2 and SATA repetitive sequences to normalise. Each experimental condition is represented by six biological replicates and analysed through at least six technical replicates, and the data in the graphs are shown as the mean ± SD. NS, not significant; *P ≤ 0.05, **P ≤ 0.01.

Discussion

RNase H2 regulation of LINE1 retroelements

RNase H2 belongs to the family of RNase H enzymes, that process the RNA moiety of RNA:DNA hybrids. RNase H2 is a heterotrimeric enzyme conserved in all eukaryotes, and it's the primary source of ribonuclease H activity in mammalian cells. Mutations in each one of the RNase H2 subunits account for over 50% of AGS patients, but the molecular mechanisms linking RNase H2 mutations to AGS pathology have not been elucidated yet.

Recent studies from our and other groups have attributed to RNase H2 the role of preserving genome integrity by controlling genomic rNTPs incorporation during DNA replication[21,24,25,26]. RNase H2 depletion in mice induces massive rNMPs incorporation in the genome, causing p53 dependent apoptosis and embryonic lethality[4,23,26]. Recently, two generated knock-in mice expressing RNase H2A, and H2B AGS mutations showed a cCAS-STING dependent interferon signature, similarly to the patients[27,28].

Strong evidence suggests that an accumulation of unprocessed, endogenous nucleic acids can trigger sensors of the immunity. The nucleic acid species accumulated in the cytoplasm of RNase H2 mutated cells are yet to be defined[2,5]. One speculative model suggests the accumulation of DNA repair by-products due to genomic rNTPs incorporation [4,27,28].

Among all the AGS-causing genes, TREX1, ADAR1, and SAMHD1 have been correlated with correct metabolism of LINE1(L1) retroelements, the most active autonomous transposable elements in human cells, able to move in the genome through an RNA intermediate[18,20,37,47]. Interestingly, also RNase H2 may regulate the mobility of endogenous L1 retroelements by degrading RNA:DNA hybrids formed during their replication. Then, another possible source of immunostimulatory nucleic acids in RNase H2 deficiency condition can be endogenous retroelements intermediates.

We decided to focus our efforts on the characterisation of a possible role of RNase H2 as an L1 retrotransposition regulator. To test this possibility, I used L1 retrotransposition assays to determine the effects of RNase H2 down-regulation on the activity of engineered L1 retrotransposons[64,65,66]. Using this assay, RNase H2 activity clearly counteracts L1 mobility (Figures 1,2).

These results are also reproduced by depleting RNase H2 using a siRNA approach, excluding any possible interference caused by adenovirus infection (Figure 3).

To further validate these preliminary data, I complemented silenced cells with a corresponding shRNA resistant subunit, testing retrotransposition mobility as previously described. As displayed in Figure 4, the rescue of RNase H2 expression restricted L1 integration efficiency, suggesting that the altered L1 mobility was specifically caused by RNase H2 knockdown. Therefore, considering the results obtained from these assays, we speculated that RNase H2 inhibits L1 mobility.

At the same time, another research group published experiments and results similar to those I obtained, RNase H2 knockdown in different human cell lines seems to cause an increase in LINE1 retrotransposition mobility[29]. However, the exact mechanism remains far from clear.

Another group, indeed, published an opposite work showing that RNase H2 has a positive role in L1 mobility. This group used the same mobility assay described in my thesis but performed on RNase H2A KO HeLa cellular clones instead of KD conditions[30]. Considering that RNase H2 KO induces embryonic lethality in mice, we cannot exclude that during clonal selection, the authors could have isolated genetic suppressors of RNase H2 activity loss, altering results of retrotransposition assays.

Since retrotransposition assays used have several critical points, such as L1 over-expression, we decided to focus our studies on L1 endogenous regulation.

First of all, we wanted to analyse L1 copy numbers in the genome of RNase H2 depleted cells to test if RNase H2 contributes to maintaining a low rate of endogenous L1 retrotransposition. To perform this evaluation, I used a qPCR approach described in Coufal, N. G. et al[67]. Unexpectedly, I detected fewer L1 copies integrated into the genome of RNase H2 silenced/mutated cells compared to controls (Figure 5). Although a small decrease in L1 CNVs was observed, it should be considered that L1 copies constitute about 17% of the human genome, and therefore we are observing a significant difference of L1 copies between our experimental conditions, especially in primary fibroblasts.

This L1 copy number decrease can be due to less L1 integration than the wt. It can also be possible that we are observing a case of ORF2-dependent L1 excision. In literature is reported that L1 ORF2 protein endonuclease activity, usually related to L1 integration, can also be responsible for L1 copies excision events[68]. Moreover, this L1 endogenous phenotype can be a result of recombinogenic events between L1 genomic sequences.

Then, I also estimated endogenous L1 mRNA levels, and in depleted/mutated RNase H2 cells, there was a significant increase compared to the controls. It remains to be clarified whether this increase depends on a loss of transcriptional repression of L1 elements, or whether the lack of RNase H2 prevents L1 RNA integration leading to its cytosolic accumulation. An increase of L1 expression levels can be consistent with results obtained from retrotrasposition assays and with an L1 inhibitory role of RNase H2. On the other hand, if we consider the L1 CNVs analysis, it seems that RNase H2 promotes L1 integration. It would be possible that RNase H2 regulates multiple steps of the L1 life

cycle, inhibiting, at the first step, L1 transcription and at the end, regulating integration of a newly synthesized L1 copy in the genome. Indeed, TREX1 provides an example of different levels of L1 regulation: retroelements can be substrates of this exonuclease, but it is also reported that TREX1 can interact with LINE1 ORF1 protein triggering its depletion[4,15,14,20].

Regarding L1 endogenous phenotypes obtained, increased expression levels, but fewer copies integrated into the genome, it could be possible that L1 intermediates are accumulated in the cytoplasm of RNase H2 depleted cells.

In order to correlate a putative L1 intermediates accumulation with cellular production of IFNα, it would be useful to modulate L1 endogenous levels. For that reason, I tested different technical approaches to silence L1: shRNAs transduction, siRNAs transfection, FANA antisense oligonucleotides (FANA ASOs) treatment, and reverse transcription inhibitors (Lamivudine or 3TC) treatment. Unfortunately, the depletion of repetitive sequences that represent about 17% of the entire genome is quite inefficient. Moreover, in primary fibroblasts, the DNA transfection is impaired.

The most L1 silencing efficiency was obtained by transfection of a combination of siRNAs or by treatment with the reverse transcriptase inhibitor Lamivudine (3TC).

Then, I began to evaluate if L1 mRNAs levels modulation can trigger IFNα levels variations. From preliminary data obtained, in RNase H2 depleted cells, the silencing of L1 also induces decreasing in IFNα levels (Figure 7A), suggesting a direct correlation between L1 expression levels and one of the key marker of the AGS pathogenesis.

We wanted to characterise this interesting observed correlation and verify an L1-dependent nucleic acids accumulation as a cause of IFNα production. In order to do that, I evaluated by immunofluorescence, whether there are differences in the content of ssDNA among: RNase H2 silenced cells, L1 and RNase H2 silenced cells and control ones. Figure 7 shows that in RNase H2 depleted cells, there is a higher ssDNA content compared to control cells, and when I silenced L1 in those cells, I restore control conditions. Therefore, just lowering L1 expression levels, also ssDNA molecules, mainly accumulated in the cytoplasm, decrease, suggesting that some of these nucleic acids accumulated in RNase H2 depleted cells derived from L1 retroelements. I will repeat this experiment to confirm these promising preliminary data. I will also perform this kind of assay using other methods of L1 silencing or reverse transcriptase inhibitors, to try to characterize the precise step of RNase H2 involvement in L1 regulation.

Similar to what observed in RNase H2 mutated cells, also mutations in other AGS genes alter the L1 number of genomic copies and L1 expression levels. As shown in Figures 8 and 9 in ADAR1 or TREX1 silenced/mutated cells, there are fewer L1 copies integrated into the genome and higher L1

mRNA levels compared to control cells. Furthermore, high L1 mRNA levels seem to correlate with increased IFNα expression (Figure 9).

However, which DNA intermediates activate the cGAS-STING response remains to be elucidated. We observed a strict relation between higher L1 expression and ssDNA accumulation, but what kind of immunostimulatory L1 intermediates are accumulated and at what level of L1 retrotransposition impairment occurs this accumulation, is still unclear.

For that reason, we are interested in evaluating the occurrence of L1 intermediates accumulation in the cytoplasm of cells carrying AGS mutated genes.

To further study an L1-dependent nucleic acids accumulation caused by mutation/silencing of AGS related genes, I am also setting up, in parallel, an extrachromosomal DNA purification protocol. Finally, from these extracts, I obtained only cytoplasmatic DNA without genomic contaminations.

I am working on the amplification and quantification of L1 sequences content of this purified DNA from RNase H2 silenced cells. It would be interesting to sequence these extrachromosomal DNA extractions in order to can have a complete analysis of these accumulated nucleic acid species.

Consisting, it has been recently observed abundant extrachromosomal DNA in human pluripotent stem cells lacking TREX1, of which endogenous L1 was a major source[17,18].

It would be interesting to extend these kinds of L1 endogenous evaluations to the others AGS-causing genes to demonstrate that the phenotypes observed in RNase H2, ADAR1, or TREX1 depleted condition are a general characteristic of AGS patients. We could demonstrate a common dysregulation of L1 mobility that can be, at least in part, responsible for the AGS pathogenesis. For that reason, we are going to test a complete panel of all the AGS-causing genes depleted cellular lines, evaluating the effects on endogenous L1 levels.

RNase T2, a putative AGS-causing gene and its role in L1 mobility

The RNase T2 belongs to the Rh/T2/S glycoprotein family and is specific for single-stranded RNAs. Recently, biallelic loss of function in this gene has been reported in a group of infants affected by a genetic encephalopathy with cerebral calcification. It is not yet clear whether these patients belong to the cases of AGS and what is the molecular role of RNase T2 in that syndrome[53]. We want to help to evaluate if RNase T2 is the eighth AGS-causing gene. I started to verify, through retrotransposition assays, if this enzyme could be involved in the regulation of L1 mobility. Similarly to other AGS-related genes, L1 luciferase retrotransposition assays show that RNase T2 could act as an L1 inhibitor (Figure 10).

Furthermore, analysing primary fibroblasts carrying RNase T2 mutation, L1 expression increases, and similarly to other AGS genes, we observed induced INFα expression (Figure 11A).

On the other hand, not all phenotypes are similar to RNase H2 because the copy numbers of the endogenous L1 seem unaltered compared to the control (Figure 11B). These data suggest that RNase T2 could be involved in L1 regulation but differently by other AGS-genes tested. It is of our interest to establish if RNase T2 is an AGS causing gene or not and to characterize its role in L1 mobility. Further analysis is necessary, but understanding these differences could help us further understand the molecular mechanisms that are the cause of AGS.

Conclusions and future perspectives

Aicardi-Goutières syndrome is an autoimmune disorder associated with the induction of a type I interferon response driven by improper endogenous nucleic acid accumulation[6]. These nucleic acid species accumulated in the cytoplasm of AGS-mutated cells are yet to be defined[2,5].

Among all the AGS-causing genes, TREX1, ADAR1, and SAMHD1 could affect the correct metabolism of LINE1 retroelements, the most active autonomous transposable elements in humans[18,20,37,47].

Considering RNase H2 enzymatic activities, it can potentially regulate the mobility of endogenous L1 retroelements by degrading RNA:DNA hybrids formed during their replication and a possible source of immunostimulatory nucleic acids in RNase H2 deficiency condition can be endogenous retroelements intermediates. Based on this hypothesis, we decided to start to characterise a possible role of RNase H2 as an L1 retrotransposition regulator.

At the beginning of my PhD project, to test this possible RNase H2 involvement in the L1 mobility process, I used L1 retrotransposition assays[64,65,66]. From these assays, it seemed clear that in RNase H2 depleted cellular conditions, mobility of L1 increases. During the first two years of my PhD, however, other researcher groups published opposite results concerning RNase H2 role in L1 mobility[29,30].

Since it is not clear in literature if the RNase H2 enzyme can promote or inhibits L1 mobility, we decided to change our approach and start to look at RNase H2 involvement in endogenous L1 regulation. From these analyses, two phenotypes emerged, an increase of L1 expression levels and less L1 copies integrated into the genome of RNase H2 silenced/mutated cells compared to controls.

We then decided to extend these endogenous L1 evaluations to two known L1 retrotransposition inhibitors, ADAR1 and TREX1[18,20,47]. In RNase H2 or ADAR1 or TREX1 silenced cells, I observed these phenotypes:

- higher L1 and IFNα expression levels;
- less L1 copies integrated into the genome.

During my PhD, I also started to characterise a role in L1 mobility of the putative AGS-causing gene RNase T2[53]. Analysing primary fibroblasts carrying RNase T2 mutation, similarly to other AGS genes, we observed an increase of L1 and INFα expression levels (Figure 11A).

On the other hand, not all phenotypes are similar to previews described because the L1 genomic copy numbers seem unaltered compared to the control (Figure 11B). These data suggest that RNase T2 could be involved in L1 regulation but in a different way respect to RNase H2, TREX1, and ADAR1. In the contest of Aicardi–Goutières syndrome, no one has performed these kinds of endogenous L1 analysis yet. We think that it is a new approach to study L1 involvement in this pathology, more comparable to the condition of patients respect to retrotransposition assays one.

We want to repeat these kinds of L1 endogenous evaluations considering the other AGS-causing genes in order to demonstrate that these L1 phenotypes observed in RNase H2, ADAR1, or TREX1 depleted condition are a general characteristic of AGS patients.

It would be really interesting to demonstrate a common correlation between impairment of L1 regulation and activation of innate immunity in the AGS pathogenesis, especially thinking about clinical therapies.

Whether we and/or other research groups will demonstrate the existence of this common thread between the AGS causing genes, it would be also reasonable thinking about a diagnostic test of this syndrome. Indeed, it would be possible to analyse L1 endogenous levels, concurrently with AGS mutation, to establish a diagnosis before birth.

Considering both of these observed L1 phenotypes, increased expression levels, but fewer copies integrated into the genome are phenotypes consistent with an accumulation of L1 intermediate in the cytoplasm of those cells carrying AGS mutated genes. Furthermore, it would be important to demonstrate whether this accumulation of nucleic acids derived from L1, or retrotransposons in general, is directly linked to the hyperactivation of the innate immunity.

With this regard, I started to evaluate nucleic acids accumulation in RNase H2 and L1 depleted conditions. From this analysis emerges that, just lowering L1 expression levels in RNase H2 silenced cells:

- we can rescue the transcriptional activation of IFNα;
- ssDNA molecules, accumulated in the cytoplasm of RNase H2 silenced cells, decrease to normal condition.

On the basis of these results, we are on the way to demonstrate that some of these nucleic acids accumulated in RNase H2 depleted cells derived from L1 retroelements and the accumulation of these L1 intermediates can be relevant for the IFNα overproduction. Indeed, our results showed that the selective L1 silencing could have an impact on IFNα expression levels, restoring normal conditions.

I will perform this kind of assay using other methods of L1 silencing and in combination with different AGS-causing genes depleted conditions, in order to verify a direct link between L1 retroelements and the pathogenesis of AGS.

Besides to demonstrate a pathologic relevance of L1 regulation in the contest of AGS syndrome, we are also interested in explaining these observed L1 phenotypes. Starting from the preliminary data obtained so far, we propose a model in which: higher L1 expression levels concurrently with less L1 integration can generate a cytoplasmatic accumulation of DNA retrotransposition intermediates. Based on our preliminary data, these L1 cytoplasmatic intermediates, can be in the form of ssDNA molecules. However, we can't exclude that other L1 DNA species, such as dsDNA or RNA:DNA hybrids, can be accumulated. In our model, we propose that these L1 intermediates can be accumulated in the cytoplasm of AGS patients' cells by inducing the innate immunity activation through the cGAS-STING pathway (Figure 12).

I still have a lot of work to do in order to understand the molecular mechanisms behind the phenotypes observed. Despite this, we hope that our data will be useful in the comprehension of molecular processes behind the Aicardi–Goutières pathogenesis and the study of therapies.

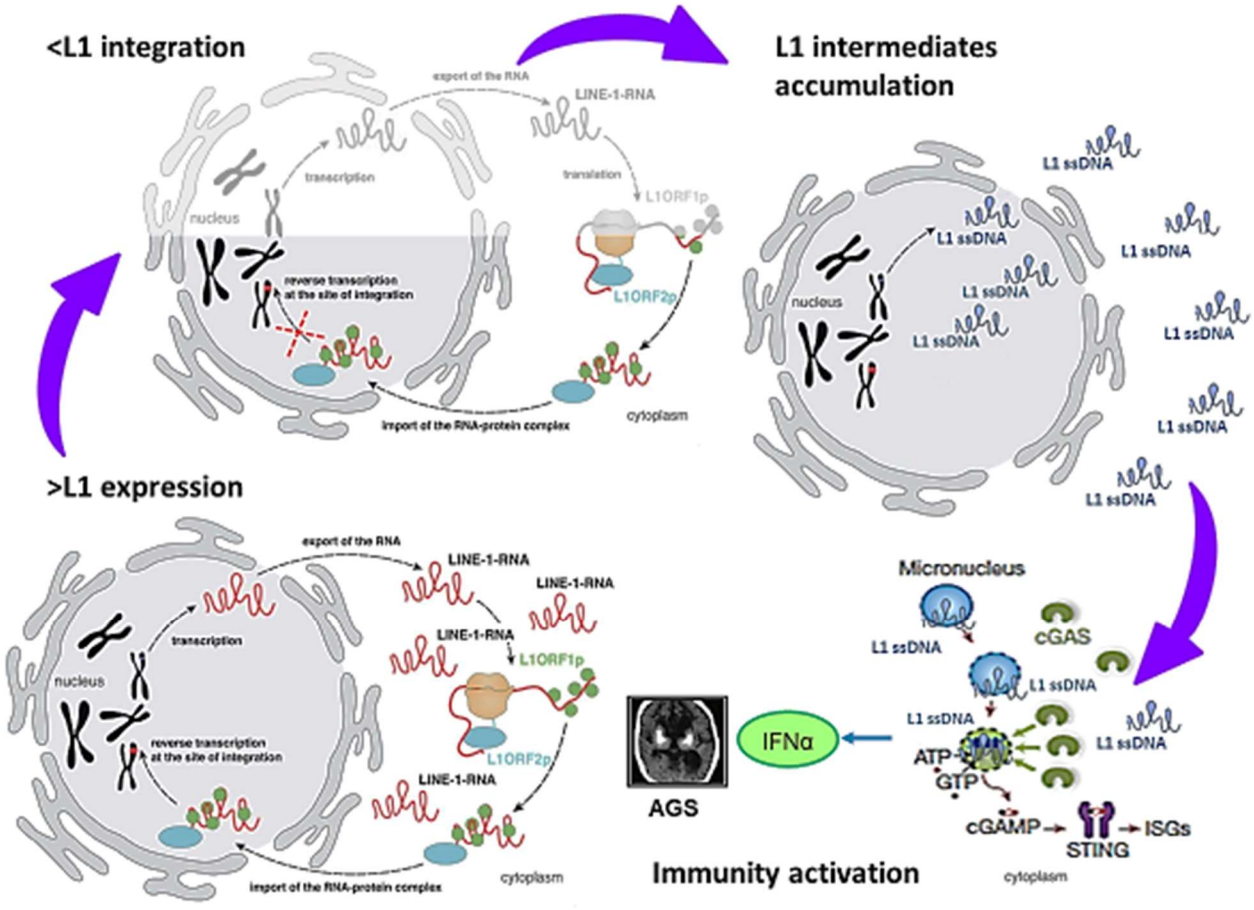


Figure 12: Model of L1 intermediates accumulation effect on AGS pathogenesis.

Materials and Methods

Cell culture and transfection

HeLa, MRC5VI and HEK293T cells were grown in Dulbecco's modified Eagle's medium (DMEM). The complete medium was supplemented with 10% FBS, penicillin, streptomycin and L-glutamine and kept at 37ºC in a humidified atmosphere with 5% of CO2. Primary fibroblasts cells derived from patients and control cells (48BR) were cultured in DMEM containing 20% FBS, penicillin, streptomycin and L-glutamine. Cells were transfected by using lipofectamineRNAiMax or lipofectamine3000 (Invitrogen) reagents according to the protocol provided by the manufacturer.

Lentiviral vectors production

shRNA sequences were cloned using EcoRI and AgeI restriction enzymes in pLKO.1-TCR cloning vector (Addgene). As control, was used SCRAMBLE shRNA cloned in pLKO.1 vector (Addgene). Lentiviral vectors were produced by transient co-transfection of pLKO.1 and packaging plasmids psPAX2 (Addgene) and pMD2.G (Addgene) in HEK293T cells. Virus was harvested at 48 hours post-transfection. Infections of cells were carried out in the presence of 10 mg/ml of polybrene. Following transduction, cells were selected adding 1 mg/ml of puromycin to the medium.

shRNA and siRNA sequences

siLuc: 5'-CGUACGCGGAAUACUUCGUU-3'
siGFP: 5'-GGCUACGUCCAGGAGCGCACCTT-3'
siRNase_H2B: 5'-GUGGAUAACGUGUUUCCAATT-3'
siL1.1: 5'-GGTATCAGCAATGGAAGA-3'
siL1.2: 5'-GGAAGATCTACCAAGCCAA-3'
shRNase H2A: 5'-CCGTTCTTCCCACCGATATTT-3'
shRNase H2B.1: 5′-GCTTCTCCACTACCTCATAAA-3′
shRNase H2B.2: 5′-ATCAAACTGTGGCAGCATTAA-3′
shADAR1: 5'-GACTGCGAAGGATAGTATATT-3'
shRNase T2: 5'-AGATCGTGGCCCTTCAATTTA-3'

L1 retrotransposition assays

For the L1 Luciferase assay, HeLa or MRC5 cells were seeded in 24-well plates, grow up to 60% of confluence and then transduced in triplicate with shRNA lentiviral constructs or alternatively, cells were transfected with siRNAs. Seventy-two hours later, cells were further transfected with the pYX014 plasmid or the pYX015 one as negative control (the last carries a L1 missense mutation that impairs retrotransposition). Four days after transfection, cells were lysed for luminescence analysis using the Dual-Luciferase® Reporter Assay System (Promega) and following the manufacturer's instruction. L1 activity was measured as the Fluc/Rluc ratio with the Ensight™ Multimode Plate Reader (PerkinElmer). In colony retrotransposition assay, HeLa or MRC5 cells were seeded and transduced as described for luciferase assay, but in this case, 72 hours post infection, cells were transfected with the pJM101/L1.3 plasmid or positive and negative controls, pcDNA3 vector or pJM105/L1.3 cassette respectively. Seventy-two hours later, cells were grown in media supplemented with G418 (600 µg/ml) and after about 14 days in G418 selection, the remaining cells were fixed with 0.4% paraformaldehyde and stained with 0.1% crystal violet to facilitate the visualization and allow the counting of the colonies formed (indicative of individual retrotransposition events). Results obtained from each experiment were normalized for transfection efficiency.

Antibodies

The following antibodies were used: anti-RNase H2A (Abcam, 1:1000), anti-RNase H2B (purified from rabbit serum from our lab, 1:500), anti-RNase H2C (Protein-Tech, 1:500), anti-Vinculin (Sigma, 1:50000). Secondary antibodies were goat anti-mouse or goat anti-rabbit conjugated to HRP (Western Blot) or to Alexa Fluor 488 or Alexa Fluor 594 (immunofluorescence).

Western Blotting

Cells were lysed in Laemmli sample buffer (60 mM Tris pH 6.8, 2% SDS, 10% glycerol, 100 mM DTT), sonicated and boiled for 5 min at 95°C. Equal amounts of whole-cell extracts were analyzed by SDS–polyacrylamide gel electrophoresis. Proteins were then transferred to Protran® nitrocellulose membrane 0.2 µm (PerkinElmer) by electroblotting or with Trans-Blot® Turbo™ System (BioRad). Non-specific antibodies binding was prevented incubating membranes in 5% (w/v) non-fat dry milk PBST 0.1% (Blocking buffer) for 1 hour at room temperature. Primary antibodies were diluted in this blocking solution and incubated overnight at 4°C. Secondary antibodies were diluted in blocking solution and incubated 1 hour at RT. HRP-conjugated secondary antibodies were detected by enhanced chemiluminescence (ECL) method, using Clarity™ Western ECL Substrate (BioRad). Chemiluminescent filters were imaged with the ChemiDocTouch™ imaging system (BioRad).

ssDNA immunofluorescence

To perform ssDNA staining, MRC5 cells grown on were fixed on ice with 4% PFA for 20 minutes and then with methanol at -20°C overnight. The next day, cells were treated with RNase A (200 µg/mL) at 37°C for 4 hours. Then, the cells were blocked with 3% BSA and incubated overnight at 4°C with the anti-ssDNA primary antibody (Millipore, MAB3299, 5µg/mL) diluted in 3% BSA. The following day, cells were incubated 1 hour at RT with secondary antibody diluted 1:1000 in PBST (0,1% Tween). Cells were mounted using ProLong Gold (Invitrogen) containing DAPI to stain nuclei. Images were acquired with a Nikon confocal A1 microscope using a ×63 oil immersion objective. All microscope settings were set to collect images below saturation and were kept constant for all images taken in one experiment. ssDNA puncta quantification for cell was performed using ImageJ.

Primer sequences

Primers used in SYBR Green qRT PCR experiments:

GAPDH_FW: TTGAGGTCAATGAAGGGGTC
GAPDH_RV: GAAGGTGAAGGTCGGAGTCA
IFNα_FW: GCCATCTCTGTCCTCCATGA
IFNα_RV: CTCTCCTCCTGCATCACACA
RNASET2_FW: ACTGGCCTGACGTAATTCAC
RNASET2_RV: CTTCTTCTGGGAGTTGAGCG
L1_5'UTR_FW: GCCAAGATGGCCGAATAGGA
L1_5'UTR_RV: AAATCACCCGTCTTCTGCGT
L1_ORF2_FW: CAAACACCGCATATTCTCACTCA
L1_ORF2_RV: CTTCCTGTGTCCATGTGATCTCA

Primers used in conjunction with TaqMan probes in experiments on genomic DNA:

L1_ORF2_FW: TGCGGAGAAATAGGAACACTTTT
L1_ORF2_RV: TGAGGAATCGCCACACTGACT
L1_5'UTR_FW: GAATGATTTTGACGAGCTGAGAGAA
L1_5'UTR_RV: GTCCTCCCGTAGCTCAGAGTAATT
SATA_FW: GGTCAATGGCAGAAAAGGAAAT
SATA_RV: CGCAGTTTGTGGGAATGATTC

qRT-PCR analysis

Total RNA was extracted from cells using the RNeasy Mini Kit (Qiagen). One microgram of RNA was used to generate cDNA using iScript™ cDNA Synthesis Kit (Biorad), and 10 ng, or 0,4 ng for L1 expression analysis, of cDNA was employed in each qRT PCR assay using the SYBR Green system (Genespin). The reactions were performed in triplicate. GAPDH gene was used as the normalization control in expression level experiments. For measure of L1 transcription, RNA samples were exhaustively digested with RNasefree DNase (Qiagen) before the synthesis of cDNA. Effectiveness of the DNase digestion was assessed in -RT PCR controls reactions using RNA dilutions instead of cDNA ones. In L1 CNVs evaluations, qRT PCR of genomic DNA was performed using the Taqman system (Life Technologies). Genomic DNA was purified with the Phenol-Chloroform method and 80 pg of gDNA was used with the previews indicated primers and TaqMan probes in L1 CNVs qRT PCR experiments. The PCR reactions were performed in at least in triplicate in expression levels analysis and at least in six replicates in L1 CNVs ones. CFX Connect Real Time System (BioRad) instrument was used for all experiments.

Bibliography

1. Rice, G. *et al.* Clinical and Molecular Phenotype of Aicardi-Goutières Syndrome. *Am. J. Hum. Genet.* **81**, 713–725 (2007).

2. Crow, Y. J. & Manel, N. Aicardi-Goutières syndrome and the type I interferonopathies. *Nat. Rev. Immunol.* **15**, 429–440 (2015).

3. Stephenson, J. B. P. Aicardi-Goutières syndrome (AGS). *Eur. J. Paediatr. Neurol.* **12**, 355–358 (2008).

4. Tolmie, J. L., Shillito, P., Hughes-Benzie, R. & Stephenson, J. B. P. The Aicardi-Goutières syndrome (familial, early onset encephalopathy with calcifications of the basal ganglia and chronic cerebrospinal fluid lymphocytosis). *J. Med. Genet.* **32**, 881–884 (1995).

5. Crowl, J. T., Gray, E. E., Pestal, K., Volkman, H. E. & Stetson, D. B. Intracellular Nucleic Acid Detection in Autoimmunity. *Annu. Rev. Immunol.* **35**, 313–336 (2017).

6. Chahwan, C. & Chahwan, R. Aicardi-Goutieres syndrome: From patients to genes and beyond. *Clin. Genet.* **81**, 413–420 (2012).

7. Crow, Y. J. *et al.* Characterization of Human Disease Phenotypes Associated with Mutations in TR E X1 , RNAS E H2A ,. *Am. J. Med. Genet. Part A* 296–312 (2015).

8. Lee-kirsch, M. A. The Type I Interferonopathies. (2017). doi:10.1146/annurev-med-050715-104506

9. Cheng, W. Y. *et al.* The cGas-sting signaling pathway is required for the innate immune response against ectromelia virus. *Front. Immunol.* **9**, (2018).

10. Ivashkiv, L. B. & Donlin, L. T. Regulation of type i interferon responses. *Nature Reviews Immunology* (2014). doi:10.1038/nri3581

11. Forster, S. Interferon signatures in immune disorders and disease. *Immunol. Cell Biol.* **90**, 520–527 (2012).

12. Crow, Y. J., Vanderver, A., Orcesi, S., Kuijpers, T. W. & Rice, G. I. Therapies in Aicardi-Goutières syndrome. *Clin. Exp. Immunol.* **175**, 1–8 (2014).

13. Kavanagh, D. *et al.* New roles for the major human 3′-5′ exonuclease TREX1 in human disease. *Cell Cycle* (2008). doi:10.4161/cc.7.12.6162

14. Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity. *Cell* **134**, 587–598 (2008).

15. Rice, G. I., Rodero, M. P. & Crow, Y. J. Human Disease Phenotypes Associated With Mutations in TREX1. *Journal of Clinical Immunology* (2015). doi:10.1007/s10875-015-0147-3

16. Bunney, P. E., Zink, A. N., Holm, A. A., Billington, C. J., & Kotz, C. M. 乳鼠心肌提取 HHS Public Access. *Physiol. Behav.* **176**, 139–148 (2017).

17. Bhoj, V. G. & Chen, Z. J. Linking Retroelements to Autoimmunity. *Cell* **134**, 569–571 (2008).

18. Thomas, C. A. *et al.* Modeling of TREX1-Dependent Autoimmune Disease using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation. *Cell Stem Cell* (2017). doi:10.1016/j.stem.2017.07.009

19. Yang, Y. G., Lindahl, T. & Barnes, D. E. Trex1 Exonuclease Degrades ssDNA to Prevent Chronic Checkpoint Activation and Autoimmune Disease. *Cell* **131**, 873–886 (2007).

20. Li, P. *et al.* Aicardi-Goutières syndrome protein TREX1 suppresses L1 and maintains genome integrity through exonuclease-independent ORF1p depletion. *Nucleic Acids Res.* **45**, 4619–4631 (2017).

21. Pizzi, S. *et al.* Reduction of hRNase H2 activity in Aicardi-Goutières syndrome cells leads to replication stress and genome instability. *Hum. Mol. Genet.* **24**, 649–658 (2015).

22. Holt, I. J. The Jekyll and Hyde character of RNase H1 and its multiple roles in mitochondrial DNA metabolism. *DNA Repair* (2019). doi:10.1016/j.dnarep.2019.06.001

23. Reijns, M. A. M. *et al.* Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* (2012). doi:10.1016/j.cell.2012.04.011

24. McElhinny, S. A. N. *et al.* Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol.* (2010). doi:10.1038/nchembio.424

25. Lazzaro, F. *et al.* RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. *Mol. Cell* (2012). doi:10.1016/j.molcel.2011.12.019

26. Hiller, B. *et al.* Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. *J. Exp. Med.* **209**, 1419–1426 (2012).

27. Mackenzie, K. J. *et al.* Ribonuclease H2 mutations induce a cGAS/STING-dependent innate immune response. *EMBO J.* **35**, 831–844 (2016).

28. Pokatayev, V. *et al.* RNase H2 catalytic core Aicardi-Goutières syndrome-Related mutant invokes cGAS-STING innate immunesensing pathway in mice. *J. Exp. Med.* (2016). doi:10.1084/jem.20151464

29. Choi, J., Hwang, S. Y. & Ahn, K. Interplay between RNASEH2 and MOV10 controls LINE-1 retrotransposition. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gkx1312

30. Benitez-Guijarro, M. *et al.* RNase H2, mutated in Aicardi-Goutières syndrome, promotes LINE-1 retrotransposition. *EMBO J.* (2018). doi:10.15252/embj.201798506

31. Beloglazova, N. *et al.* Nuclease activity of the human SAMHD1 protein implicated in the aicardi-goutières syndrome and HIV-1 restriction. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M112.431148

32. Hrecka, K. *et al.* Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* (2011). doi:10.1038/nature10195

33. Goldstone, D. C. *et al.* HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* (2011). doi:10.1038/nature10623

34. Lahouassa, H. *et al.* SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat. Immunol.* (2012). doi:10.1038/ni.2236

35. Rice, G. I. *et al.* Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. *Nat. Genet.* (2009). doi:10.1038/ng.373

36. Maelfait, J., Bridgeman, A., Benlahrech, A., Cursi, C. & Rehwinkel, J. Restriction by SAMHD1 Limits cGAS/STING-Dependent Innate and Adaptive Immune Responses to HIV-1. *Cell Rep.* (2016). doi:10.1016/j.celrep.2016.07.002

37. A., H. *et al.* The SAMHD1-mediated block of LINE-1 retroelements is regulated by phosphorylation. *Mob. DNA* **9**, 11 (2018).

38. Samhd, A. S. *et al.* NIH Public Access. **4**, 1108–1115 (2014).

39. Coquel, F., Neumayer, C., Lin, Y. L. & Pasero, P. SAMHD1 and the innate immune response to cytosolic DNA during DNA replication. *Current Opinion in Immunology* (2019). doi:10.1016/j.coi.2018.09.017

40. Coquel, F. *et al.* SAMHD1 acts at stalled replication forks to prevent interferon induction. *Nature* (2018). doi:10.1038/s41586-018-0050-1

41. Bass, B. L. RNA Editing by Adenosine Deaminases That Act on RNA. *Annu. Rev. Biochem.* (2002). doi:10.1146/annurev.biochem.71.110601.135501

42. Rice, G. I. *et al.* Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type i interferon signature. *Nat. Genet.* **44**, 1243–1248 (2012).

43. Patterson, J. B. & Samuel, C. E. Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Mol. Cell. Biol.* (1995). doi:10.1128/mcb.15.10.5376

44. Hartner, J. C., Walkley, C. R., Lu, J. & Orkin, S. H. ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. *Nat. Immunol.* (2009). doi:10.1038/ni.1680

45. Pestal, K. *et al.* Isoforms of RNA-Editing Enzyme ADAR1 Independently Control Nucleic Acid Sensor MDA5-Driven Autoimmunity and Multi-organ Development. *Immunity* (2015). doi:10.1016/j.immuni.2015.11.001

46. Liddicoat, B. J. *et al.* RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science (80-.).* (2015). doi:10.1126/science.aac7049

47. Orecchini, E. *et al.* ADAR1 restricts LINE-1 retrotransposition. *Nucleic Acids Res.* **45**, 155–168 (2017).

48. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **35**, 41–48 (2003).

49. Rice, G. I. *et al.* Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type i interferon signaling. *Nat. Genet.* (2014). doi:10.1038/ng.2933

50. Acquati, F. *et al.* Loss of function of Ribonuclease T2, an ancient and phylogenetically conserved RNase, plays a crucial role in ovarian tumorigenesis. *Proc. Natl. Acad. Sci. U. S. A.* (2013). doi:10.1073/pnas.1222079110

51. Thorn, A. *et al.* Structure and activity of the only human RNase T2. *Nucleic Acids Res.*

(2012). doi:10.1093/nar/gks614

52. Henneke, M. *et al.* RNASET2-deficient cystic leukoencephalopathy resembles congenital cytomegalovirus brain infection. *Nat. Genet.* (2009). doi:10.1038/ng.398

53. Tonduti, D. *et al.* Encephalopathies with intracranial calcification in children: Clinical and genetic characterization. *Orphanet J. Rare Dis.* (2018). doi:10.1186/s13023-018-0854-y

54. Ariumi, Y. Guardian of the human genome: Host defense mechanisms against LINE-1 retrotransposition. *Frontiers in Chemistry* (2016). doi:10.3389/fchem.2016.00028

55. Goodier, J. L. Restricting retrotransposons: A review. *Mobile DNA* (2016). doi:10.1186/s13100-016-0070-z

56. Feschotte, C. & Pritham, E. J. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu. Rev. Genet.* (2007). doi:10.1146/annurev.genet.40.110405.090448

57. Volkman, H. E. & Stetson, D. B. The enemy within: Endogenous retroelements and autoimmune disease. *Nature Immunology* (2014). doi:10.1038/ni.2872

58. Luning Prak, E. T. & Kazazian, H. H. Mobile elements and the human genome. *Nature Reviews Genetics* (2000). doi:10.1038/35038572

59. Hancks, D. C. & Kazazian, H. H. Active human retrotransposons: Variation and disease. *Current Opinion in Genetics and Development* (2012). doi:10.1016/j.gde.2012.02.006

60. Viollet, S., Monot, C. & Cristofari, G. L1 retrotransposition: The snap-velcro model and its consequences. *Mob. Genet. Elements* (2014). doi:10.4161/mge.28907

61. Suarez, N. A., Macia, A. & Muotri, A. R. LINE-1 retrotransposons in healthy and diseased human brain. *Developmental Neurobiology* (2018). doi:10.1002/dneu.22567

62. Goodier, J. L. & Kazazian, H. H. Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites. *Cell* (2008). doi:10.1016/j.cell.2008.09.022

63. Khalid, M. *et al.* Line-1: Implications in the etiology of cancer, clinical applications, and pharmacologic targets. *Mutation Research - Reviews in Mutation Research* (2018). doi:10.1016/j.mrrev.2018.09.003

64. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).

65. Doucet, A. J. *et al.* Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet.* (2010). doi:10.1371/journal.pgen.1001150

66. Xie, Y., Rosser, J. M., Thompson, T. L., Boeke, J. D. & An, W. Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res.* (2011). doi:10.1093/nar/gkq1076

67. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* (2009). doi:10.1038/nature08248

68. Erwin, J. A. *et al.* L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat. Neurosci.* (2016). doi:10.1038/nn.4388