

SISSA

PhD course in Functional and Structural Genomics

Neuroscience Area

PhD thesis

Intron retention dysregulation in Aicardi-Goutières syndrome

Thesis submitted for the degree of *"Philosophiae Doctor"*

Student:

Nicolò Gualandi

Supervisor:

Prof. Remo Sanges

Academic year 2021-2022

Table of Contents

1 ABSTRACT....1

2 INTRODUCTION....3

2.1 Aicardi-Goutières syndrome: a molecular overview....3

2.2 AGS shares molecular similarities with systemic lupus erythematosus....4

2.3 STING-associated vasculopathy with onset in infancy (SAVI) and Type I interferon-mediated autoinflammation due to DNase II deficiency syndrome (IMD) are characterized by aberrant IFN response similarly to AGS....6

2.4 Endogenous nucleic acid sensing and cGAS-STING pathway....7

2.5 AGS causative genes....9

2.5.1 TREX1....9

2.5.2 RNASEH2 A-B-C....10

2.5.3 ADAR1....11

2.5.4 IFIH1....11

2.5.5 SAMHD1....11

2.6 DNA damage and DNA damage response....12

2.6.1 DNA damage response mechanisms....14

2.6.1.1 Nucleotide excision repair (NER)....14

2.6.1.2 Mismatch repair (MMR)....14

2.6.1.3 Base excision repair (BER)....15

2.6.1.4 Non-homologous end joining (NHEJ)....15

2.6.1.5 Homologous recombination (HR)....16

2.7 DNA damage and DNA damage response are involved in the generation of extrachromosomal DNA in AGS....17

2.8 Transposable elements....19

2.8.1 Class I elements (Retrotransposons)....20

2.8.1.1 LTR retrotransposons....21

2.8.1.2 non-LTR retrotransposons....21

2.8.1.2.1 LINE....22

2.8.1.2.2 SINE....22

2.8.2 Class II....23

2.8.3 Transposable elements and host interaction: control of retrotransposition by host cells factors....26

2.9 TEs activity is associated with cytosolic DNA generation, DNA damage and IFN response....30

2.10 mRNA Splicing....33

2.10.1 Alternative splicing....35

2.10.1.1 Intron retention: mechanisms and outcomes....37

2.10.1.2 Intron retention: regulation....40

2.10.1.3 Intron retention is involved in different physiological processes....43

2.10.1.4 Intron retention in neurodegenerative and autoimmune disease....45

2.10.2 Principal challenges in detecting IR from RNA-seq data....48

3 AIM OF THE STUDY....50

4 MATERIALS AND METHODS....51

4.1 Data collection....51

4.2 Building reference genome and mapping....52

4.3 Reads per gene quantification....53

4.4 Differential gene expression analysis....53
4.5 IFN score....53
4.6 PCA analysis....54
4.7 Alternative splicing quantification and differential alternative splicing analysis....54
4.8 Intron retention quantification and IRscore quantification....54
4.9 Identification of differentially retained introns....55
4.10 Enrichment analysis....55
4.11 Correlation analysis....55
4.12 IR features and RNA binding motifs sites enrichements....56
4.13 Proliferation score....56
4.14 Bulk RNA-seq cell-type enrichment....56

5 RESULTS AND DISCUSSION....58

5.1 AGS, SAVI, IMD and SLE samples are characterized by an increased IFN signaling...58
5.1.1 AGS subjects are characterized by aberrant IFN gene expression that is influenced by the administration of RT inhibitors....58
5.1.2 AGS subjects are characterized by gene expression dysregulation that recapitulates the major symptoms of the disease....60
5.1.3 AGS, SAVI and IMD are accomunated by a positive IFN signature....63
5.1.4 SLE B-cells subtypes samples are characterized by increased IFN and ISGs expression....66
5.1.5 IFN signaling is mainly supported by astrocytes in AGS neuronal samples....67
5.2 Alternative splicing is impaired in AGS samples....72
5.2.1 AGS and SLE samples are characterized by increased IR which is not directly associated with IFN signaling....75
5.2.2 Inspection of differential IR in AGS and SLE....78
5.2.3 Genes containing differentially retained introns belong to specific pathways....83
5.2.4 IR may be associated with changes in gene expression of hosting genes in AGS whole blood samples....87
5.2.5 Retained introns are characterized by specific features....89
5.2.6 IR may be modulated by specific RNA binding protein....93
5.2.7 AGS whole blood samples are characterized by expansion of immune system cells associated with increased IR....96
5.2.8 IR may be associated with the proliferation / differentiation decision....98

6 CONCLUSIONS....101

7 BIBLIOGRAPHY....104

List of Figures

Figure 1: AGS causative genes are involved in nucleic acid processing.....8
Figure 2: AGS causative genes impact TEs activity at different stages.....12
Figure 3: Different types of DNA damage are repaired by specific pathways.....17
Figure 4: DNA damage and repair is a source of extrachromosomal DNA.....19
Figure 5: Transposable elements classification.....24
Figure 6: Transposable elements replication mechanisms.....25
Figure 7: LINE1 structure and retrotransposition mechanism.....26
Figure 8: LINE1 activity influences the genome in different ways.....29
Figure 9: Transposable elements have both harmful and beneficial roles.....30
Figure 10: LINE1 sequences made up a consistent fraction of extrachromosomal DNA in TREX1 mutated cells.....33
Figure 11: Splicing is a tightly regulated process.....35
Figure 12: Alternative splicing events. Different types of alternative splicing events and their effect on transcripts sequences.....37
Figure 13: Intron retention participates in different biological processes.....40
Figure 14: Intron retention is regulated by different genomic features.....42
Figure 15: *xcell* 64 human cell type compendium.....57
Figure 16: AGS samples are characterized by aberrant IFN response that is influenced by treatment with RT inhibitors.....60
Figure 17: AGS samples are characterized by gene expression alterations reflecting the phenotypic hallmark of the disease.....63
Figure 18: AGS, SAVI and IMD are accomunated by a positive IFN signature.....65
Figure 19: SLE subjects are characterized by increased IFN signaling.....67
Figure 20: AGS gene expression alterations are cell-type specific.....70
Figure 21: AGS gene expression alterations are cell-type specific.....71
Figure 22: Alternative splicing is impaired in AGS samples and IR changes are the most abundant among all AS events.....74
Figure 23: AGS and SLE samples are characterized by increased IR that is not rescued by RT inhibitors treatment.....77
Figure 24: AGS whole blood samples are characterized by differential IR.....80
Figure 25: Example of differential IR between controls and AGS samples.....81
Figure 26: AGS neuronal samples and SLE DN cells are characterized by differential intron retention.....82
Figure 27: Overlap between genes affected by significantly increased IR in all the dataset analyzed.....83
Figure 28: Genes undergoing increased IR are enriched in RNA metabolism, IFN signaling and Neuronal development pathways.....86
Figure 29: IR may be associated with changes in transcript abundance.....89
Figure 30: Retained introns are characterized by specific features.....92
Figure 31: IR may be modulated by specific RNA binding proteins.....96
Figure 32: Increased IR is associated with increased B-cell content in AGS whole blood samples.....98
Figure 33: Increased IR is associated with reduced cell proliferation.....100

List of Tables

Table 1: Summary of AGS causative genes.....4
Table 2: Summary of AGS, SAVI and IMD causative genes with complete gene name. 7
Table 3: Summary of genes involved in DNA damage sensing and repair.....13
Table 4: Summary of dataset used.....52
Table 5: Number of different splice sites sequences in differentially retained introns in AGS whole blood samples associated with downregulated, non-deregulated or upregulated genes.....92

Appendices

Table 1: Summary of genes used for computing the interferon signature
Table 2: Summary of genes used for computing the proliferation signature

List of Abbreviations and Symbols

AGS	Aicardi-Goutières syndrome
CFS	Cerebrospinal fluid
TEs	Transposable elements
dNTPs	deoxynucleoside triphosphates
DDR	DNA damage response
SLE	Systemic lupus erythematosus
IFN	Interferon
ISGs	Interferon stimulated genes
DBS	Double-strand break
MMR	Mismatch repair
BER	Base excision repair
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
HR	Homologous recombination
RT	Reverse transcriptase

1 Abstract

Aicardi Goutières syndrome is an autosomal recessive, severe and progressive inflammatory disorder, characterized by defects in the central nervous system and chronic cerebrospinal fluid lymphocytosis with elevated expression of interferon responsive gene representing the strong autoimmune component of the disease. Nowadays the precise mechanism at the basis of the onset and progression of Aicardi Goutières syndrome is still partly unknown and more efforts are needed to shed light on the different altered pathways

Intron retention is an alternative splicing mechanism where an intron, instead of being removed by the splicing machinery, is maintained in the mature transcript. Intron retention has been overlooked so far but it is now gaining attention due to its newly recognized involvement in immune response and neuronal development. In fact, many recent findings have identified an unappreciated relation between intron retention dysregulation and the onset of neurodegenerative and autoimmune diseases.

In this scenario I have decided to investigate the possible involvement of intron retention in Aicardi Goutières syndrome. Moreover, I have extended this study to another relevant autoimmune disease, systemic lupus erythematosus, which share a certain degree of similarity with Aicardi Goutières syndrome. I have discovered that Aicardi Goutières syndrome whole blood samples are characterized by increased intron retention and this can be further associated with the aberrant activation of immune system cells, in particular B-cells. The results obtained have been further extended to FACS isolated systemic lupus erythematosus B-cells subtypes thus confirming the possible involvement of intron retention in the onset and progression of Aicardi Goutières syndrome related autoimmune diseases. Generally, intron retention has been associated with a reduced cell proliferation at the expense of a more differentiated and activated cell state. Intron retention can thus act as a mechanims capable of sustaining the aberrant immune system response observed in Aicardi Goutières syndrome and systemic lupus erythematosus. Finally, the analysis of Aicardi Goutières syndrome neuronal samples have revealed that aberrant intron retention can be also observed in other cell types. This evidence further extend the possibility that

disregulation of intron retention can be involved in neural cell differentiation in Aicardi Goutières syndrome.

This study identifies a novel and yet undescribed phenomenon in Aicardi Goutières syndrome and systemic lupus erythematosus. Although more in-depth investigation are needed, I propose that intron retention could be used as a marker for disease identification and a target that can lead to the development of new treatments for reducing the symptom of these diseases.

2 Introduction

2.1 Aicardi-Goutières syndrome: a molecular overview

Aicardi-Goutières syndrome (AGS) is a rare genetic, early onset, neurodevelopmental disorder characterized by increased expression of interferon stimulated genes (ISGs) (Yanick J. Crow 1993). AGS was firstly described in 1984 when Jean Aicardi and Françoise Goutières reported eight children from five families with an early-onset encephalopathy characterized by basal-ganglia calcification, white-matter abnormalities, and a chronic Cerebrospinal fluid (CSF) lymphocytosis (Aicardi and Goutières 1984). AGS usually results in a severe and irreversible mental retardation and developmental regression. AGS is usually associated with high morbidity and mortality and currently, there is no known cure for AGS and available treatments are aimed at symptom reduction (Y J Crow et al. 2014). AGS is a genetically heterogeneous disorder frequently inherited in an autosomal recessive manner and, up to date, is reported to be mainly caused by mutation in 5 different genes: ADAR1, SAMHD1, RNASEH (complex RNASEH2 A-B and C), IFIH1 and TREX1 (Table 1) (Yanick J. Crow 1993). Interestingly these genes are involved in nucleic-acids processing and a failure in these mechanisms can result in inappropriate activation of the innate immune system, mimicking a congenital viral infection (Xiao 2009). Besides this, the absence of pathogenic infections in AGS subjects leads to the hypothesis that ISGs induction could be triggered by the accumulation of endogenous extrachromosomal nucleic acids in the nucleus and/or the cytoplasm wich trigger the activation of the innate immune system (Yanick J. Crow and Rehwinkel 2009). Currently it is believed that these nucleic acids could arise as a result of either chronic DNA damage, unprocessed nucleic acid species or processing of transposable elements-derived nucleic acids. However, the source and the consequences of extrachromosomal DNA accumulation in AGS are only partially understood. In this introduction I will describe the DNA damage pathways and the transposable elements (TEs) activity before placing them in the context of AGS. I will then describe the AGS causative genes with particular attention to their interaction with both DNA damage pathways and TEs activity. Finally, I will introduce the alternative splicing (AS)

process with a special focus on intron retention (IR), nowdays emerging as a fundamental factor involved in the onset and progression of different neurodegenerative and autoimmune diseases.

Gene name	Inheritance	Human phenotypes	Protein function
TREX1	Autosomal recessive or autosomal dominant	AGS	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	Autosomal recessive or autosomal dominant	AGS, DSH, BSN, spastic paraparesis and CNP	Hydrolytic deamination of adenosine to inosine in dsRNA
IFIH1	Autosomal dominant	Various neuroimmunological and non-neurological phenotypes, including AGS, spastic paraparesis, CNP and SMS	Cytosolic sensor of dsRNA

Table 1: Summary of AGS causative genes. The table reports the list of genes involved in the onset of AGS along with a brief description and the inheritance type. AGS, Aicardi-Goutières syndrome; BSN, bilateral striatal necrosis; CLL, chronic lymphocytic leukaemia; CNP, complete non-penetrance; SMS, Singleton–Merten syndrome; DSH, dyschromatosis symmetrica hereditaria; FCL, familial chilblain lupus (adapted from (Yanick J. Crow and Manel 2015)).

2.2 AGS shares molecular similarities with systemic lupus erythematosus

Autoimmunity is defined as the incapability of the immune system to distinguish between self and non-self (Rosenblum, Remedios, and Abbas 2015). Systemic lupus erythematosus (SLE) is an autoimmune disease mainly characterized by the loss of immune self-tolerance leading to the over-production of self-antibodies. These have been found in many different parts of the body like skin, hematopoietic, joints and kidneys but also in the central nervous system (Kaul et al. 2016; de Vries et al. 2010). SLE is characterized by the production of large quantities of antibodies directed against ubiquitous self-antigens, particularly double-stranded DNA, and by a high

interferon (IFN) production (Goulielmos et al. 2018). In fact, like AGS, SLE is associated with an increased type I IFN signaling (Yanick J. Crow and Rehwinkel 2009). Moreover, mutations in TREX1, IFIH1 and SAMHD1 genes have been associated also with SLE (M A Lee-Kirsch, Wolf, and Günther 2014; Justiz Vaillant, Goyal, and Varacallo 2022; Fredi et al. 2015; Min Ae Lee-Kirsch et al. 2007). In particular heterozygous mutations in TREX1 have been associated with a cutaneous subtype of SLE and heterozygous mutations in the TREX1 gene are considered the most common cause of monogenic SLE identified to date (Yanick J. Crow and Rehwinkel 2009) while homozygous mutations of TREX1 are instead strongly associated to AGS. SLE is characterized by both genetic and phenotypic heterogeneity that makes the disease difficult to study and characterize. In this scenario, significant insight into SLE molecular mechanisms have been achieved from the study of monogenic diseases. AGS has been reported to share many common molecular similarities with SLE in particular for the autoimmune component of the disease. For example studies on AGS samples have paved the way for the identification of novel pathways of intracellular nucleic acids metabolism in SLE. Those findings have indeed revealed that the induction of autoimmunity can be initiated, also in SLE subjects, by immune recognition of endogenous nucleic acids originating from processes such as DNA replication, DNA repair or endogenous retroelements activity (M A Lee-Kirsch, Wolf, and Günther 2014). The final outcome of both syndromes is the aberrant IFN signature defined as a continued production of different IFNs that over-stimulates the immune response. In fact type I IFNs (IFN-α, IFN-β) have a pronounced immune-stimulatory effects that promote the loss of B-cell and T-cell tolerance, dendritic cell activation and autoantibody production (Shlomchik, Craft, and Mamula 2001). This in turn can cause cutaneous chilblain lesions, arthritis, anti-nuclear antibodies, reduced complement and haematological abnormalities that can be observed in both diseases (Lebon et al. 1988; Tolmie et al. 1995). These evidence suggest probable roles for AGS causative genes in the onset and progression of SLE confirming that common pathways, involved in the processing of nucleic acids that are produced during cell metabolism, can cause an aberrant immune system activation leading to both AGS and SLE.

2.3 STING-associated vasculopathy with onset in infancy (SAVI) and Type I interferon-mediated autoinflammation due to DNase II deficiency syndrome (IMD) are characterized by aberrant IFN response similarly to AGS

Autoinflammatory disorders are characterized by the activation of the immune system that leads to the generation of a continuous inflammatory state without the presence of detectable infections (Nigrovic, Lee, and Hoffman 2020). The discovery that the aberrant activation of the IFN signaling can lead to the onset of many different diseases has paved the way to an intense research for phenothypes sharing a common aberrant IFN response. Among them, SAVI and IMD share several commonalities with AGS and SLE (H. Kim, Montealegre Sanchez, and Goldbach-Mansky 2016). SAVI is a rare autoinflammatory disease characterized by systemic inflammation, interstitial lung disease and severe cutaneous vasculopathy (Y. Wang, Wang, and Zhang 2021). SAVI is caused by gain-of-functions mutations in the STING1 gene that cause the continuous activation of the IFN pathways resulting in continuous IFN and ISG signaling leading to systemic inflammation (figure 1) (Y. Wang, Wang, and Zhang 2021). Even though both are characterized by aberrant IFN signaling, AGS and SAVI differ in the fact that AGS causative mutations are capable of impair nucleic acid metabolism while SAVI mutations directly activate the IFN pathways (H. Kim, Montealegre Sanchez, and Goldbach-Mansky 2016). IMD is characterized by loss of function mutations in the DNASE2 associated with an aberrant IFN signaling (Rodero et al. 2017). DNASE2 is a lysosomal protein involved in the degradation of nucleic acid, mainly arising from apoptotic cells and nuclei expelled from erythroid precursors. The consequent accumulation of DNA in lysosome have been linked with the activation of the IFN pathway observed in IMD subjects. This leads to the insurgence of different clinical symptoms like membrano-proliferative glomerulonephritis, liver fibrosis and severe neonatal anemia (Rodero et al. 2017).

Gene name	Gene symbol	Disease	Protein function
Three Prime Repair Exonuclease 1	TREX1	AGS	3′–5′ DNA exonuclease acting by degrading extrachromosomal DNA
Ribonuclease H2 Subunit A	RNASEH2A	AGS	A catalytic component of the RNase H2 complex that acts on the RNA portion of RNA–DNA hybrids by degrading RNA in RNA-DNA hybrids. May act by reducing the amount of extrachromosomal DNA generated by reverse transcriptase activity.
Ribonuclease H2 Subunit B	RNASEH2B	AGS	A non-catalytic component of the RNase H2 complex
Ribonuclease H2 Subunit C	RNASEH2C	AGS	A non-catalytic component of the RNase H2 complex
SAM And HD Domain Containing Deoxynucleoside Triphosphate Triphosphohydrolase 1	SAMHD1	AGS	dNTP triphosphohydrolase triphosphatase and ribonuclease activity. May help to reduce the pool of avaiable dNTPs and thus reducing the amount of extrachromosomal DNA
Double-Stranded RNA-Specific Adenosine Deaminase	ADAR1	AGS	Hydrolytic deamination of adenosine to inosine in dsRNA. Can help reducing the immunostimulatory effect of dsRNA.
Interferon Induced with Helicase C Domain 1	IFIH1	AGS	Cytosolic sensor of dsRNA. Gain of function mutations cause the continuous activation of the IFN signaling pathway.
Stimulator of Interferon Response cGAMP Interactor 1	STING1	SAVI	Stimulates the IFN response upon activation by cGAMP. Gain of function mutations leads to continuous activation of IFN signaling.
Deoxyribonuclease 2	DNASE2	IMD	Hydrolyzes DNA under acidic conditions and mediates the breakdown of DNA during erythropoiesis and apoptosis. Mutations in this protein have been associated with accumulation of unprocessed DNA.

Table 2: Summary of AGS, SAVI and IMD causative genes with complete gene name. The table reports the list of genes involved in the onset of AGS, SAVI and IMD along with the full gene name and the encoded protein function in the context of DNA/RNA processing.

2.4 Endogenous nucleic acid sensing and cGAS-STING pathway

The cytosolic accumulation of self DNA, observed in AGS and SLE subjects, activates the cGAS/STING-mediated type I IFN response that leads to systemic inflammation which is at the basis of both diseases (Yan 2017). cGAS is a mainly cytoplasmic protein with the capability of recognize and bind DNA (Hopfner and Hornung 2020). cGAS protein is not capable to distinguish between self and non-self DNA. To avoid sensing of endogenoud DNA from cGAS the cell relies upon spatial separation and the activity of host DNAses. Indeed, on one hand, host DNA is mostly found in mitochondria and nucleus and this compartmentalization provides the basic mechanism to avoid self-DNA recognition by cGAS. On the other hand, the capability of TREX1 and other exonucleases to degrade DNA in the cytoplasm ensures that self-DNA debris, which might end up in the cytoplasm, are quickly degraded before they are sensed by cGAS (Yu and Liu 2021). The DNA sensor cGAS is activated by DNA

through direct binding which triggers conformational changes that led to the conversion of GTP and ATP into cGAMP. cGAMP acts as an endogenous second messenger able to induce the translocation of STING, a transmembrane protein localized in the endoplasmic reticulum (ER), from the ER to the Golgi apparatus. Following translocation to the Golgi, STING interacts with TBK1 that activates the transcription factor IRF3 through phosphorylation. Activated IRF3 dimerizes and translocates to the nucleus to regulate the transcription of IFN (Figure 1) (Motwani, Pesiridis, and Fitzgerald 2019). Type I IFN is then secreted in the extracellular matrix to work as an autocrine, paracrine or systemic messenger. The secreted type I IFN is able to bind to the IFNAR family of extracellular receptors and to trigger the expression of more than 300 ISGs that contributes to activate the innate immunity responses (A. J. Lee and Ashkar 2018; Schneider, Chevillotte, and Rice 2014).

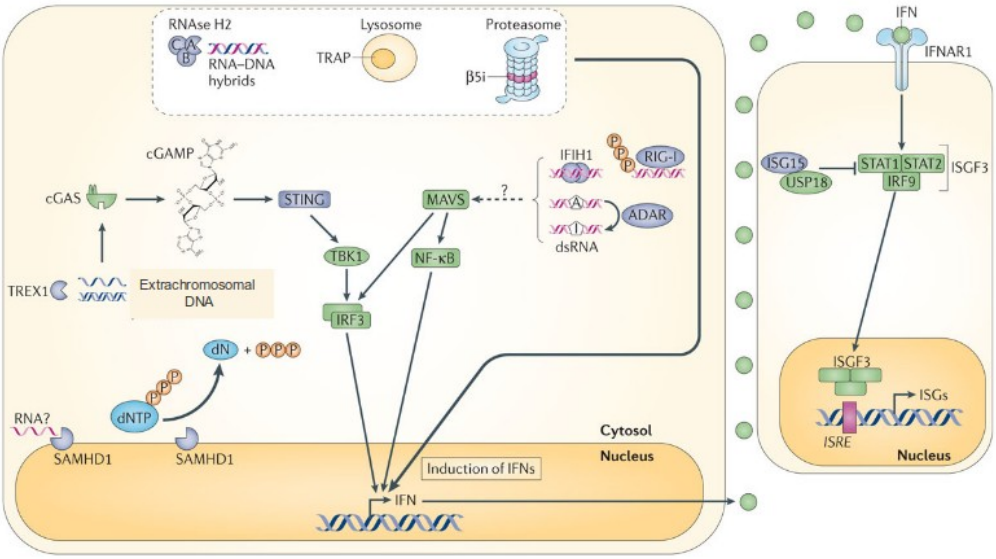


Figure 1: AGS causative genes are involved in nucleic acid processing. AGS causative genes are involved in the processing of nucleic-acids and a failure in those mechanisms can cause the cytosolic accumulation of extrachromosomal DNA. The extrachromosomal DNA, in the cytosol, is recognized by the cGAS protein that, through the STING-TBK1-IRF3 cascade, causes the synthesis of IFNs. IFNs work both as an autocrine, paracrine or systemic messenger inducing the transcription of ISGs. Interferon (IFN); Interferon regulatory factor 3 (IRF3); Interferon regulatory factor 9 (IRF9); Serine/Threonine-Protein Kinase (TBK1); Nuclear factor kappa B (NF-κB); Mitochondrial antiviral-signaling protein (MAVS); Interferon stimulated gene (ISG); Interferon sensitive response element (ISRE) adapted from (Yanick J. Crow and Manel 2015).

2.5 AGS causative genes

AGS onset has been associated with mutations in 7 different genes. Here the functions of AGS causative genes will be briefly introduced and their main functions in AGS context briefly explained.

2.5.1 TREX1

The three prime repair exonuclease 1 (TREX1) gene is composed by a single open reading frame located on chromosome 3p21.31 that encodes for the major 3'-5' cellular exonuclease which can digest both ssDNA and, with less affinity, dsDNA (Simpson et al. 2020). The TREX1 protein is composed by an N-terminal domain with exonuclease activity and a leucine-rich C-terminal domain required for its association with the endoplasmic reticulum (Lindahl et al. 2009a). TREX1 is able to prevent the accumulation of extrachromosomal DNA and in turn type I IFN activation. Moreover, its activity has been also linked to different pathways including TEs retrotransposition control and DNA damage response (P. Li et al. 2017). In 2007 Yang and colleagues discovered that TREX1 deficient cells accumulate abnormal ssDNA species leading to chronic activation of the ataxia telangiectasia mutated (ATM) kinase-dependent DNA damage checkpoint and defective G1/S transitions (Yang, Lindahl, and Barnes 2007). Furthermore, it has been observed that in normal S-phase and in response to genotoxic stress, TREX1, at least partly, redistributes to the cell nucleus in which it is able to metabolize ssDNA species arising from DNA replication (Lindahl et al. 2009a; Yang, Lindahl, and Barnes 2007). Taken together these evidence support the idea that TREX1 safeguards the cells from the accumulation of ssDNA species arising during DNA replication and suggest a possible and crucial involvement of this gene in the cell cycle progression. Another source of extrachromosomal DNA metabolized by TREX1 is produced during the DNA damage and DNA repair pathways. In this context DNAs byproducts arose during DNA repair are degraded by TREX1 preventing their accumulation and the consequent IFN response (Demaria and Vanpouille-Box 2017). These findings suggest that TREX1 is capable of degrading ssDNA generated from the processing of aberrant DNA replication intermediates and DNA repair, therefore attenuating DNA damage checkpoint signalling and preventing pathological immune activation. Recently, it has been demonstrated an enrichment of retrotransposons

sequences, mainly derived by retrotransposons belonging to the LINE1 family, through direct sequencing of cytosolic DNA from TREX1$^{-/-}$ cells (figure 10) (Thomas et al. 2017). This result candidates LINE1 retrotranscription activity as one of the main sources of accumulated extrachromosomal DNA in cells lacking TREX1 activity. Moreover it has been reported that TREX1 is capable, through an exonuclease-independent mechanism, to trigger LINE1 ORF1p depletion and, in turn, to prevent the retrotransposon-induced DNA damage response (figure 2) (P. Li et al. 2017). All these findings suggest that TREX1 can prevent the cell-intrinsic initiation of autoimmunity through the degradation of extrachromosomal DNA and the prevention of its generation. TREX1 mutations are thus associated with extrachromosomal DNA accumulation and aberrant immune response, as observed in AGS and SLE subjects.

2.5.2 RNASEH2 A-B-C

RNase H2 gene family encode respectively for 3 proteins called RNaseH2A, RnaseH2B and RnaseH2C. These 3 proteins form the major heterotrimeric ribonuclease H complex in mammalian cells (K. Kojima et al. 2018). The protein complex has endonuclease activity and is mainly involved in processing RNA molecules in RNA:DNA hybrids. The protein complex is mainly involved in Okazaki fragments processing, retroelements mobilization and genome integrity maintenance (Shaban et al. 2010). It has been demonstrated that the RNaseH2 complex plays an essential role in the removal of ribonucleotides that are misincorporated into genomic DNA during cell replication thus preventing the accumulation of detrimental mutations (Kind et al. 2014). RNASEH mutated cells have impaired cell cycle progression associated with chronic activation of post-replication repair mechanisms (Pizzi et al. 2015). Interestingly the RNaseH2 complex is able to prevent the formation of functional RNA–DNA hybrids generated during LINE1 retrotranscription, and specifically interfere with the retrotransposition mechanism (figure 2) (Genovesio et al. 2011; Benitez-Guijarro et al. 2018; Lim et al. 2015; J. Choi, Hwang, and Ahn 2018).

2.5.3 ADAR1

ADAR1 is an RNA editing enzyme that is able to convert Adenosine to Inosine through hydrolytic deamination in dsRNA and RNA:DNA hybrids (Lamers, van den Hoogen, and Haagmans 2019; Bass 2002). This modification has been linked both to changes in protein sequences, as inosine is recognized as guanine by ribosomes and in RNA secondary structures stability as I-U basepairs is less stable than A-U (Bass 2002). ADAR1 has the capability of editing dsRNA derived by Alu (SINEs) and LINE1 reducing their immuno-stimulatory effect and reducing the type I IFN response (figure 2) (Mannion et al. 2014; Orecchini, Frassinelli, and Michienzi 2017). This indicates a possible role for ADAR1 as a suppressor of type I IFN signaling (Rice et al. 2012). Moreover, it has been demonstrated that ADAR1 is able to inhibit LINE1 retrotransposition probably at a post-transcriptional level with different mechanisms (Orecchini, Frassinelli, and Michienzi 2017).

2.5.4 IFIH1

The interferon-induced with helicase C domain 1 (IFIH1) acts as a cytoplasmic sensor of viral nucleic acids and plays a major role in sensing viral infection and in the activation of a cascade of antiviral responses including the induction of type I IFNs and proinflammatory cytokines (Rice et al. 2014). In AGS IFIH1 protein is implicated in the aberrant sensing of nucleic acids that cause an hyperactivation of innate immunity. AGS-causing mutations in the IFIH1 are the only known gain-of-function mutations suggesting that these mutants bind dsRNAs more avidly and tightly than wild-type (Oda et al. 2014). Nowadays the nature of dsRNA ligands of this protein is still unknown.

2.5.5 SAMHD1

SAM domain and HD domain-containing protein 1 is a protein that in humans is encoded by the SAMHD1 gene. It exhibits phosphohydrolase activity by converting dNTPs to inorganic phosphate and a 2’-deoxynucleoside (Powell et al. 2011). SAMHD1 also has nuclease activity (Beloglazova et al. 2013). SAMHD1 is involved in controlling the pool of available dNTPs. In this way it can control reverse

transcriptase activity, DNA replication and genome stability. In fact, it has been proved that SAMHD1 is a negative regulator of LINE1 retrotransposition mainly acting through the inhibition of the ORF2p-mediated LINE1 reverse transcription by directly reducing ORF2p level (figure 2). In fact mutations in this gene have been related to defective LINE1 inhibition (Zhao et al. 2013, 1). Furthermore, depletion of SAMHD1 in fibroblasts was reported to result in increased genomic DNA damage, due to dNTP imbalances, leading to intrinsic activation of IFN signaling (Kretschmer et al. 2015).

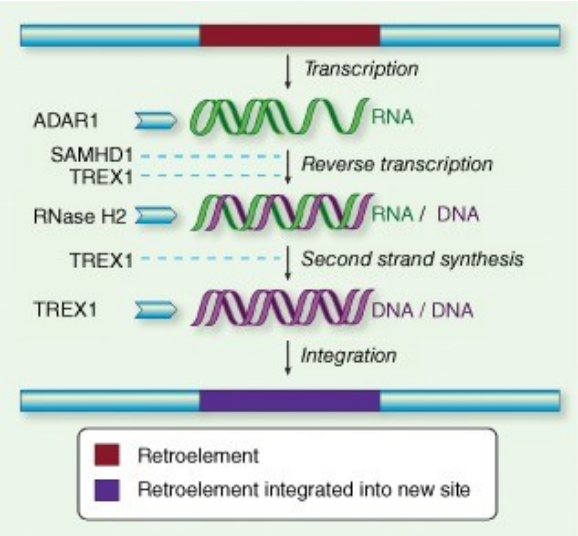


Figure 2: AGS causative genes impact TEs activity at different stages. Different AGS causative genes can interfere with TEs activity at different stages (Y J Crow et al. 2014).

2.6 DNA damage and DNA damage response

Cells are continuously subjected to many different endogenous and exogenous insults that can damage DNA causing modifications and mutations. Although considered the basis of evolution, the vast majority of these insults lead to unwanted consequences that can cause the onset of several diseases (Alberts et al. 2002). To counteract these hazards, cells have evolved specific mechanisms, collectively named DNA damage response (DDR), to detect, signal and repair damaged DNA (Giglia-Mari, Zotter, and Vermeulen 2011). DNA repair is thus a fundamental aspect of cell biology that ensure the correct maintenance and delivery of the genome to the next generations (Jackson and Bartek 2009). These processes can be summarized in 3 main steps that are

common to all mechanisms: sensing of damaged DNA, signaling and repair (Harper and Elledge 2007). The principal mechanisms of DNA repair are the mismatch repair (MMR), base excision repair (BER), nucleotide excision repair (NER), non-homologous end joining (NHEJ) and homologous recombination (HR). In this paragraph, the 5 different pahways will be brielfy and generally introduced before placing them in the context of AGS (Table 3).

Gene name	Gene symbol	Pathway
Xeroderma Pigmentosum Complementation Group C	XPC	NER
RAD23 Homolog B	RAD23B	NER
Transcription Initiation Factor II H	TFIIH	NER
Cockayne Syndrome Group A	CSA	NER
Cockayne Syndrome Group B	CSB	NER
mutS homolog 2	MSH2	MMR
mutS homolog 3	MSH3	MMR
mutS homolog 6	MSH6	MMR
Exonuclease 1	EXO1	MMR
PMS1 homolog 2	PMS2	MMR
mutL homolog 1	MLH1	MMR
DNA polymerase δ	POLδ	MMR
DNA ligase 1	LIG1	MMR
Apurinic/Apyrimidinic Endonuclease	AP	BER
DNA polymerase I	POL I	BER
X-ray Repair Cross Complementing 6	XRCC6	NHEJ
X-ray Repair Cross Complementing 5	XRCC5	NHEJ
DNA-dependent Protein kinase	DNA-Pkcs	NHEJ
X-ray Repair Cross Complementing 4	XRCC4	NHEJ
Non-homologous end Joining Factor 1	NHEJ1	NHEJ
DNA ligase 4	LIG4	NHEJ
MRE11 homolog	MRE11	HR
RAD51 recombinase	RAD51	HR
Nibrin	NBS1	HR
RB Binding Protein 8 Endonuclease (CtIP)	RBBP8	HR
Heterotrimeric Replication Protein A	RPA	HR
Breast Cancer type 2 Susceptibility	BRCA2	HR

Table 3: Summary of genes involved in DNA damage sensing and repair. The table reports the list of genes involved in the DNA sensing and repair along with the full gene name and the pathway in which the gene is involved.

2.6.1 DNA damage response mechanisms

2.6.1.1 Nucleotide excision repair (NER)

NER is the pathway of election for resolving bulky DNA lesions like pyrimidine dimers generated following UV light exposures. Uncorrected lesions can inhibit both DNA and RNA polymerase, cause misreading during replication and transcription, or lead to replication arrest (Goodsell 2001). NER has been divided into two major branches: global genome NER (GG−NER) and transcription coupled NER (TC−NER). In GG-NER DNA lesions are detected by Xeroderma pigmentosum complementation group C (XPC) and RAD23 homolog B (RAD23B) that are able to detect local single strand DNA (ssDNA) and DNA elix distortions mainly caused by pyrimidine dimers. Upon recognition, a multiprotein complex composed of more than 20 proteins binds to the lesion and recruits the transcription initiation factor II H (TFIIH) complex, a transcription initiation and repair factor composed of ten protein subunits that perform the excision and gap filling (Chatterjee and Walker 2017). The TC-NER pathway is similar to the GG-NER but here the DNA damage is sensed and signaled by the stalling of the RNA polymerase II that recruits the Cockayne syndrome group A (CSA) and Cockayne syndrome group B (CSB) proteins. The following steps are similar to those of the GG-NER (figure 3A) (Table 3)(Chatterjee and Walker 2017).

2.6.1.2 Mismatch repair (MMR)

MMR is a mechanism to detect and repair mismatched bases and small insertions/deletions generated during DNA replication and recombination (G.-M. Li 2008). After mismatch recognition by the MutSα, a heterodimer composed by mutS homolog 2 (MSH2) and mutS homolog 6 (MSH6), and insertions/deletions recognition by MutSβ, a heterodimer composed by mutS homolog 2 (MSH2) and mutS homolog 3 (MSH3), the complex MutLα, composed by the heterodimer mutL homolog 1 (MLH1) and PMS1 homolog 2 (PMS2), which is recruited to DNA lesions, performs a single-stranded strand-specific DNA nick on the newly synthesized DNA strand, that contains the mismatch (Kadyrov et al. 2006). After this step, the exonuclease 1 (EXO1) protein, with its exonuclease activity, degrades the damaged strand and POLδ synthesize the new DNA filament using the other strand as a

template. Finally, DNA ligase 1 (LIG1) performs the final ligation step (figure 3A) (Table 3) (Y. Zhang et al. 2005).

2.6.1.3 Base excision repair (BER)

BER detects and repair damages arisen from oxidation, deamination and alkylation. Damaged bases are firstly recognized by a base-specific DNA glycosylase enzyme that mediates base removal by leaving an abasic site. The apurinic/apyrimidinic endonuclease (AP) then perform a nick in DNA between the sugar and the phosphate group. Then DNA polymerase I synthesizes the new DNA filament using the other filament as template. In the end, ligases bind the newly synthetized filament (figure 3A) (Table 3) (Krokan and Bjørås 2013).

2.6.1.4 Non-homologous end joining (NHEJ)

NHEJ is a DNA repair mechanism that is able to repair double-strand breaks. This is one of the predominant repair pathways in mammalian cells and plays a fundamental role in repairing harmful double-strand breaks. The mechanism, active throughout the cell cycle, performs the joining of DNA ends without the need for a template strand and thus is considered an error-prone repair mechanim (Chang et al. 2017). The majority of double-strand breaks are composed of incompatible ends that preclude direct ligation and necessitate further processing. Firstly both ends of the double-strand breaks are recognized by x-ray repair cross complementing 6 (XRCC6) / x-ray repair cross complementing 5 (XRCC5) heterodimer that, in turn, recruits the catalytic subunit of the DNA-dependent protein kinase (DNA-Pkcs). Most of the time the double-strand break ends are incompatible so they are either resected with the degradation of short regions of the 5′ and/or 3′ forming overhangs to generate small regions of microhomology (≤4 nucleotides) between the strands that can facilitate end-joining or filled in by polymerases. In the end, the 2 compatible strands are ligated by a ligase complex composed by x-ray repair cross complementing 4 (XRCC4), non-homologous end joining factor 1 (NHEJ1) and DNA ligase 4 (LIG4) (Brandsma and van Gent 2012). It is worth mentioning that NHEJ is *per-se* a mutagenic pathway. In fact, the original DNA sequences can be recomposed only if the double-strand break creates “compatible” ends that can be directly rejoined. In all the other cases, when

ends are not compatible, the action of exonucleases and polymerases will introduce base substitutions, insertions and deletions (figure 3A) (Table 3) (Kuhfittig-Kulle et al. 2007).

2.6.1.5 Homologous recombination (HR)

HR is a pathway able to repair double-strand breaks. The main difference with the NHEJ is that this pathway require a sequence identical to the template to repair. The identical template sequence, in eucaryotic cells, is the sister chromatid. It appears clear that this pathway can take place only in S and G2 phases of the cell cycle when the DNA has been replicated. HR is thus active, in higher eucaryotes, only in S and G2 cell cycle phases to avoid recombination between non-identical sequences that could lead to the generation of indels (figure 3B) (X. Li and Heyer 2008). HR starts with resection of both ends of the double-strand break by a protein complex composed by MRN (MRE11 Homolog, Double Strand Break Repair Nuclease (MRE11), RAD51 recombinase (RAD51) and nibrin (NBS1)) and RB binding protein 8 endonuclease (RBBP8) that generate 3' free ends by degrading only the 5' strand. The ssDNA is then bound and protected by the heterotrimeric replication protein A (RPA) protein complex that maintains DNA in a linear form and prevents the formation of secondary structures. Then breast cancer type 2 susceptibility (BRCA2) and RAD51 form a protein complex that is able to recruit the identical sequences on the sister chromatid. After that, the same protein complex mediates the so-called "strand invasion" step in which the ssDNA from one chromatid invades the same sequences on the sister chromatid. At this point, the DNA end is extended using the intact sequence as template and the second end of the broken DNA is captured. Finally, the junctions are resolved to give a precisely repaired DSB (figure 3A) (Table 3) (Brandsma and van Gent 2012).

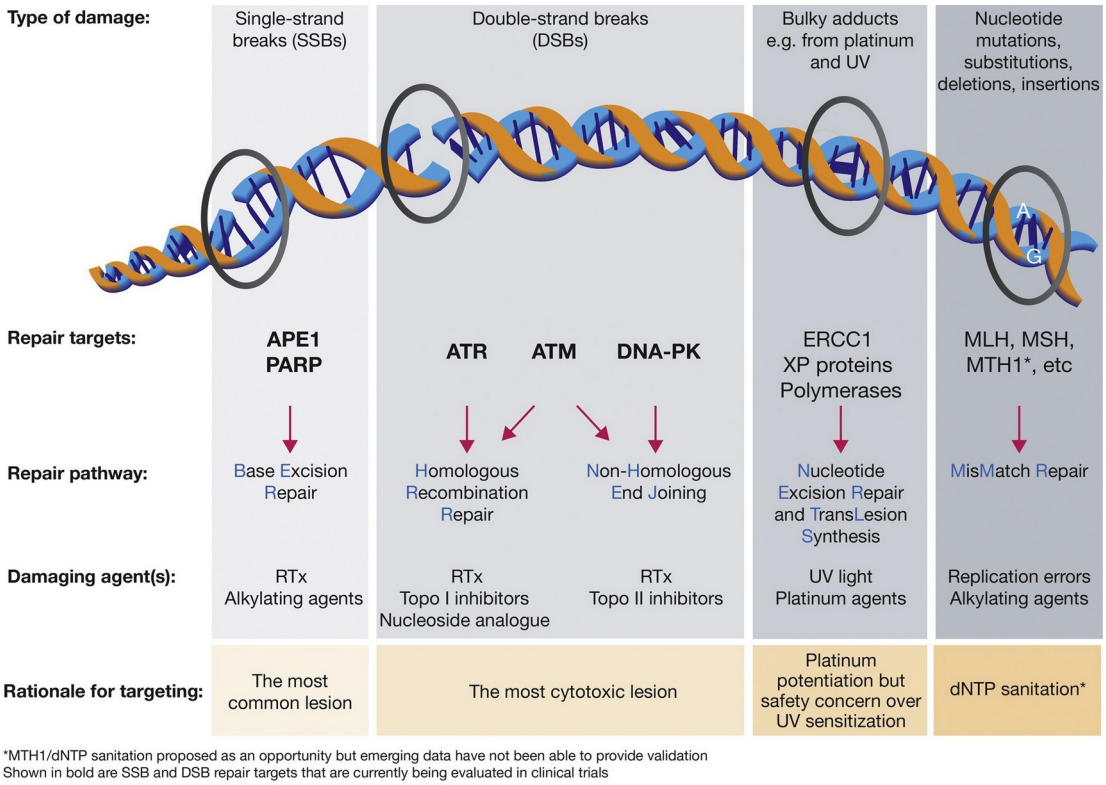


Figure 3: Different types of DNA damage are repaired by specific pathways. DNA repair pathways and activation of specific cellular responses including the rationale behind the choice (adapted from (O'Connor 2015)).

2.7 DNA damage and DNA damage response are involved in the generation of extrachromosomal DNA in AGS

Recently it has been hypothesized that both DNA damage and DNA repair can be a source of both extrachromosomal single strand DNA (ssDNA) and double strand (dsDNA) (T. Li and Chen 2018). This evidence has paved the way for a possible relationship between DNA damage, DNA repair and AGS. Interestingly the DNA damage and repair pathways seem to be involved, although probably in different ways, in the onset and progression of AGS independently of the causative gene. Recent research have provided strong evidence that genomic DNA damage leads to cGAS activation through the generation of extrachromosomal DNA species that stimulate the IFN response (T. Li and Chen 2018). Indeed, the cytosolic DNA sensing pathway appears to be linked with DNA damage and innate immunity. In 2015, Lim and colleagues have reported that defects in TREX1, RNASEH2, and SAMHD1 trigger the DNA damage response and are associated with increased genomic instability (Lim et

al. 2015). At this regard, different groups have independently reported that defects in any of the 3 subunits of the RNASEH2 complex lead to chronic DNA damage and increased IFNs response (Hiller et al. 2012; Pokatayev et al. 2016; Mackenzie et al. 2016). Furthermore, Aditi *et al.* have observed an increased genomic instability independent of type I IFN signaling in RNASEH2B depleted cells (Aditi et al. 2021). They, and other groups, demonstrated that RNASEH depletion induces the misincorporation of rNTPs in genomic DNA that leads to genomic instability (Aditi et al. 2021; Potenski et al. 2019). Moreover, both human cells depleted for RNASEH2 or derived from AGS subjects show impaired cell cycle progression associated with chronic activation of post-replication repair and genome instability (Pizzi et al. 2015) that promotes the formation of micronuclei, which, in turn, recruit cGAS to activate type I IFN signaling (Bartsch et al. 2017; Mackenzie et al. 2017). Genome instability has been also independently observed in primary fibroblast derived form AGS subjects with mutation in SAMHD1 (Kretschmer et al. 2015, 1). The authors have demonstrated that SAMHD1 mutation causes genomic instability through an imbalance of dNTP concentrations that leads to the activation of type I IFN signaling (Kretschmer et al. 2015). One of the main functions of the TREX1 exonuclease is the control of the amount of extrachromosomal DNA to avoid an unwanted IFN response. In fact, cells lacking TREX1 suffer from an overload of ssDNA fragments as well as chronic IFN induction and autoinflammatory phenotypes (Hasan et al. 2013). I have already reported that ssDNA can arise as consequence of both DNA damage and DNA repair. In fact, it has been reported that an intracellular accumulation of abnormal ssDNA species in TREX1 deficient cells leads to chronic activation of the ataxia telangiectasia mutated (ATM) kinase-dependent DNA damage checkpoint and defective G1/S transitions (Yang, Lindahl, and Barnes 2007). Moreover, it has been observed that in normal S-phase and in response to genotoxic stress, TREX1, at least partly, redistributes to the cell nucleus in which it is able to metabolize ssDNA species arising from DNA replication (Lindahl et al. 2009b). These findings suggest that TREX1 is capable of degrading ssDNA generated from the processing of aberrant DNA replication intermediates or DNA damage response, therefore, attenuating DNA damage checkpoint signaling and preventing pathological immune activation (Yang, Lindahl, and Barnes 2007). Taken together these observations suggest a strong interplay between DNA damage response pathways and extrachromosomal DNA

generation candidating those pathways as pivotal players in DNA damage–induced inflammatory response in AGS (figure 4).

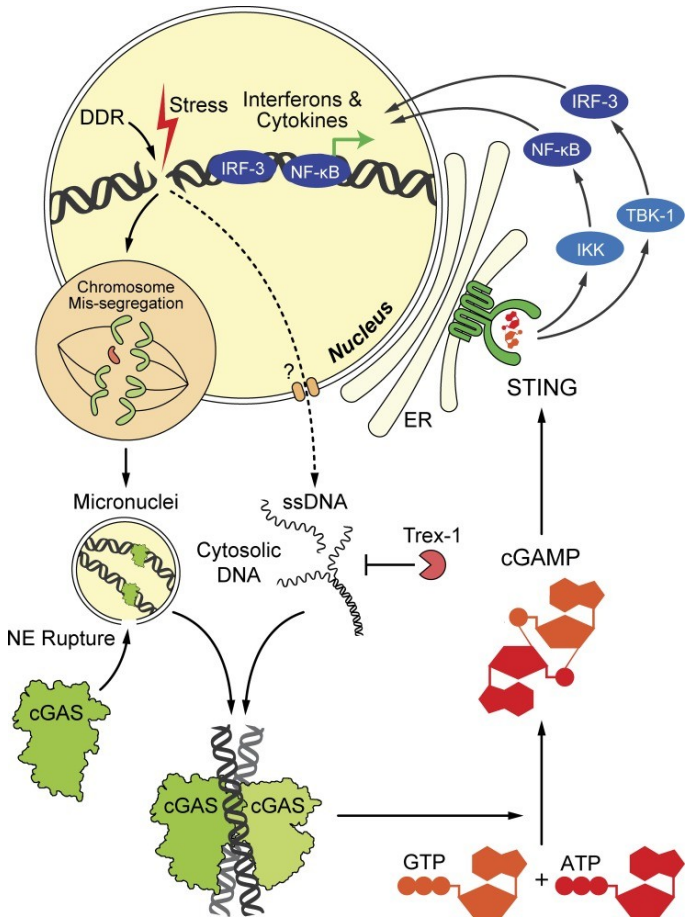


Figure 4: DNA damage and repair is a source of extrachromosomal DNA. Genomic instability and DNA repair can trigger IFN response due to the accumulation, when degradation systems (such as TREX1) do not work, of ssDNA species that activate the cGAS-STING pathway (T. Li and Chen 2018).

2.8 Transposable elements

Transposable elements (TEs) are repetitive sequences that comprise nearly half of the human genome (Lander et al. 2001; Ayarpadikannan and Kim 2014). Their main characteristic is the ability to change their location within the genome and they are virtually found in all eukaryotes (Kidwell 2002; Wessler 2006). They were firstly described in 1940 by Barbara McClintock and for decades have been considered as

"junk DNA" (McClintock 1965). Nowadays it is widely accepted that TEs contribute to the evolution of many different processes like gene expression regulation and splicing (figure 9) (Payer and Burns 2019; Reilly et al. 2013; Saleh, Macia, and Muotri 2019; Chuong, Elde, and Feschotte 2017). TEs contain several regulatory sequences with the capability of influencing the genomic environments located near their insertions (Ali, Han, and Liang 2021). One of the most important effects of TEs insertions is exerted on gene expression. In fact, many TEs harbor regulatory sequences that can act as transcriptional promoters, enhancer or silencer (Sundaram and Wysocka 2020). In particular intact long interspersed nuclear elements (LINE) and long terminal repeats (LTR) elements contain a PolII promoter that can initiate the transcription of genes located downstream. For example, the insertion of an ERV element upstream of the AMY1 gene leads to its tissue-specific expression (Samuelson et al. 1990). In addition, LINE1 elements also contain an active antisense promoter that can drive transcription from the antisense strand to generate a chimeric transcript composed of the LINE1 5′ UTR followed by the adjacent genomic sequence (Speek 2001; Pinson et al. 2022). TEs are therefore a source of DNA binding site sequences. In this scenario, they also recruit different epigenetic factors influencing the local chromatin status with effect on nearby genes expression (figure 8 and figure 9) (Ohtani and Iwasaki 2021). These evidence suggest that TEs insertions harbor the potential to influence many cellular processes. TEs are divided into 2 categories based on the transposition mechanism, sequence similarities and structural relationships: class I or retrotransposons, which transpose using an RNA intermediate, class II or DNA transposons, which do not use any RNA intermediates (figure 5) (Wicker et al. 2007).

2.8.1 Class I elements (Retrotransposons)

Class I elements are the most abundant class in the human genome and comprise about the 40% of it (Goodier 2016). They move by a "copy and paste" mechanism involving the transcription of the element into an RNA intermediate that is then reverse transcribed by a TE-encoded reverse transcriptase and inserted at a new genomic site (figure 5 and figure 6). Due to this mechanism, the original locus remains unchanged and a new copy of the elements is inserted elsewhere in the genome thus increasing the number of copies in the genome (Havecker, Gao, and Voytas 2004). According to the

sequence features the class I TEs are divided into long terminal repeats (LTR) retrotransposons and non-LTR retrotransposons.

2.8.1.1 LTR retrotransposons

LTR retrotransposons sequences are remnants of past viral infections and are characterized by the presence of 2 LTR flanking an internal coding region that encodes for the gag and pol genes (figure 5). The *gag* gene encodes a structural protein capable of forming virus-like particle that represent the discrete space in which retrotranscription occur. The *pol* gene instead encodes a protease, a reverse transcriptase, an RNaseH and an integrase (Havecker, Gao, and Voytas 2004). The 5' LTR repeat contains an internal promoter while the 3' LTR contains a polyadenylation signal. LTR retrotransposons are present with about 450,000 copies and occupy about the 10% of the human genome (Piriyapongsa et al. 2007). A full-length element is about 6-9 kb long (Chuong, Elde, and Feschotte 2017). Nowadays human LTR elements cannot autonomously retrotranspose due to the accumulation of mutations and are thus considered as inactive elements (K. K. Kojima 2018; Mills et al. 2007; Richardson et al. 2015). The mechanism of retrotransposition begins with the transcription, by the cellularly encoded RNA polymerase II, of the LTR element mediated by the promoter sequence located in the 5' LTR region. The LTR-derived mRNA is then exported to the cytoplasm in which encodes for the gag and pol proteins. The gag protein gives rise to the virus-like particle (VLP) in which the LTR mRNA is reverse-transcribed through multiple steps giving rise to a double-stranded cDNA molecule. In the end, the cDNA enters the nucleus where is integrated into the host DNA by the LTR-mediated activity of the integrase protein (figure 6) (Havecker, Gao, and Voytas 2004).

2.8.1.2 non-LTR retrotransposons

Non-LTR retrotransposon is the most common TEs class in the human genome and is the only class of TEs known to still be currently active in the human genome (figure 5) (Mills et al. 2007; Han 2010). They mobilize through a process called target-primed reverse transcription (TPRT) (figure 6) (Finnegan 1997). This class can be divided into 2 main families depending on sequences, genomic structure and the capability to

autonomously retrotranspose. Long interspersed nuclear elements (LINEs) represent the family that can autonomously retrotranspose while short interspersed nuclear elements (SINEs) is the non-autonomous family (figure 5) (Kapitonov, Tempel, and Jurka 2009).

2.8.1.2.1 LINE

LINE are repetitive sequences of which LINE1 is the most abundant and autonomous subfamily in humans (Ostertag and Kazazian 2001). LINE1 made up about the 17% of the human genome and, although the majority of copies are inactive because fragmented, diverged and subfunctionalized sequences, there are about 100 full-length potentially active copies that can autonomously retrotranspose (figure 5) (Ostertag and Kazazian 2001; Sassaman et al. 1997). A full-length LINE1 element is about 6 kbp long and is composed of a 5' UTR region that contain an internal promoter, a 3' UTR region with a polyadenilation site and 2 non-overlapping ORF (ORF1 and ORF2) (figure 7). ORF1 encodes for a protein with RNA-binding and chaperone activity that are both necessary for RNA stabilization and retrotransposition (Martin 2006). The ORF2 protein encodes for a protein with both endonuclease and reverse transcriptase (RT) activity (Feng et al. 1996). The consensus target site for LINE1 endonuclease is AT rich and represented by the 5'- TT | AAAA -3' consensus on the plus strand (corresponding to a minus strand 3'- AA | TTTT -5' where the vertical bar represents the cleavage site) (Szak et al. 2002). LINE1 activity can also cause the retrotransposition of other non-autonomous transcribed TEs by providing the necessary protein machinery (figure 8) (Richardson et al. 2015).

2.8.1.2.2 SINE

SINEs are repetitive, up to 500-600 bp long, non-coding sequences that occupy about the 13% of the human genome (Maquat 2020). They are transcribed by cellular RNA polymerase III and are not capable of autonomously retrotransposing but can retrotranspose due to the presence of cellular encoded reverse transcriptase proteins like the LINE1 encoded ORF2p (figure 8) (Richardson et al. 2015; Dewannieux, Esnault, and Heidmann 2003). A SINE sequence is commonly composed of a promoter region, a middle body and a 3' terminal region. SINEs promoter region

displays some similarities with different types of small noncoding RNAs like tRNA, 7SL RNA, or 5S rRNA (figure 5) (Piskurek et al. 2003). The origin of SINE bodies is not easy to trace, although a region descending from LINEs in a large fraction of SINE families has been identified. The tail is a sequence of variable length consisting of simple repeats and often contains a polyadenylation signal that is necessary to promote non-autonomous retrotransposition by providing a binding site for LINE1 RT (Kramerov and Vassetzky 2011).

2.8.2 Class II

DNA transposons comprise about 3 % of the human genome and most of them mobilize through a “cut and paste” mechanism involving the excision of the element and the reintegration in a new genomic site (figure 5) (Goodier 2016). With few exceptions, no active DNA transposons are known in mammals (Platt, Vandewege, and Ray 2018).

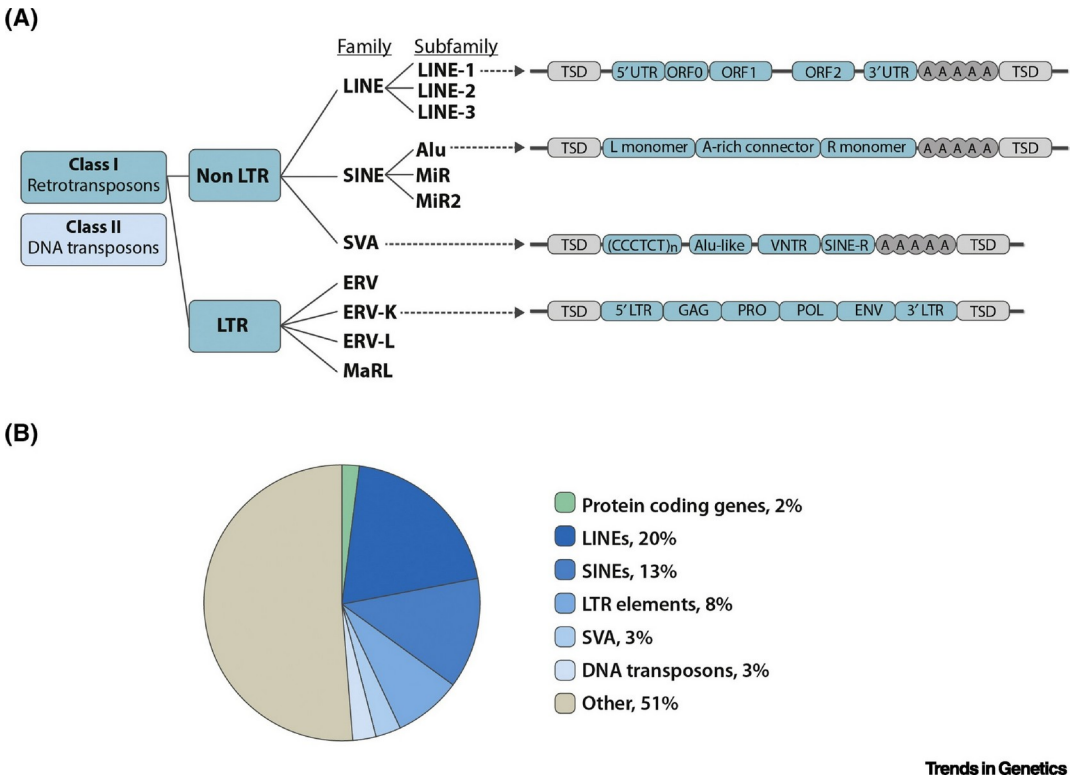


Figure 5: Transposable elements classification. (A) Transposable elements are divided into Class I and Class II elements based on the mechanism of replication and sequence features. Class I retrotransposons are further divided into LTR and non-LTR elements that comprise LINE, SINE and SVA elements. (B) Transposable elements occupy about the 50% of the human genome, of them, LINE is the most abundant family (Jönsson et al. 2020).

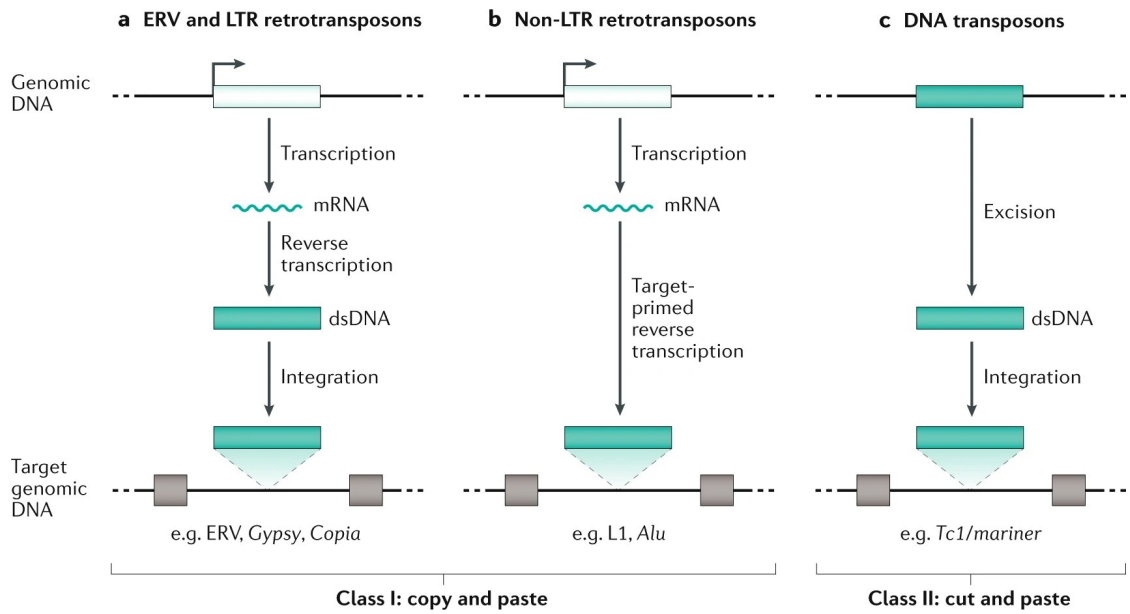


Figure 6: Transposable elements replication mechanisms. (a) Class I retroelements of the ERV and LTR class replicate through a copy-and-paste mechanism. Their replication mechanism involves the generation of an RNA intermediate and a reverse transcription step to generate a dsDNA that is integrated into a newly genomic position. (b) Non-LTR retrotransposons, comprising LINE and SINE family, replicate through a copy-and-paste mechanism. After transcription the RNA serves as a template for the target-prime reverse transcription reaction. (c) Class II transposons replicate through a copy-and-paste mechanism in which the sequence is excised from the DNA and inserted in a newly genomic location (Lanciano and Cristofari 2020).

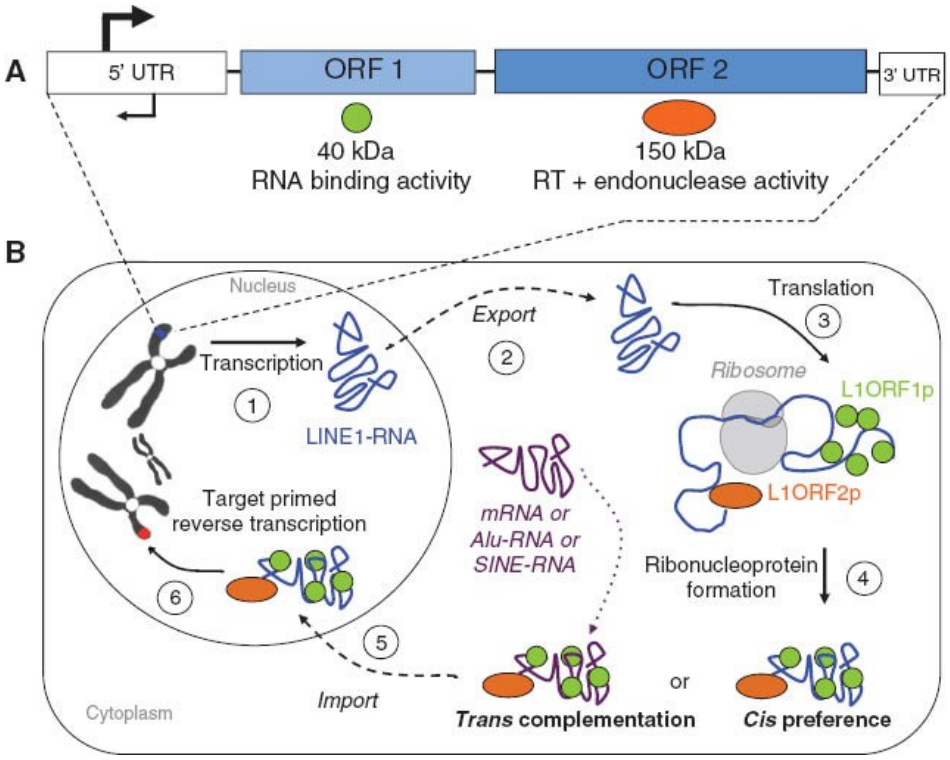


Figure 7: LINE1 structure and retrotransposition mechanism. (A) LINE1 elements are composed of a 5'UTR regions containing an internal promoter, a central region encoding for ORF1p and ORF2p and a 3'UTR region with a polyadenylation signal. (B) The mechanism of LINE1 retrotransposition involves the nuclear transcription and export of LINE1 RNA into the cytosol. Here the encoded ORF1p and ORF2p bind to the element and form an RNP particle that is imported to the nucleus. Here the element is inserted in the newly genomic location with the TPRT mechanism. ORF1p and ORF2p can also lead to the retrotransposition of other TEs as they can also bind other sequences like Alu (Bodak, Yu, and Ciaudo 2014).

2.8.3 Transposable elements and host interaction: control of retrotransposition by host cells factors

TEs occupy about the 40% of the human genome and many of them have been co-opted to perform a plethora of diverse regulatory functions (figure 9) (Reilly et al. 2013). The human brain harbours the highest level of TEs activity and it has been reported that a single neuron can accumulates up to ~13 new somatic LINE1 insertions suggesting a possible role for LINE1 in neuroplasticity (Reilly et al. 2013). Apart from the human brain, higher LINE1 expression has been mainly restricted to germ-line tissues but also detected in nearly all human somatic tissues suggesting possible roles of TEs in physiological functions (Belancio et al. 2010). However,

uncontrolled TEs activation has been related to the onset and progression of many different neurological disorders (figure 9) (Saleh, Macia, and Muotri 2019). In fact their ability to be transcribed and replicate has been recognized to bring both useful and deleterious effects. Cells have thus evolved many distinct mechanisms to control TEs expression and retrotransposition (Cosby, Chang, and Feschotte 2019; Tam, Ostrow, and Gale Hammell 2019). These regulatory mechanisms include both transcriptional and post-transcriptional regulations and involve, but are not limited to, small non-coding RNA-mediated TEs silencing, DNA methylation, chromatin modification and RNA editing. One of the most famous examples of host defense mechanisms to restrict retrotransposition is the piRNA pathway. piRNAs are small non-coding RNAs with sequences complementary to TEs that are processed and loaded on an Argonaute protein complex. These ribonucleoprotein complexe recognizes complementary target mRNAs triggering a reduction in their abundance both through post-transcriptional or transcriptional mechanisms (Czech et al. 2018). DNA methylation is another powerful tool that cells use to avoid TEs expression. In this context it has been demonstrated that methylation of TEs might reduce their expression and thus mitigate their harmful activity (Jansz 2019). For instance, Patchsung *et al.* have reported that treatment of cells with siRNA directed against SINE sequences increases their methylation levels, lowers the endogenous DNA damage and increases DNA resistance to DNA-damaging agents (Patchsung et al. 2018). Interestingly also LINE1 hypomethylation has been linked to genomic instability and DNA damage (X. Zhang, Zhang, and Yu 2020). In fact a lower methylation level of LINE1 is commonly reported in most cancers leading to genomic instability and alteration of gene expression (Kitkumthorn and Mutirangura 2011). In 2019 Jonsson and colleagues report that the inactivation of the DNMT1 gene, which encodes for the main enzyme responsible for DNA methylation maintenance, in human Neural Progenitor Cells, results in a global loss of DNA CpG-methylation leading to specific transcriptional activation of LINE1 elements (Jönsson et al. 2019). Another important TEs control activity is carried out through RNA editing (Krestel and Meier 2018). RNA editing is a process that converts adenosines to inosines in double-stranded RNAs and is mainly performed by ADAR proteins (Nishikura 2010). Elegant experiments have demonstrated both direct and indirect roles of RNA-editing enzymes in participating in LINE1 repression. For instance, Orecchini *et al.* have reported a

novel function of ADAR1 as suppressor of LINE1 retrotransposition through the direct binding of the basal LINE1 RNP complex by the ADAR1 protein highlighting the role of ADAR1 as a regulator of LINE1 life cycle (Orecchini, Frassinelli, and Michienzi 2017; Orecchini et al. 2017). Interestingly also the sole binding of single-stranded retrotransposon transcripts by RNA editing enzymes, alternatively to their effective enzymatic deamination, may have a homeostatic effect on retrotransposon turnover (Krestel and Meier 2018). Another powerful tool used by cells to mitigate TEs activity is histone modification. SINEs, LINEs and SVAs are in fact typically marked by histone H3 methylation at Lys9 that is a known repressive chromatin marker (Kondo and Issa 2003). Finally, also zing finger proteins play important roles in controlling TEs retrotransposition activity. For example, it has been reported that the ZAP protein, a member of the poly (ADP-ribose) polymerase (PARP) family that can be induced by IFN signaling (Hayakawa et al. 2011), is able to restrict the insertion of non-LTR TEs in cell culture through degradation of their RNA intermediates by recruiting degradation proteins to TEs RNPs. Moreover ZAP overexpression in *HeLa* cells inhibits the retrotransposition of engineered human LINE1 and SINE elements and, accordingly, the siRNA-mediated depletion of endogenous ZAP leads to an increase in human LINE1 retrotransposition confirming a role of this protein in TEs homeostasis control (Moldovan and Moran 2015). Interestingly also many AGS causative genes are reported to be involved in the control of TEs activity confirming the fundamental role of TEs dysregulation in AGS pathogenesis (Tam, Ostrow, and Gale Hammell 2019). Recently, it has been demonstrated an enrichment of retrotransposons sequences, mainly derived by retrotransposons belonging to the LINE1 family, through direct sequencing of cytosolic DNA of TREX1$^{-/-}$ cells (Stetson et al. 2008; Thomas et al. 2017). Curiously TREX1 protein is able, through an exonuclease-independent mechanism, to trigger LINE1 ORF1p depletion thus preventing the retrotransposon-induced DNA damage response (P. Li et al. 2017). Moreover, as for TREX1, it has been proved that SAMHD1 can acts as a negative regulator of LINE1 retrotransposition mainly acting through the inhibition of the ORF2p-mediated LINE1 reverse transcription by directly reducing ORF2p levels. Mutations in this gene have in fact been related to defective LINE1 inhibition (Zhao et al. 2013). Another proposed way in which SAMHD1 can interfere with retrotransposition is by degrading the dNTP pool in non-dividing cells to levels below that necessary for a correct reverse

transcription (Goldstone et al. 2011; Gramberg et al. 2013; Lahouassa et al. 2012). Furthermore, the RNASEH2 complex is able to prevent the formation of functional RNA–DNA hybrids, known to be generated during LINE1 retrotransposition, and specifically interfere with the retrotransposition mechanism (Genovesio et al. 2011; Benitez-Guijarro et al. 2018; Lim et al. 2015; J. Choi, Hwang, and Ahn 2018). Finally, the ADAR1 enzyme, which can edit dsRNA derived by SINEs and LINE1 transcription, inhibits TEs ability to induce type I IFN response (Mannion et al. 2014; Orecchini, Frassinelli, and Michienzi 2017). Taken together these findings highlight TEs control as a fundamental mechanisms to protect cells from different insults and report a strict relation between TEs control and AGS causative genes.

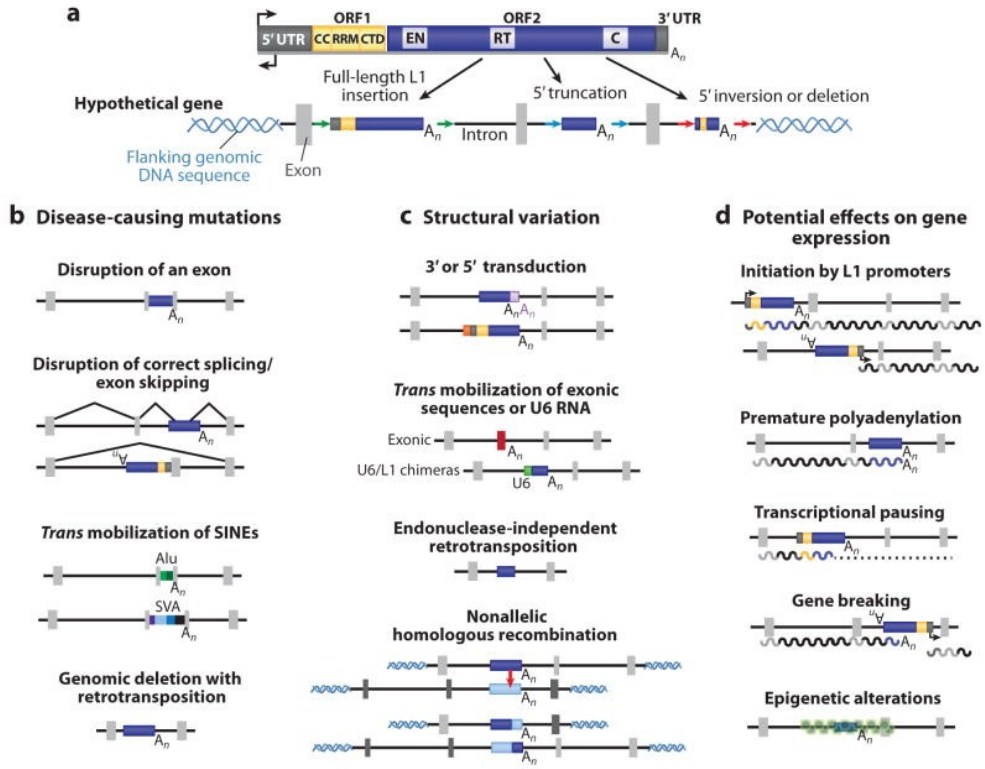


Figure 8: LINE1 activity influences the genome in different ways. (a) LINE1 structures. LINE1 structure is composed by a 5' UTR, ORF1, ORF2 and a 3'UTR. (b) LINE1 activity is involved in the generation of disease-causing mutations. LINE1 insertion, for example, can disrupt exonic sequences, influences the correct splicing or cause the mobilization of other TEs like Alu. (c) LINE1 activity is involved in the generation of structural variations. (d) LINE1 insertions can influence gene expression by providing additional promoter sequences, interfering with transcription or inducing epigenetic alterations (Beck et al. 2011).

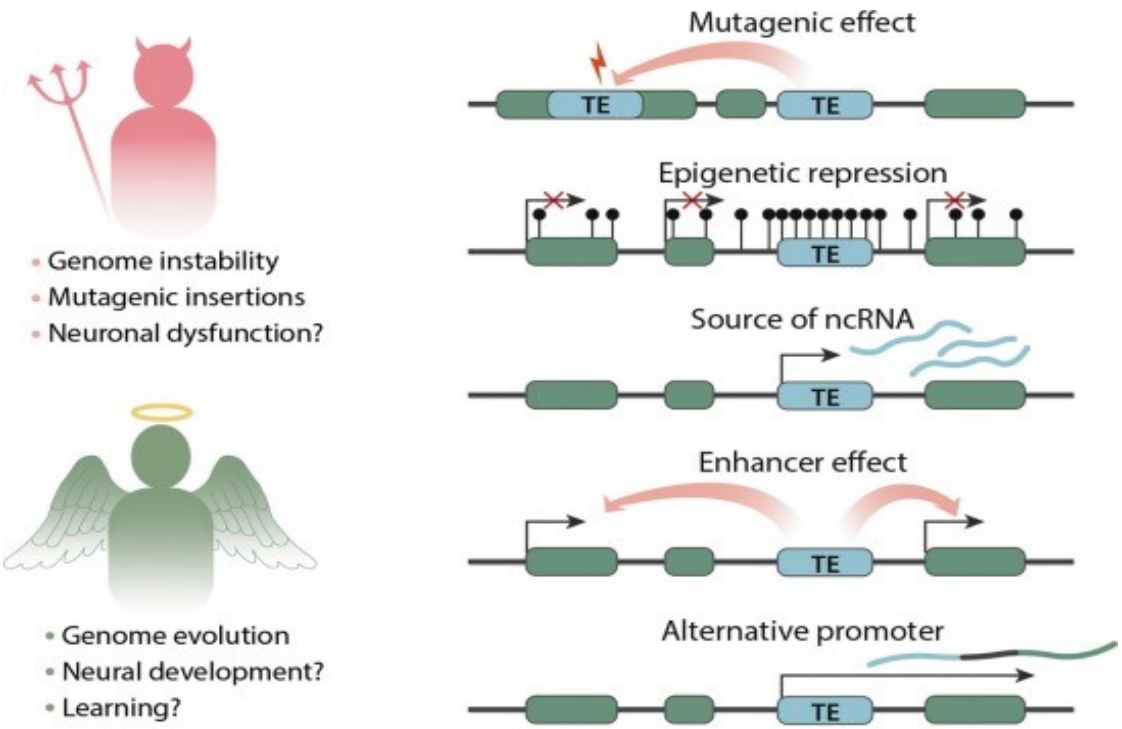


Figure 9: **Transposable elements have both harmful and beneficial roles.** TEs activity can lead to deleterious effects causing genomic instability and mutagenic insertions that lead to neuronal dysfunction and autoimmune disease. On the other hand, TEs activity also participates in genome evolution and is involved in neuronal development (Jönsson et al. 2020).

2.9 TEs activity is associated with cytosolic DNA generation, DNA damage and IFN response

TEs activity is involved in different cellular processes among which extrachromosomal DNA generation is one of the most important in AGS. Several investigations have been conducted to better understand the relationship between TE activity and AGS causative genes. Indeed, it has been demonstrated that TREX1-deficient human astrocytes are overloaded with DNA from retrotranscribed LINE1 extrachromosomal copies capable to promote inflammation through type I IFNs secretion (Thomas et al. 2017). Interestingly treatment of those cells with reverse transcriptase inhibitors rescues the inflammation phenothype suggesting that reverse transcriptase activity is involved in the generation of the increased number of

extrachromosomal DNA from LINE1 copies (figure 10) (Thomas et al. 2017). The fact that reverse transcriptase enzyme could actually reverse-transcribe LINE1 mRNAs also in the cytosol and not exclusively in the nucleus is still under debate (Volkman and Stetson 2014). Recently, in 2021, Fukuda and colleagues identify that Alu undergoes LINE1-mediated reverse transcription via self-priming in the cytoplasm, independent of retrotransposition, providing the first direct evidence of cDNA synthesis in this cellular compartment in human for non-LTR TEs (SINEs and LINEs) (Fukuda et al. 2021). Taken together these evidence confirm that the TEs activity participates to the generation of extrachromosomal DNA that is at the basis of the aberrant type I IFN response observed in AGS (Yanick J. Crow and Rehwinkel 2009). As previously reported, the generation of extrachromosomal DNA can also be the consequence of the activity of DNA repair enzyme.The uncontrolled activation of TEs has been linked to an increase in DNA damage and consequent DNA repair (figure 9) (Peze-Heidsieck et al. 2022; Farkash and Prak 2006; Kines et al. 2016). A first relation between these 2 apparently unrelated processes has been found by Erwin and colleagues. They report that the overexpression of LINE1 in human embryonic stem cells differentiated into hippocampal neurons led to an increase in DNA double strand breaks (DSBs) marked by γ-H2AX (a specific histon modification) which was abolished when the LINE1 endonuclease and reverse transcriptase domains were inactivated (Erwin et al. 2016). More recently it was also demonstrated that heterochromatin destructuration following loss of methylation results in LINE1 derepression that, in turn, can induce DNA damage (Ravel-Godreuil et al. 2021). Other evidence supporting this hypothesis came from a study published in 2018 in which Blaudin de Thè *et al.* highlight that LINE1 overexpression triggers oxidative stress-induced DNA strand breaks that are significantly reduced by LINE1 activity inhibition through different methods both *in vitro* and *in vivo* (Blaudin de Thé et al. 2018). In 2006 it was reported that LINE1 activity causes genome instability identified from an accumulation of γ-H2AX foci, an early response to DNA double strand breaks, in association with an abnormal cell cycle progression through a G2/M accumulation and apoptosis induction. This observation can be explained by the evidence that LINE1 retrotransposition is sensed as a DNA-damaging event through the creation of DNA breaks by the LINE1-encoded endonuclease (Belgnaoui et al. 2006). These results were further confirmed by Gasior and colleagues that demonstrated that LINE1

activity leads to a high level of genomic DSBs formation and that the number of LINE1-induced DSBs is greater than the predicted numbers of successful insertions, suggesting a significant degree of inefficiency during the integration process. This highlight that also the ORF2p endonuclease activity alone could contribute to DSBs formation and that the repair of lesions, by cellular error-prone mechanism, contributes to genomic instability representing a possible driving force for secondary mutations (Gasior et al. 2006; Farkash et al. 2006). Uncontrolled TEs activity appears thus to have multiple consequences on different cellular processes of which DNA damage and consequent genomic instability appear to be of major importance. TEs can induce DNA damage directly by sustaining the activity of the LINE1-encoded endonuclease or also by inducing different cellular stresses that are in turn linked with DNA damage.

In summary, there is an intimate connection between TEs activity and the generation of extrachromosomal DNA. The activity of LINE1-encoded reverse transcriptase can sustain the cytoplasmic retrotranscription of LINE1 and, possibly, other TEs sequences giving rise to the accumulation of extrachromosomal DNA. On the other hand, TEs activity has been linked to increasing DNA damage that can produce, as byproducts, DNA species that, unless correctly degraded, can accumulate in the cytosol. It appears thus clear that TEs activity is linked to aberrant IFN production.

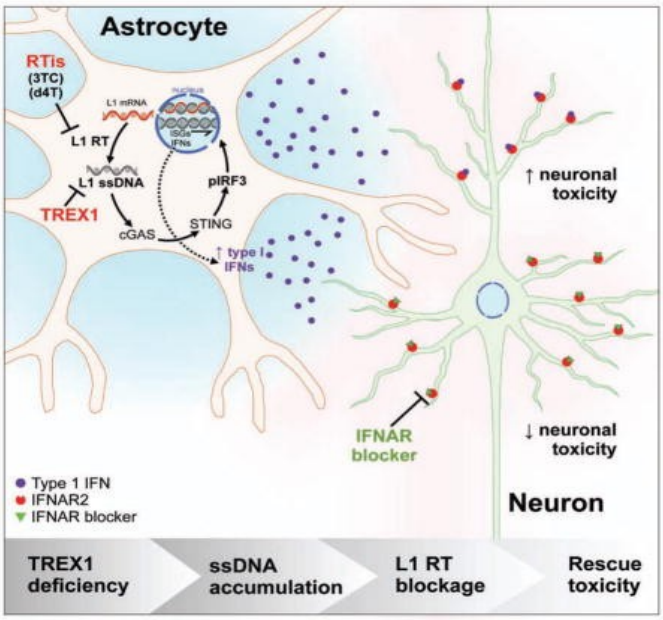


Figure 10: LINE1 sequences made up a consistent fraction of extrachromosomal DNA in TREX1 mutated cells. Extrachromosomal DNA is, at least in part, derived from LINE1 sequences. Extrachromosomal DNA accumulation triggers an IFN response that leads to heightened cell toxicity. Interestingly the treatment with reverse transcriptase inhibitor is able to decrease the aberrant IFN response to near wild-type levels suggesting involvement of RT activity in extrachromosomal DNA generation (Thomas et al. 2017).

2.10 mRNA Splicing

Constitutive splicing is a transcriptional and post-transcriptional RNA modification in which introns are removed from the mature transcript and exons are ligated one by one. The majority of eukaryotic genes are transcribed into precursor messenger RNA (pre-mRNA) molecules in which the exons are separated by the presence of nucleotide sequences called introns (Cooper 2000). Messenger RNA (mRNA) splicing normally occurs in the cell nucleus, even if nowadays some cases of cytosolic splicing have started to be documented (Caceres and Misteli 2007; Denis et al. 2005), and involve the participation of many different ncRNAs and proteins. During the splicing process, introns are removed from the mature transcripts and exons are ligated together. This step is accomplished by a series of precisely regulated events by a supramolecular complex called spliceosome. The spliceosome is composed by a large number of proteins and ncRNAs in which the most abundant components are the small nucleolar ribonucleoprotein (snRNPs) (Wilkinson, Charenton, and Nagai 2020). These

ribonucleoproteins are capable of recognizing and binding to specific regulatory sequences and guiding the excision of the intron from the transcript followed by exon ligation to form a mature RNA. Splicing is mainly guided by regulatory sequences present in both exons and introns. The most important regulatory sequences in splicing are the 2 splice sites. A “canonical” splice signal is composed by a 3’SS (often GU), a 5’SS (often AG) and a branch sites usually located in a polypyrimidine sequence near the 3’end of an intron (figure 11A and B) (Wahl, Will, and Lührmann 2009). Even though this sequence structure has been described as a “basal” splicing signal many variations exist in nature, and the final outcome of splicing is frequently the results of a balance between the basal spliceosome and the activity of many different RNA-binding proteins that can enhance, repress or modify the splicing outcome in a cell type-dependent context (Fu and Ares 2014). Elegant studies have highlighted the series of minimal processes needed for splicing to occur: intron excision initiates with the recognition of the 5’SS by the U1 snRNPs complex followed by the identification of the branch point by the U2 snRNPs. After this first step, the spliceosome assembly continues with the recruiting of the U4, U5 and U6 RNPs to the transcripts to form the so-called “mature spliceosome complex”. The first splicing reaction is represented by the attack of the 2’-hydroxyl group on the branch point adenosine to the phosphodiester group at the 5’SS. This creates a 5’SS blunt end and a so-called “intron lariat” in which the 5’ phosphate of the first nucleotide in the intron splice site, usually a guanine, is ligated to the 2’-oxygen of the branch point adenosine. Then the newly exposed 3’-hydroxyl group of the 5’ exon attacks the phosphodiester group of the 3’ exons thus ligating the two exons and releasing the intron lariat structures (figure 11C) (Wilkinson, Charenton, and Nagai 2020).

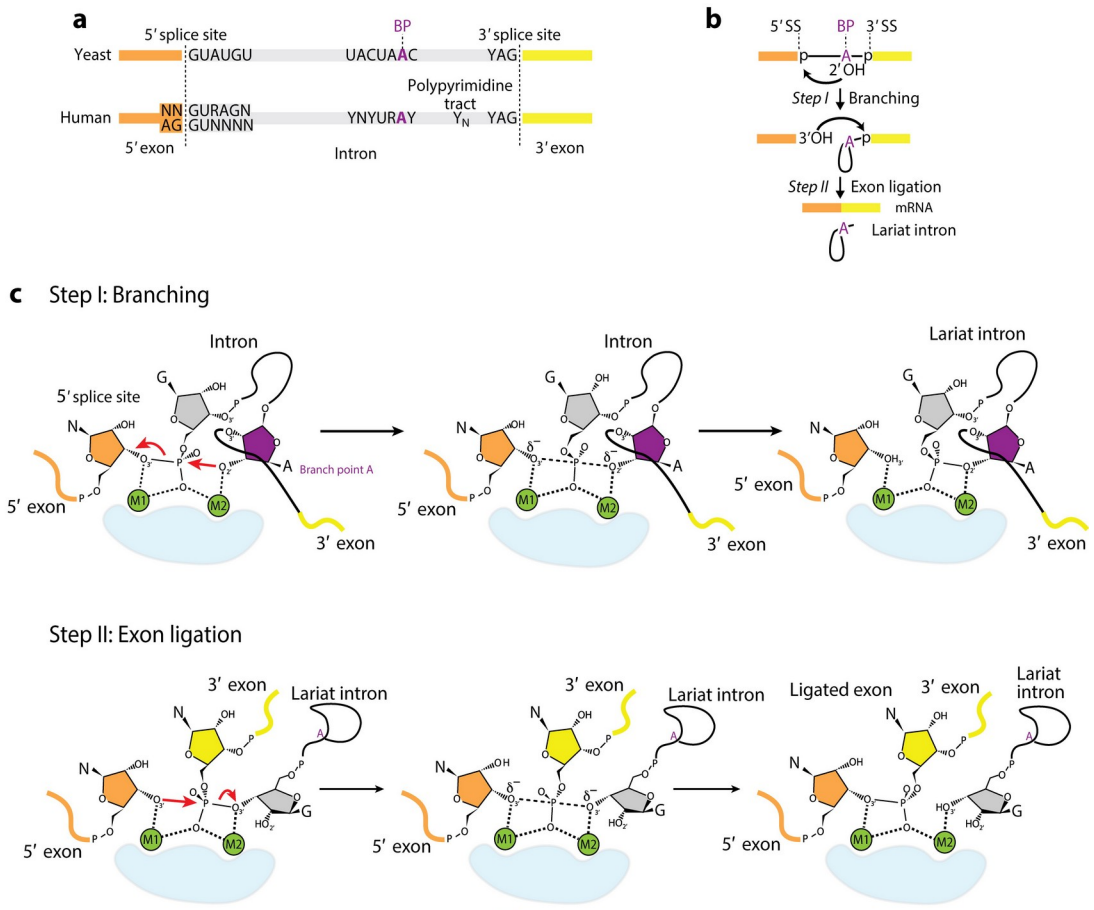


Figure 11: Splicing is a tightly regulated process. (A) Consensus sequences for splicing in humans and yeast. (B) Introns are removed by the splicing machinery by 2 transesterification reactions. (C) Schematic representation of the splicing process. Adapted from (Wilkinson, Charenton, and Nagai 2020).

2.10.1 Alternative splicing

Besides the "canonical" splicing mechanism alternative splicing (AS) has been described as a mode of RNA splicing that allows the pre-mRNA to form different mature transcripts while increasing the transcriptome and proteome diversity without affecting the number of genes (Y. Wang et al. 2015). AS plays important roles in cellular differentiation and organism development, and it is thus finely regulated at both transcriptional and post-transcriptional levels by a variety of factors (Y. Lee and Rio 2015; Y. Wang et al. 2015; Dvinge 2018). In particular, AS is very common in

humans in which affects more than the 60% of genes confirming its important regulatory role (E. Kim, Magen, and Ast 2007). AS comprise 5 different events, in addition to canonical splicing, that are named exon skipping (ES), alternative 5' splice site (A5S), alternative 3' splice site (A3S), intron retention (IR) and mutually exclusive exons (MXE) (figure 12) (Y. Wang et al. 2015). Exon skipping cause the ligation of 2 non-consecutive exons thus eliminating from the final transcripts one or more exons. It represents the most common AS type in humans and mammals with about the 30% of genes affected (Y. Wang et al. 2015; E. Kim, Magen, and Ast 2007). Both A5S and A3S occur when the sites of splicing are different from the canonical one at the 5' or at the 3' ends respectively and can lead to changes in the coding sequences by adding or removing nucleotides at the ends of the exons (Y. Wang et al. 2015). MXE is characterized by the inclusion of one and only one of 2 consecutive exons (Y. Wang et al. 2015). IR occurs when introns, normally removed from the mature transcript by the splicing machinery, are maintained in the mature transcript (Zheng et al. 2020). Among these types of AS, IR has been the most overlooked in mammals and was historically classified as a splicing error. Recently, the advent of newer technologies has paved the way for a deeper understanding of this important phenomenon. In fact, recent studies have demonstrated that IR is involved in different processes such as gene expression regulation, protein stability and protein isoform generation among others both during normal development as well as in response to stress and disease (Jacob and Smith 2017).

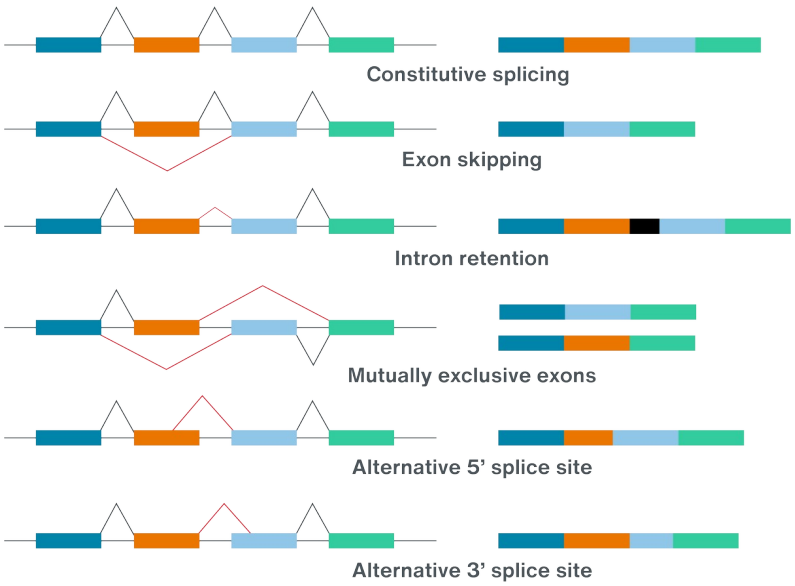


Figure 12: Alternative splicing events. Different types of alternative splicing events and their effect on transcripts sequences. source: *https://nanoporetech.com/applications/investigation/splice-variation*

2.10.1.1 Intron retention: mechanisms and outcomes

IR is an alternative splicing process in which an intron, instead of being removed, is retained in the mature transcript. Despite being originally overlooked in animals, IR has recently gained interest due to its recognized role in gene expression regulation, protein abundance control and association with complex diseases (Monteuuis et al. 2019; Jacob and Smith 2017). Today we know that IR is a widespread and evolutionary conserved mechanism present in both health and disease (Zheng et al. 2020), affecting a wide range of genes with cell type-specific patterns (Braunschweig et al. 2014). One of the main functions of IR is the regulation of transcript abundance. Indeed IR plays important roles in regulating the number of transcripts mainly through the induction of the non-sense mediated decay pathway (NMD). NMD is mainly induced by the inclusion, in the mature transcript, of an intron bearing a premature termination codon (PTC). In NMD pathways the transcripts that include a PTC are eliminated to avoid the production of truncated and non-functional proteins (Schweingruber et al. 2013). In this way IR has been associated with a reduction in

transcript and protein abundance (Y and Bt 2014). One of the first evidence of transcript downregulation from IR comes from the experiments of Kienzle *et al.* that have demonstrated that the retention of an intron in the EBNA3 gene leads to a reduction in its transcript and protein without affecting the transcriptional rate. This suggests that the decreased transcript abundance is not due to a transcriptional downregulation but is the consequence of an increased transcript degradation (Kienzle et al. 1999). In line with this, it has been reported that decreased IR is associated with increased transcript abundance in different contexts. For instance in activated T-cells the decreased intron retention in key genes involved in development and proteasome activity leads to their upregulation allowing a fast response to extracellular stimuli like infections. Generally speaking, it has been observed that the majority of the genes upregulated in activated T-cells are accompanied by a significant reduction in IR (Ni et al. 2016). IR results therefore be involved in the maturation of T-cells. Contrary to what is generally expected, IR can also cause an increase in transcripts abundance mainly by reducing their degradation (Fu 2017). In this context IR can induce the nuclear detention of affected transcripts thus protecting them from NMD, which occurs in the cytoplasm, and favoring their accumulation. Those transcripts can be spliced and exported from the nucleus to the cytoplasm in response to an external stimuli. This gives the cell the unique capability of mounting a faster response with respect to the one obtained beginning with *de-novo* transcription especially for genes with long introns (Boutz, Bhutkar, and Sharp 2015). Interestingly micro-RNA (miRNA) target prediction algorithms have revealed a significantly higher density of putative miRNA binding sites in sequences of retained introns compared to non-retained introns (Schmitz et al. 2017). In humans, for instance, nearly half of the known miRNAs are encoded within the introns of protein-coding genes implying a potential association between intronic miRNA and their harboring genes. In this situation intronic miRNA may target their host transcripts in the intronic portion suggesting another way in which intron retention can be involved in the regulation of gene expression (Steiman-Shimony, Shtrikman, and Margalit 2018; Buckley et al. 2014). Moreover, intron-retaining transcripts might act as competing endogenous RNAs thus indirectly regulating transcript abundance of other genes sharing the same set of miRNA binding sites (Schmitz et al. 2017). Another intriguing function of IR, though only reported in a few cases, is its ability to encode protein domains. In fact the

existence of retained introns in frame within the mature transcript has been demonstrated. In this case a retained intron can participate to the generation of new protein isoforms codifing for additional protein domains. During vascular lesion in rats, a process controlled by smooth muscle cell migration and extracellular matrix secretion, the production of an alternative isoform of the id3 transcription factor, known as id3a, is generated by the inclusion of a 115 bp length intron encoding for an additional 29 amino acids C-terminus domain. This new isoform, which is expressed during the late stages of lesion formation, is characterized by additional DNA binding modules that help to limit the pathological proliferation of smooth muscle cells (Forrest et al. 2004). To identify the possible protein domains encoded by retained introns at a genome-wide scale Galante and colleagues have performed large-scale analysis of intron retention for a set of 21,106 known human genes. Using the Pfam database they have found that 2 retained introns completely encode for a protein domain (Mistry et al. 2021; Galante et al. 2004). Moreover, they have reported that many protein domains are encoded by a combination of exon/retained-intron/exon. In such manner introns can, instead of encoding a complete protein domain, participate to the inclusion of a protein domain portion further expanding the proteome complexity (Galante et al. 2004). An intriguing hypothesis, proposed by T. P. Buckley is summarized in the evidence that a non-negligible number of intron-retaining transcripts are stored in the cytosol, awaiting for specific signal, before undergoing cytosolic splicing (Denis et al. 2005; Caceres and Misteli 2007). Those introns can harbour many different features (e.g., splicing sequences, ID targeting elements, miRNAs and miRNA binding sites) and elements needed to create multiple subpopulations of RNAs under specific cellular conditions. Such an RNA has been referred to as a “sentinel RNA” that could serve as a master transcript from which several other variants could be derived via cytosolic splicing in response to specific stimulation (Buckley et al. 2014). This mechanism allows the cells to increase transcriptome plasticity and allow them to quickly modulate their responses based on external stimuli.

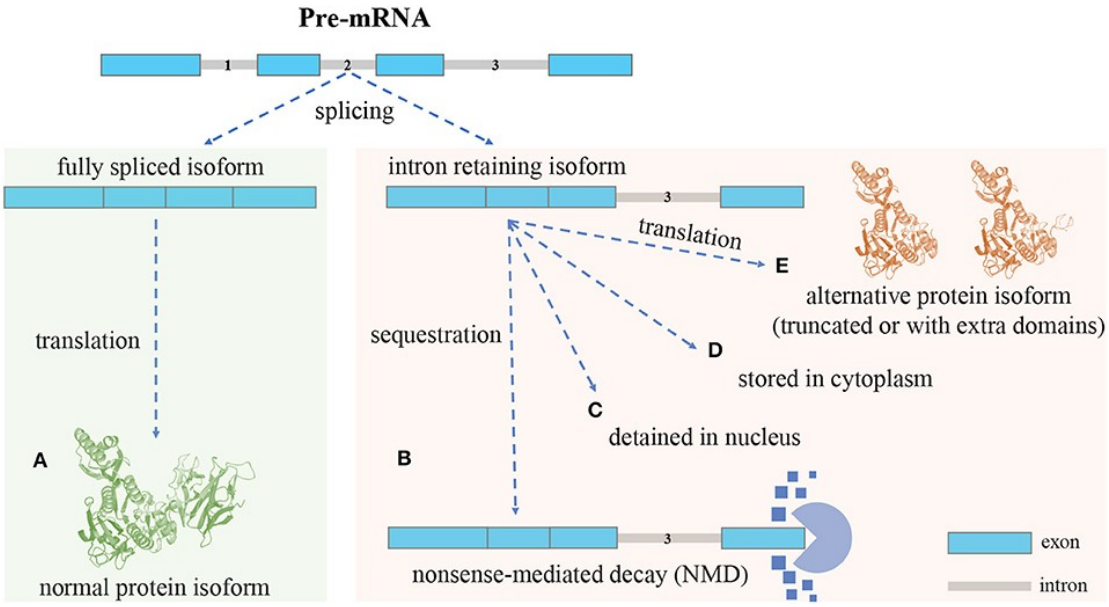


Figure 13: Intron retention participates in different biological processes. (A) The spliced RNA isoforms are exported to the cytoplasm for translation. When some introns are not removed from the mature transcript the transcript undergoes intron retention. (B) In most cases transcripts harboring retained introns are degraded by NMD. (C) Some transcripts harboring retained introns are detained in the nucleus waiting for external stimuli to be spliced and exported. (D) In the case of cytoplasmic splicing, intron-retaining transcripts are shuttled to the cytoplasm for preservation and may be subject to further splicing. (E) Intron-retaining transcripts sometimes can escape NMD degradation and are translated into protein isoforms. Those protein isoforms often result in truncated domains or can include extra domains formed by the amino acid sequences translated from retained introns (Zheng et al. 2020).

2.10.1.2 Intron retention: regulation

Different studies have tried to shed light upon the precise factors that regulate IR. Accumulating evidence have revealed that IR is controlled both by *cis* and *trans*-acting elements (Monteuuis et al. 2019). Firstly IR is regulated by the activity of splicing factor proteins. Generally speaking, an increased global IR has been related to a decreased expression of splicing factor expression (Wong et al. 2013; Ullrich and Guigó 2020). For example, knock-down of U2AF1, a pivotal splicing factor, has been associated with a global increase in IR levels (Yao et al. 2020). Interestingly different experiments have revealed an enrichment of retained introns in transcripts encoding for splicing factors genes thus suggesting the existence of an autoregulatory feedback mechanism. In this scenario, an increased IR in splicing factor would decrease their respective protein level which, in turn, increases the levels of IR. This autoregulatory

feedback mechanism has been associated with the maintenance of a differentiated or quiescent cellular state (Jacob and Smith 2017). Another factor associated with increased IR is the splice site strength. In fact retained introns can be characterized by the presence of weaker splice sites, with respect to constitutively spliced introns, that would increase their retention also upon a slightly decrease of splicing efficiency (Braunschweig et al. 2014). Even though IR can be partially explained by a global decrease in splicing activity in transcripts characterized by weaker splice sites, this feature *per-se* cannot explains a large fraction of reported IR events. In fact the majority of IR events appear to target specific introns in selected genes suggesting that some forms of more precise regulation must occur (Sakabe and de Souza 2007). Different authors report that another hallmark of retained introns is the intron length. Indeed retained introns are, on average, shorter than non-retained introns (Sakabe and de Souza 2007). IR can also be influenced by both chromatin accessibility and nucleosome occupancy. Actually, increased chromatin accessibility has been related to the retention of introns in a cell-specific manner (Petrova et al. 2021). Another commonly reported feature of retained introns is the GC content. Retained introns have, on average, a higher GC content compared to their non-retained counterparts. This evidence has paved the way for the demonstration that IR is associated with high CpG density and often with decreased DNA methylation possibly precluding the correct recruitment of splicing factors (Wong et al. 2017). Moreover, also transcriptional dynamics can affect the level of IR. In fact a slow RNA polII transcription rate has been related to a most efficient recognition of splice sites and a "more efficient" splicing with the reduction of IR (Wong et al. 2017; Fong et al. 2014). The most important *cis-acting* players in IR control are the RNA-binding proteins. These proteins can bind specific intronic motifs influencing the level of IR or regulate the fate of intron-retaining transcripts (Monteuuis et al. 2019; Middleton et al. 2017). Finally, mutations affecting the functionality of some important splicing components are associated with an increase in IR as they reduce the processivity of the splicing machinery (Zheng et al. 2020; Monteuuis et al. 2019). One interesting hypothesis is that TEs activity can influence intron retention. In the plant *Gossypium raimondii*, where IR is the most common AS event, TEs insertions are present in more that 40% of retained introns while their frequency in non-retained introns is significantly lower. TEs insertions near the 3′ splice site of the intron might impact the efficiency of intron

splicing by decreasing RNA secondary structure flexibility or by altering the distribution of branch points (Monteuuis et al. 2019). Although firstly reported in plant genomes this mechanism is conceivable also in mammals. In fact, in humans, introns contain a large number of both LINE1 and Alu elements. Thee elements can cause IR by interfering with gene transcription (Park et al. 2012; Kaer and Speek 2012). In particular, it has been proven that intronic LINE1s and other retroelements in tandem orientation with respect to gene transcription could cause transcriptional interference of the host gene and induce IR (Kaer et al. 2011). The discovery of an increasing number of factors capable of influencing IR has confirmed that cells have evolved various regulatory mechanisms to fine-tune this important mechanism (figure 14).

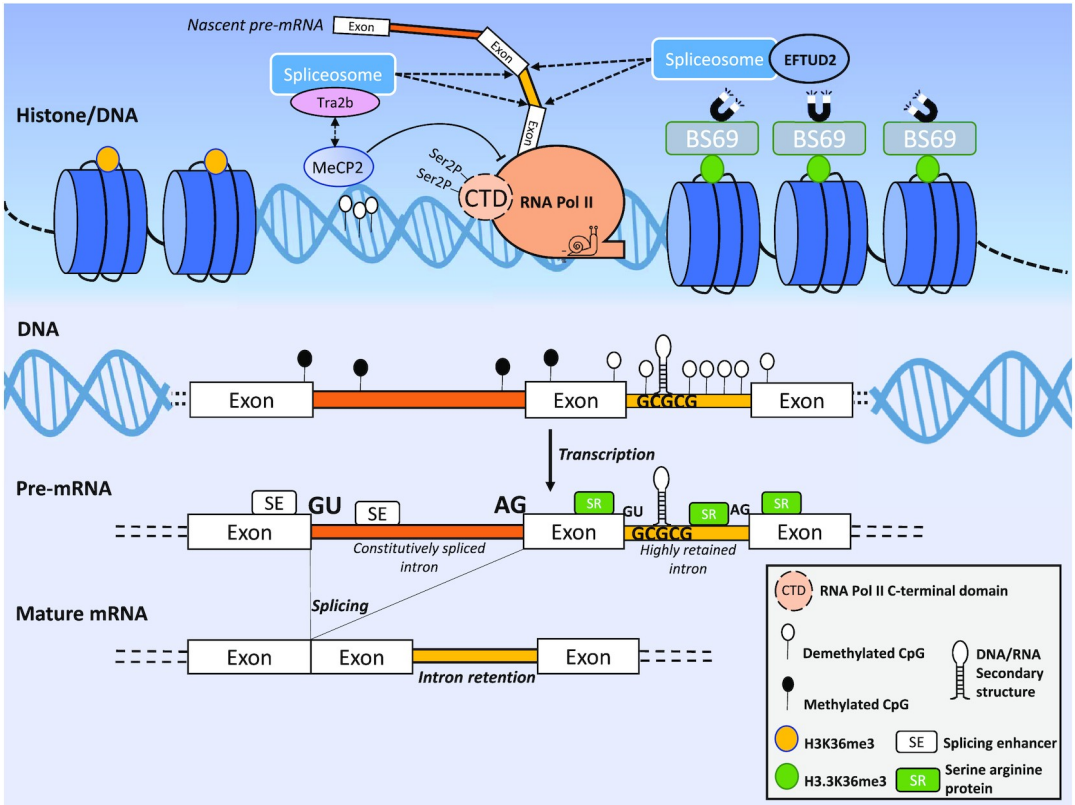


Figure 14: Intron retention is regulated by different genomic features. The figure summarizes the different factors and hallmarks influencing the level of IR. The upper blu shaded part reports epigenetic features while the light-shaded blue lower part DNA methylation and sequence features (Monteuuis et al. 2019).

2.10.1.3 Intron retention is involved in different physiological processes

Recent studies have started to explore the effect of IR in different scenarios discovering that IR plays pivotal roles in the development and activation of the immune response by modulating the differentiation and activation of many immune system cells (Wong et al. 2013; Ullrich and Guigó 2020; Song et al. 2022; Green et al. 2020; Frankiw et al. 2019). IR has thus been mainly investigated in relation to the immune system. In 2020, Green *et al.*, have discovered that macrophage development and activation is coordinated by IR in inflammatory genes (Green et al. 2020). They have produced transcriptomic and proteomic paired data and have reported that a decreased IR is associated with an increased expression of transcript and encoded protein fundamental for monocyte to macrophage differentiation. In particular they have identified that CXCL2 and NFKBIZ transcripts undergo a dramatic decrease in IR during macrophage activation with a concomitant protein upregulation. Interestingly the CLCX2 intron-retaining isoform accumulates during early exposure to a proinflammatory stimuli. They have hypothesized that cells accumulate unspliced intron-retaining isoforms to prepare for subsequent developmental or functional states, such as being able to respond quickly to proinflammatory stimuli without starting a more time consuming *de-novo* transcription (Green et al. 2020). In 2022 Song and colleagues have confirmed and expanded these findings reporting that IR is the major AS event during monocyte differentiation and confirming the previous results obtained by Green *et al.* (Song et al. 2022). An increasing number of evidence have identified IR as an important regulator of monocyte, B-cell and granulocyte differentiation and activation. In fact the IR level increases during B-cells differentiation from the bone marrow precursor to cells residing in secondary lymphoid organs to rapidly decrease upon B-cells affinity maturation (O'Grady et al. 2022; Ullrich and Guigó 2020). Moreover, the IR dynamics in B-cells seems to be conserved between human and mouse confirming the important role of IR in B-cells development (Ullrich and Guigó 2020). Further enrichment analysis has revealed that genes affected by IR mainly belong to splicing, cell cycle and inflammatory response. More specifically they have discovered that the level of IR in introns of a given transcript correlates with the expression of the hosting transcripts. In particular, it has been reported that negatively correlated genes belong to the splicing factors category while positive correlations contain NF-kB and other inflammatory-related genes. This strongly suggests that the

relation of IR with gene expression depends on gene function (Ullrich and Guigó 2020). IR physiologically controls gene expression also during granulocyte differentiation. In fact an increased IR has been observed during granulocyte differentiation. The group that has identified this mechanism has reported that IR specifically regulates the expression of 86 key genes for granulocytes differentiation by inducing NMD in a tightly regulated fashion (Wong et al. 2013). In general, it has been reported an inverse correlation between the level of intron retention and the cell proliferative state even if some exceptions occur. For example in neutrophil it has been reported a positive correlation between IR and the proliferative stage suggesting that IR can have a pleiotropic effect based on the affected genes in a cell type-specific manner (Ullrich and Guigó 2020). In addition to its extensive role in regulating immune system development and activation, many recent research projects have identified a role for IR in neuronal system development. In 2021 Mauger *et al.* identified a novel transcription-independent mechanism capable of rapidly modifying the transcriptome of neuronal cells in response to external stimuli. They have proposed a mechanism in which neuronal stimulation activates a specific splicing program capable of removing specific introns from intron-retaining transcripts in the nucleus, thereby mobilizing a pool of mRNA to the cytoplasm for subsequent translation. They have discovered that those intron-retaining transcripts are normally stored in the cell nucleus, and protected from degradation. These introns have been defined as detained introns. Detained introns can have half-lives of over an hour that exclude the possibility that they rapidly undergo NMD but are instead maintained in the nucleus and protected from degradation (Boutz, Bhutkar, and Sharp 2015). Upon neuronal stimulation, the intron-retaining transcripts undergo rapid intron excision, with concomitant mature mRNA export. This mechanism is thus able to rapidly increase the pool of mature mRNA ready for translation in a way faster with respect to an *ab-initio* transcriptional event. This method is ideal for neuronal genes, which are typically longer than other genes and require longer times to complete transcription (McCoy and Fire 2020; Mauger, Lemoine, and Scheiffele 2016). This evidence suggests that IR can also be involved in the control of neuronal plasticity. In support of this hypothesis, it has been reported that the Staufen2 gene, a known dsRNA binding protein, implicated in dendritic mRNA localization, is able to bind a group of transcripts at a retained-intron located in the gene's 3'UTR. The strongest binding has

been found in the Calm3 transcripts, which encode for the Calmodulin protein, which is involved in neuronal plasticity. Interestingly, the absence of an intron in the 3'UTR of the Calm3 mature transcript, which is normally found in dendrites, causes its mislocalization, impairing neuronal functions (Sharangdhar et al. 2017). Another interesting example has been reported in the study published by Yap *et al.*. They have discovered that increased intron retention during neuronal differentiation is linked with reduced transcript and protein levels. In this context the authors have reported that genes encoding critical presynaptic proteins are transcribed at a detectable level in both neurons and non-neuronal cells. However, in non-neuronal cells, the PTBP1 protein represses the correct splicing leading to IR. This, in turn, triggers nuclear detention and degradation. After PTBP1 reduction, during neuronal differentiation, introns are removed leading to the fast accumulation of translation competent mRNAs allowing the fast and precise production of protein in the appropriate developmental context (Yap et al. 2012). The examples provided till now, establish IR as a crucial player in various physiological processes like immune system activation and development as well as neuronal development and plasticity.

2.10.1.4 Intron retention in neurodegenerative and autoimmune disease

IR has been extensively detailed in its physiological roles but a growing body of evidence have associated changes in IR with different neurological and autoimmune diseases (Sznajder et al. 2018; Ong and Adusumalli 2019; Adusumalli et al. 2019). Recenlty, it has been reported an association between IR and Alzheimer disease (AD) from 3 independent research groups. They have reported that AD samples are characterized by a significant increase in the rate of IR at genes involved in RNA processing and protein homeostasis (Ong and Adusumalli 2019; Adusumalli et al. 2019; H.-D. Li et al. 2021). In this context, IR has been also shown to promote cellular senescence. In fact, through the analysis of polyA+ RNA-seq data of human replicative senescence models, it has been highlighted that IR is abundant and dynamically regulated during senescence with IR levels negatively correlated with expression alteration of corresponding genes in line with what is observed in AD (Yao et al. 2020). The discovery that XAB2 depletion, a human splicing factor, leads to the induction of cellular senescence has suggested an interplay between splicing and cellular senescence. Further investigation identified that XAB2 depletion leads to

increased IR in the POLR2A gene, the largest RNA polymerase II subunits, with a concomitant reduction in both its transcript and protein levels and a concomitant reduction of global transcriptional efficiency that, in turn, induces cellular senescence (Hou et al. 2019, 2). IR has been also found to be widespread across a series of cancer transcriptomes (Dvinge and Bradley 2015). For example, Zhang *et al.* have demonstrated that adenocarcinoma samples are characterized by more than 2000 tumor-specific IR events (Q. Zhang et al. 2014). In 2015 Dvinge and Bradley have analyzed the extent of IR in 805 matched tumor and normal control samples from 16 distinct cancer types and have reported that abnormal RNA splicing, characterized by widespread intron retention, is common across cancers even in the absence of mutations directly affecting the RNA splicing machinery (Dvinge and Bradley 2015). Another hallmark of different tumors is the increased mutational burden defined as the number of somatic mutations acquired during cell life. Tumors are in fact often characterized by an increased mutational rate with respect to normal samples. It is worth noting that IR can be induced by mutations. Mutations affecting splicing donor or splicing acceptor sites, for example, can affect the splice sites recognition thereby inducing the intron inclusion. This has relevant consequences on transcript and protein outcome (Yadegari et al. 2016). In 2018, Sznajder and colleagues have tested the hypothesis that GC-rich intronic microsatellite expansions are able to trigger host IR in different diseases like myotonic dystrophy type 2 (DM2), *C9orf72*-linked amyotrophic lateral sclerosis with frontotemporal dementia (C9-ALS/FTD), Fuchs endothelial corneal dystrophy (FECD), Friedreich's ataxia (FRDA), and spinocerebellar ataxia type 10 (SCA10). They have demonstrated that IR is detectable in affected tissues and peripheral blood lymphocytes and have proposed that IR screening constitutes a rapid and inexpensive biomarker for intronic repeat expansion disease. They have demonstrated that GC-rich intronic microsatellite expansions induce IR and that these retention events are not developmentally regulated. Interestingly, IR occurs only at GC-rich, and not at AT-rich intronic expansion where the mutation is located more distally from the nearest splice site. These results are consistent with the hypothesis that GC-rich microsatellite expansions exert an inhibitory effect on splicing by altering RNA structure and access of splicing factors to intronic regulatory regions (Sznajder et al. 2018). Intron retention has been independently observed by different groups also in amyotrophic lateral sclerosis (ALS), a neurodegenerative disease that induces the

degeneration of upper and lower motor neurons which pathogenetic mechanism is still partially to be determined. It has been reported that a substantial number of ALS subjects are subjected to EAAT2, a functional glutamate transporter, protein loss. Lin *et al.* have demonstrated that the EAAT2 protein loss in ALS may be due to aberrant mRNA splicing. They have reported that about 65% of all postmortem specimens harbour one or both of the intron 7 retention and exon 9 skipping transcripts. Interestingly aberrant *EAAT2* mRNAs were restricted to motor cortex and spinal cord, the primary regions of neurodegeneration confirming a possible specificity in the association to the disease (Lin et al. 1998). Another important evidence resides in the fact that ALS has been associated with the insurgence of mutation in some RNA processing genes that may play a role in the regulation of IR. Luisier and colleagues, during the analysis of transcriptomes derived from in vitro differentiation of motor neuron from subjects iPSCs, have identified a prematurely increased intron retention in many genes among which one of the most important is SFPQ, a key regulator of axonal development. They have reported that SFPQ intron-retaining transcript is exported to the cytoplasm and that SFPQ protein binds to its retained intron. This produces a nuclear loss of SFPQ in ALS. They have thus demonstrated that increased intron retention in SFPQ promotes the nuclear loss of SFPQ protein suggesting a possible role of in ALS (Luisier et al. 2018). At this regards another group identifies that, in C9orf72 GGGGCC containing hexanucleotide expansion, one of the most common causes of ALS, the repetitive RNA is known to bind and sequester the hnRNPH splicing regulator. This, in turn, results in widespread intron retention, affecting more than 2000 transcripts, with impaired protein homeostasis in pathways involved in ALS etiology, including proteasome and autophagy system (Q. Wang et al. 2020). Aside from motor neurons, also astrocytes have been implicated in ALS degeneration. Opposite to what was observed for ALS motor neurons in ALS astrocytes it has been highlighted a decreased intron retention with respect to normal counterparts. Here transcripts with decreased IR are associated with increased gene expression and are over-represented in reactivity processes including cell adhesion, stress response and immune activation (Ziff et al. 2021). Recently, also a work depicting for the first time changes in alternative splicing events in SLE has been published (Papanikolaou, Bertsias, and Nikolaou 2021). The authors, using whole blood samples from SLE subjects, have identified a huge number of differentially

retained introns in genes enriched for metabolism and histone acetylation leading to the first evidence of splicing deregulation in SLE (Papanikolaou, Bertsias, and Nikolaou 2021). The evidence that IR plays a fundamental role in the regulation of immune system response and neuronal system development in addition to its increasingly frequent association with autoimmune and neurodegenerative diseases candidates IR as an interesting disease marker with the potential to providing pathways for new innovative treatments.

2.10.2 Principal challenges in detecting IR from RNA-seq data

The quantification of IR from RNA-seq data faces different difficulties that must be addressed to achieve an accurate and reliable IR quantification. One of the major issues in IR quantification is that, unless observed for other AS events, the sequence of an IR event is identical to those of the pre-mRNA sequence. This issue can be solved by using polyA+ selected transcripts to enrich the samples for processed and mature RNAs (Vanichkina et al. 2018). The use of cytoplasmic polyA+ libraries has been also hypothesized and has resulted in an accurate IR quantification despite the loss of some potentially important nuclear detained mature RNAs (Tyzack et al. 2021). Furthermore, some intronic reads can result from genomic DNA contaminations. This issue has been largely resolved by ensuring that the majority of reads originate from "correct" splicing between exons or that the ratio of exonic to intergenic reads exceeds a certain threshold (Middleton et al. 2017). Moreover, intron annotations can overlap with the annotation of other transcripts, inflating IR quantification so that care must be taken to consider only unique intron regions. From a computational point of view, the use of Illumina short reads makes it difficult to specifically map the reads to unambiguous splice junctions (Grabski et al. 2021). Furthermore, introns are known to contain a high number of repetitive sequences and low mappability regions, which can prevent an accurate and unambiguous mapping of Illumina short reads (Broseus and Ritchie 2020). Nonetheless, numerous approaches have been developed to mitigate these issues and achieve a rather accurate IR quantification from RNA-seq data. Many of these methods contextually measure both the number of intronic reads and the number of reads mapped at the exon-introns boundaries, while controlling for correct splicing using the number of exon-exon reads. Finally, it must be considered that IR quantification typically provides only a temporal snapshot of the process. In fact

introns can be retained in a mature and processed RNA or can represent the accumulation of specific RNA isoforms that can be rapidly spliced post-transcriptionally (Jacob and Smith 2017).

3 Aim of the study

The aim of the study is to investigate the transcriptomic variations in different subjects affected by Aicardi Goutières syndrome and other interferonopathies to gain insight into the molecular mechanisms at the basis of the increased interferon response. The primary functions of interferon signalling is the stimulation of the immune response. Considering that the activation of the immune system has been associated to extensive changes in alternative splicing I aimed at investigating possible dysregulation in alternative splicing with particular attention to intron retention and its relation with immune system activation. This study points out a possible impairment of alternative splicing in Aicardi Goutières syndrome and systemic lupus erythematosus. The identification of altered pathways associated with those diseases can pave the way to the discovery of new diagnostic strategies and the development of therapeutic approaches toward symptom amelioration.

4 Materials and Methods

4.1 Data collection

This work has been entirely based on publicly available datasets. This choice is at firstly dictated by the fact that it is not easy to investigate a big cohort of subjects for a rare disease and secondly because it is likely that AGS causative mutations have a cell-type specific effect. In fact recently it has also been reported that different cell types have different roles in the onset and progression of AGS and thus a comprehensive analysis of different datasets, from different cellular sources, is needed (Giordano et al. 2022; Cuadrado et al. 2015). The first dataset used is from whole blood of AGS subjects with mutations in one of the different genes associated with AGS (Table 1). Moreover, the authors also provide RNA-seq data from stimulator of interferon genes (STING)-associated vasculopathy with onset in infancy (SAVI) disease and Type I interferon-mediated autoinflammation due to DNase II deficiency (IMD) syndrome. SAVI is an interferonopathy caused by gain-of-function mutations in the STING1 gene (Dai et al. 2020). IMD is an interferonopathy caused by mutations in the DNASE2 gene (Rodero et al. 2017). The dataset is composed of 43 samples: 6 subjects with mutations in TREX1, 3 with mutations in RNASEH2A, 7 with mutations in RNASEH2B, 5 with mutations in RNASEH2C, 5 with mutations in SAMHD1, 4 with mutations in ADAR1, 2 with mutations in IFIH1, 3 with mutations in STING1 and 3 with mutations in DNASE2 in addition to 5 healthy controls. Moreover, for a subset of subjects, it is also provided RNA-seq data sampled during treatment with a combination of three reverse transcriptase (RT) inhibitors (abacavir, lamivudine, and zidovudine). For these subjects, RNA-seq have been performed at 12 months after RT inhibitors treatment and at 6 months after treatment suspension (months 18) (project accession PRJEB41073 and PRJEB27958). The second dataset used is from 5 different B-cells subtypes derived from 9 SLE subjects and 12 healthy controls. Cells were FACS sorted and isolated according to B-cells surface antigens: resting naive (rN) (CD19+, IgD+, CD27−, MTG−, CD24+, CD38+), transitional 3 (T3) (CD19+, IgD+, CD27−, MTG+, CD24mid/+CD38 −), activated naive (aN) (CD19+, IgD+, CD27−, MTG+, CD24−, CD38−), isotype-switched memory (SM) (CD19+, IgD−, CD27+) and double-negative (DN) (CD19+, IgD−, CD27−, CXCR5−) (project accession PRJNA484963) (Scharer et al. 2019). The last dataset used is from iPSC cells derived

from AGS subjects with mutations in TREX1 or RNAEH2B genes and is composed by 4 AGS samples and 2 controls. The cells were extracted from subjects and in-vitro differentiated into NSC, astrocytes and neurons (project accession PRJNA797375,) (Table 4) (Giordano et al. 2022).

Raw fastq files were downloaded from ENA-EBI. Quality of the reads was assessed using fastqc. Reads from the PRJEB41073 dataset where additionally trimmed using trimmomatic (Bolger, Lohse, and Usadel 2014) to remove low-quality reads and illumina adapters.

Dataset	Cell type	mRNA type	Library type	Sequencing type	N of sample	N of subjects	treatment
PRJEB41073	Whole blood	polyA+	Stranded	Paired-end	43 (5 controls, 3 SAVI, 3 IMD, 32 AGS)	43 (5 controls, 3 SAVI, 3 IMD, 32 AGS)	None
PRJEB27958	Whole blood	polyA+	Stranded	Paired-end	27 (22 AGS and 5 controls)	13 (8 AGS and 5 controls)	RT inhibitors
PRJNA484963	B-cells subtypes	polyA+	Stranded	Paired-end	84	9 SLE and 12 controls	None
PRJNA797375	iPCS derived NSC, astrocytes and neurons	polyA+	Unstranded	Paired-end	18 (6 sample per cell type of which 2 control per cell type)	4 AGS and 2 controls	None

Table 4: Summary of dataset used. The table summarize the different datasets used in this work along with information on the type of mRNA used, the library preparation and the number of samples in each dataset.

4.2 Building reference genome and mapping

Reference human genome was downloaded from NCBI (version hg38, fasta format) and a genome index was generated using STAR (Dobin et al. 2013) (genomeGenerate mode). Reference gene annotation was downloaded from gencode (gtf format v37). Mapping was performed using STAR with default parameters (--twoPassMode Basic, --sjdbOverhang max(reads length) – 1). The percentage of mapped reads was retrieved from STARLog outputs files for each sample analyzed.

4.3 Reads per gene quantification

The number of reads per gene was quantified using the htseq (Anders, Pyl, and Huber 2015) software with default parameters and with –stranded parameter according to RNA-seq library type.

4.4 Differential gene expression analysis

Differential gene expression analysis was performed using DESeq2 (Love, Huber, and Anders 2014) with default parameters by comparing disease samples *vs* healthy controls. Briefly, the gene expression matrix, in which rows represent genes and columns samples, have been constructed using the "DESeqDataSetFromMatrix" function. Then the size factor and library size normalization have been performed with the "DESeq" function. Results were then extracted using the "results" function by comparing the disease samples *vs* healthy controls.

4.5 IFN score

The IFN-score was computed from the expression of 32 IFN-responsive genes as reported in Rice *et al.* (Rice et al. 2013). The gene list was retrieved from Giordano *et al.*, and is reported here (IFI27, IFI44L, IFIT1, ISG15, RSAD2, SIGLEC1, IFIT3, SERPING1, IFITM1, IRF7, STAT1, C1QA, CXCL10, EPSTI1, GBP1, HERC5, HERC6, IFI44L, IFIT2, IFI6, IFIT5, LAMP3, OAS1, MX1, OAS2, OAS3, OASL, PLSCR1, RTP4, SOCS1, SPATS2L, and USP18) (Giordano et al. 2022). Briefly, for each gene in each sample, a Z-score was computed as follows:

$$\frac{Gene\ expression - mean\left(Gene\ expression\ of\ CTR\right)}{standard\ deviation\left(Gene\ expression\ of\ CTR\right)}$$

were "Gene expression" is the normalized number of reads for a particular gene. The normalized number of reads were computed starting from raw reads counts per gene using the DESeq2 default median of ratios approach (using the counts function with the normalized = TRUE parameter) to account for both genes dispersion and different sequencing depth across samples. The IFN-score for each sample is represented by the

mean of the 32 genes Z-scores and is used as a representation of the level of activation of the IFN pathway.

4.6 PCA analysis

Samples PCA analysis were performed on gene expression data with the “plotPCA” function on vst transformed data using the top 500 genes showing the highest level of variance. Samples PCA for IR was computed on the IRratio of variable introns (variance >= 0.02 across all samples) using the “prcomp” R function with the parameters “scale” and “center” set to TRUE.

4.7 Alternative splicing quantification and differential alternative splicing analysis

Alternative splicing (AS) and differential AS analysis were performed using the rMATS software with default parameters (Shen et al. 2014). Differentially AS events were defined with an adjusted p-value < 0.05 and a deltaPSI >= 0.1. The PSI value it is defined as the ratio of the relative abundance of all isoforms containing a certain intron over the relative abundance of all isoforms of the gene. Events supported by less than 3 reads were discarded.

4.8 Intron retention quantification and IRscore quantification

Intron retention was quantified using the IRFinder software (Middleton et al. 2017) with default parameters. The number of retained introns in each dataset was counted according to the following filters (IRratio >= 0.1 and IntronDepth >= 3 and Warnings = “-” and No overlap with known features). The IRratio quantifies the portion of transcriptional activity traversing a given intron not removed by the splicing mechanisms. It is therefore a measure representing, for a given gene, the percentage of transcripts retaining the introns with respect to the total number of transcripts. For the AGS whole blood dataset the filter on IRratio was set to 0.01 because the analysis is performed on bulk RNA-seq comprised of a mixture of different cell types that can dilute the IRscore. The IRscore was expressed as the number of retained introns normalized on the total number of mapped reads and was used as an estimate of the global level of IR in each sample analyzed.

4.9 Identification of differentially retained introns

Differentially retained introns were identified comparing the IRratio distribution of 2 groups of samples analyzed using a t-test, similarly to what previously reported in (Ullrich and Guigó 2020). The IRratio of non-measurable introns (labelled as "LowSplicing", "LowCover", MinorIsoform") was substituted with a NA value and not considered in the t-test. The "LowSplicing" label indicates that there are not enough reads supporting the correct splicing, the "LowCover" label indicates that the sequencing depth for an event is too low and the "MinorIsoform" labels indicates that the event is not the major or most common splicing outcome among all transcripts of that gene (Middleton et al. 2017). Only introns correctly quantified in at least the 50% of the samples (non NA value) in each of the group under comparison were considered. P-values were corrected for multiple hypothesis testing using the "fdr" method and results were considered significant with an adjusted p-value < 0.1. Differentially retained introns in the PRJNA797375 was identified using the audic and claverie test (Tiňo 2009) as suggested by the IRFinder manual for dataset with a low number of biological replicates using the script *analysisWithLowReplicates.pl* (Middleton et al. 2017).

4.10 Enrichment analysis

Enrichment analyses were performed using the GprofileR R package (Kolberg et al. 2020) with default parameters using the REACTOME database (Fabregat et al. 2018). Every enrichment analysis was performed using as gene universe the list of expressed genes defined as genes passing the independent filters of DESeq2 and thus considered as expressed. Results were filtered for pathways with less that 1000 annotated genes to remove too general and uninformative pathways. Results with an FDR < 0.05 were considered statistically significant.

4.11 Correlation analysis

Correlations analyses were performed in R with the cor.test() function. P-values were then corrected for multiple hypothesis testing using the "fdr" method.

4.12 IR features and RNA binding motifs sites enrichements

IR hallmarks, defined as intron length, intronic GC content and splice sites strength were extracted with a custom python script using the Biopython library (Cock et al. 2009) and the hg38 human genome fasta sequence. Enrichment for specific RNA binding sites in retained introns was computed as Z-score by comparing the mean number of binding sites in retained introns against the mean of 100 randomizations of randomly chosen introns. The binding site sequences were retrieved from CISBP-RNA databse (Ray et al. 2013). Binding site sequences were searched in retained introns using regular expression taking advantage of the *re* python library. Splice site strength was assessed using MaxentScan (Yeo and Burge 2004). For assessing the strength of 5' splice sites was used the sequences composed by 3 exonic bases and 6 intronic bases while for 3' splice sites 3 exonic bases and 20 intronic bases as suggested in the software manual.

4.13 Proliferation score

The proliferation score have been computed using the GSVA R package (Hänzelmann, Castelo, and Guinney 2013) with default parameters. GSVA transforms the gene expression matrix from a gene by sample matrix to a gene-set by sample matrix, thereby allowing the evaluation of pathway enrichment for each sample. The GSVA package has been used to compute a score based on the expression levels of genes annotated in the "GO_BP_PROLIFERATION" ontology downloaded from the Gene ontology database (Gene Ontology Consortium 2004).

4.14 Bulk RNA-seq cell-type enrichment

The Bulk RNA-seq data cell-type enrichment was performed using the *xcell* software with default parameters (Aran, Hu, and Butte 2017). The gene signature was retrieved from xcell online database (source: https://xcell.ucsf.edu) and comprise 64 different immune and stromal cell types (figure 15). The gene expression input matrix was transformed into fragments per kilobase of transcript per million mapped fragments (FPKM) as suggested in the *xcell* software manual.

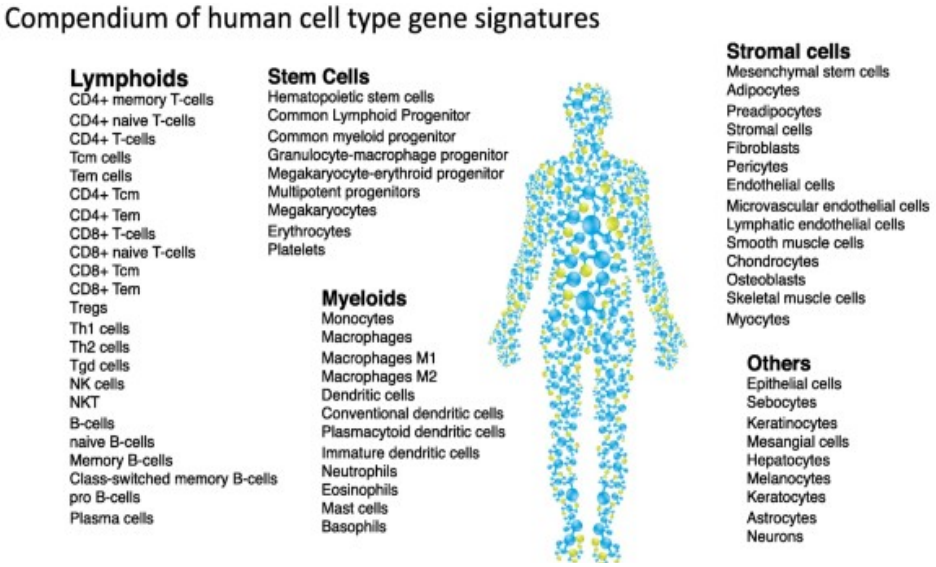


Figure 15: *xcell* 64 human cell type compendium

xcell compendium of 64 hman different cell types divided into 5 different cell-types families. Adapted from (Aran, Hu, and Butte 2017).

5 Results and Discussion

5.1 AGS, SAVI, IMD and SLE samples are characterized by an increased IFN signaling

AGS is an autoinflammatory and neurodegenerative disorder characterized by the accumulation of extrachromosomal DNA in the cytosolic fraction of cells. In the cell cytosol, the extrachromosomal DNA is firstly recognized by the cGAS protein that, upon DNA binding, is able to dimerize and through the STING pathway trigger the transcription of interferon (IFN) (Yu and Liu 2021) which is thought to be involved in AGS onset and progression (Rice et al. 2013). Nowadays the precise mechanisms behind the extrachromosomal DNA generation are still only partially identified but it has been hypothesized to be supported by LINE1 reverse transcriptase (RT) activity (Thomas et al. 2017). In this scenario, treatment of AGS cells with RT inhibitors have been identified as good candidate for decreasing the IFN signaling, therefore confirming RT activity as one of the mechanism for extrachromosomal DNA generation (Rice et al. 2018). IFN signaling is primarily controlled at transcriptional level through the activity of different inducible transcription factors that can modulate the abundance of IFN and IFN stimulated genes (ISGs) transcripts (H. Li et al. 2007). In this scenario, I aimed to explore the AGS, SAVI, IMD and SLE transcriptome to gain insight into the IFN gene dysregulation by using subjects RNA-seq data.

5.1.1 AGS subjects are characterized by aberrant IFN gene expression that is influenced by the administration of RT inhibitors

To quantify the IFN pathway activation in AGS subjects I have performed gene expression quantification from RNAseq data. This allows me to quantify the ISGs expression and assess the impact of RT treatment on ISGs expression levels. In this scenario, treatment with RT inhibitors can decrease ISGs expression by reducing the amount of accumulated extrachromosomal DNA that, in turn, triggers the activation of the IFN response (De Cecco et al. 2019). To understand the impact of RT treatment on ISGs expression 8 AGS subjects were orally administrated with a combination of three nucleoside analogue RT inhibitors (abacavir, lamivudine, and zidovudine) for 12

months. Blood collection and RNA-seq have been performed at 0 months (untreated), 12 months after RT inhibitors treatment (treated), and at 6 months after treatment suspension (18 months, post-treatment). PCA analysis on gene expression of AGS subjects and healthy controls highlights a difference between AGS and control samples but is not able to separate untreated AGS from treated AGS samples (figure 16A). This suggests that gene expression levels in AGS untreated subjects differ from controls but are more similar to those observed in AGS treated subjects, supporting the idea that RT inhibitor treatment does not have a global impact on gene expression levels but may indirectly influence the expression of a specific subset of genes. Interestingly, PCA analysis performed on the selected 32 genes belonging to the IFN signature, is able to separate treated AGS subjects from untreated ones suggesting a possible influence of RT inhibitors treatment on the expression of the 32 ISGs and therefore on the IFN response (figure 16B). The heatmap reported in figure 16C, performed on the 32 IFN-stimulated genes, shows that AGS treated and healthy controls cluster in a separate branch with respect to AGS untreated and post-treatment subjects further confirming that RT inhibitors treatment influence the expression levels of ISGs. This result suggests that AGS gene expression signature differs from those of healthy controls and that the treatment with RT inhibitors causes only minor, but crucial, changes in gene expression. These changes are not reflected in the global gene expression signature but mainly affect the expression of ISGs and therefore inflammation which, in subjects under treatment with RT inhibitors, is strongly reduced as reported in the literature (Rice et al. 2018).

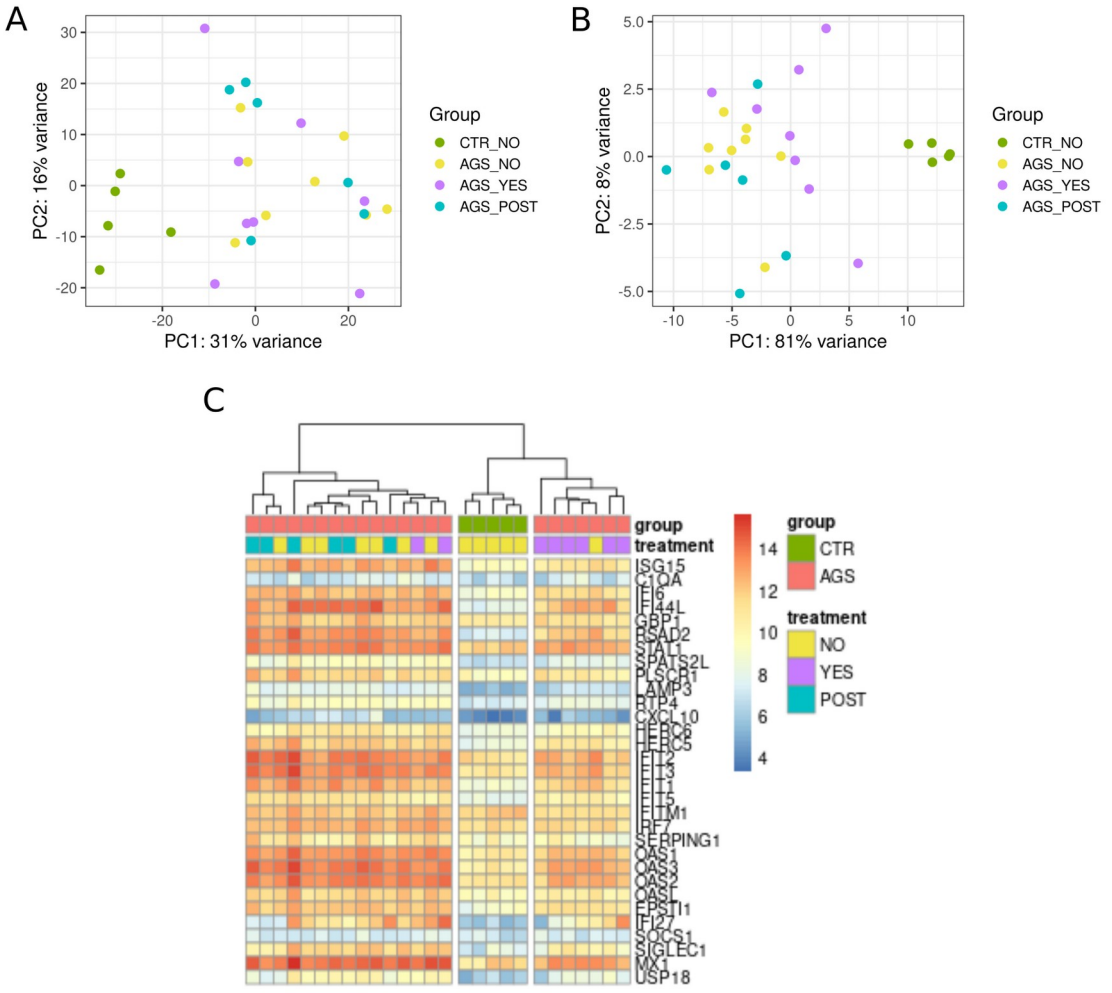


Figure 16: AGS samples are characterized by aberrant IFN response that is influenced by treatment with RT inhibitors. (A) PCA analysis on vst corrected gene expression data from AGS subjects untreated, treated or 6 months after treatment. The treatment consist in the oral administration of RT inhibitors. CTR_NO = untreated controls, AGS_NO = untreated AGS samples, AGS_YES = treated AGS samples, AGS_POST = AGS samples 6 months after treatment suspension. (B) PCA analysis on vst corrected gene expression data of 32 selected IFN-stimulated genes. (C) Heatmap with dendrogram of vst corrected gene expression data of 32 selected IFN stimulated genes in AGS subjects at 3 different treatment stages and healthy controls.

5.1.2 AGS subjects are characterized by gene expression dysregulation that recapitulates the major symptoms of the disease

To identify gene expression changes in AGS samples I have performed differential gene expression analysis on a larger cohort of subjects affected by different interferonopathies comprising AGS, SAVI and IMD subjects. Differential gene expression analysis, comparing AGS samples *vs* healthy controls, reveals 1629

upregulated and 630 downregulated genes (according to an adjusted p-value < 0.05 and |log2FoldChange| > 0.58) (figure 17A). Interestingly the most significant upregulated gene is the RSAD2 gene, an IFN inducible antiviral protein, that plays fundamental roles in the innate immune response (Fitzgerald 2011). Moreover, I have also observed a significant upregulation of other important IFN-stimulated genes like IFI44 and IFI44L (Busse et al. 2020) that are known to increase their expression upon IFN stimulation. Furthermore, I have observed that the majority of genes belonging to the IFI gene family are upregulated (Landolfo et al. 1998). The IFI gene family is composed of genes whose expression is known to be induced by IFNs and can participate in many different regulatory functions among which cell growth, immunomodulation, host resistance to tumors and viral infections (Asefa et al. 2004). High expression levels of ISGs have been reported in SLE and other inflammatory diseases may represent an indication of increased IFN signalling and cytosolic DNA accumulation (Baechler et al. 2003; Choubey et al. 2010). Taken together these results confirm an over-activation of the IFN response in AGS subjects whole blood samples. Among the most significant downregulated genes I have observed CPSF7, a subunit of the cleavage factor I 3′ end processing complex, involved in alternative splicing suggesting a possible impairment of the splicing process in AGS whole blood samples (Y. Zhu et al. 2018). To shed light upon the deregulated pathways in AGS I have performed pathways enrichment analysis on both upregulated and downregulated genes using the R package GprofileR and the REACTOME pathway database (adjusted p-value < 0.05, ordered query false) (Kolberg et al. 2020; Fabregat et al. 2018). Pathway enrichment analysis reveals that “interferon alpha/beta signaling” and “interferon signaling” (adjusted p-value = $1.66*10^{-10}$ and $1.84*10^{-14}$ respectively) are the most significant pathways enriched among upregulated genes (figure 17B). Moreover is interesting to notice that also the "cell cycle checkpoints” pathway is significantly enriched among upregulated genes suggesting a possible imbalance in cell cycle progression in AGS subjects (adjusted p.value = $1.45*10^{-7}$). These findings are consistent with the literature: it has been reported that several AGS causative genes may impact cell cycle progression (Yang, Lindahl, and Barnes 2007; Batalis et al. 2021, 1; F. Zhang and Rabinovici 2007). Regarding the downregulated genes, they result enriched in the “circadian clock” pathway (adjusted p-value = 0.003), “neuronal system” (adjusted p-value = 0.03) and “chromatin organization” (adjusted p-value =

0.03) (figure 17C). Interestingly the downregulation of neuronal genes can be in agreement with some of the neurological symptoms observed in AGS subjects like the loss of motor and cognitive skills and intellectual disability. The fact that the blood transcriptome can reflect some of the neurological abnormalities has been discussed in the case of Alzheimer and other related neurodegenerative diseases (Naughton et al. 2015; Maes et al. 2007). Indeed many of the significantly altered genes in the blood of Alzheimer subjects are known to be critical for neurological functions (Naughton et al. 2015). The enrichment of downregulated genes in the "chromatin organization" pathway can be in line with recent evidence that cells from AGS subjects are characterized by a global epigenetic imbalance characterized by genome-wide loss of methylation (Lim et al. 2015). Finally, an increasing body of evidence has related autoimmunity and neurodegeneration with the activity of circadian clock genes (Shivshankar et al. 2020; X.-L. Wang and Li 2021; Xiang et al. 2021). Taken together, these results confirm that AGS is characterized by an aberrant IFN activation representing a fundamental hallmark of the disease. Interestingly type I IFN can be a damaging factor for the developing central nervous system (Y J Crow et al. 2014). This evidence candidate the aberrant IFN signaling as a primary factor for the neurological symptoms observed in AGS (Y J Crow et al. 2014). In fact I have also observed that downregulated genes are significantly enriched in pathways related to the neuronal system, in agreement with the neurological symptoms observed in AGS. These results then candidate the IFN production as a fundamental player in AGS etiology confirming and extending what has already been reported about this disease.

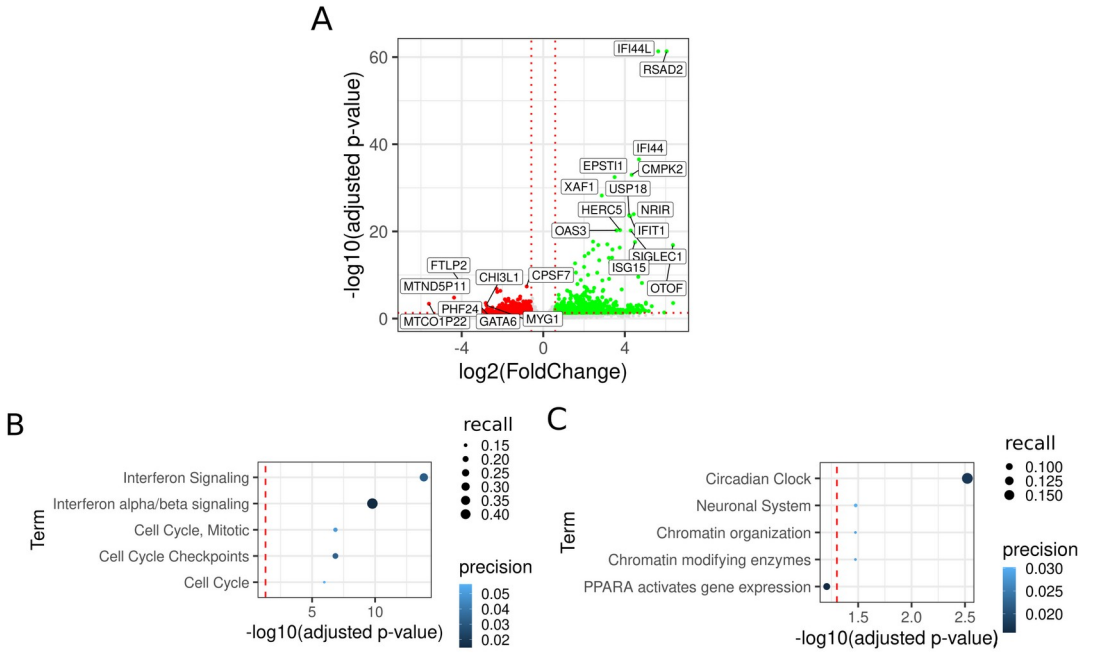


Figure 17: AGS samples are characterized by gene expression alterations reflecting the phenotypic hallmark of the disease. (A) Volcano plot reporting the differentially expressed genes between AGS and control samples. Green points represent the significantly upregulated genes while red point the significantly downregulated genes. (B) Enrichment analysis on upregulated genes in AGS samples. The x-axis reports the -log10 of the adjusted p-value. The point dimension represents the recall while the precision is represented by the color of the point. The recall indicates the proportion of functionally annotated genes that the query recovers while the precision represents the proportion of genes in the input list that are annotated to the function. (C) Enrichment analysis on downregulated genes in AGS samples.

5.1.3 AGS, SAVI and IMD are accomunated by a positive IFN signature

To specifically quantify the IFN signaling from gene expression data and to obtain a reliable and precise quantification of the IFN pathway activation I have computed the IFN-score based on the expression level of 32 selected IFN-related genes (see Materials and Methods). This score can be used as a quantitative index for sample-specific IFN response. Interestingly, while comparing untreated, treated, and 6 months after RT treatment AGS subjects, I have observed that the treatment with an RT inhibitor significantly reduces the IFN response, with a subsequent re-increase of the IFN-score at 6 months following treatment discontinuation (figure 18A and 18B). Anyway, the treatment with RT inhibitors is not able to completely restore the IFN-score to wild-type levels probably suggesting that other players, in addition to the RT

activity, are involved in the stimulation of IFN production in AGS (figure 18B). This result confirms that AGS subjects are characterized by an increased IFN response that is reduced by RT inhibitor treatment suggesting an involvement of RT activity in triggering the IFN signaling. Anyway, as observed in figure 18B, the treatment with RT inhibitors does not reverse the IFN score to control levels suggesting that probably the induction of the aberrant IFN response is multifactorial. In fact, in addition to RT activity, also DNA damage and DNA repair pathways have been proposed to participate in the generation of extrachromosomal DNA that can, in turn, stimulate IFN production (Giordano et al. 2022; Cristini et al. 2022). To quantify the IFN signaling in other interferonopathies, sharing common tracts with AGS, I have extended the analysis to a bigger cohort composed by AGS, SAVI and IMD subjects. The PCA analysis performed on the top 500 variable genes show a certain separation between AGS and controls samples while the other interferonopathies cluster together with controls (figure 18C). Conversely, PCA analysis performed on the expression levels of the 32 IFN-stimulated genes reveal a clear separation between controls samples and samples derived from the other interferonopathies confirming that the IFN pathway is altered in all AGS, SAVI and IMD subjects (figure 18D). To quantify the IFN signaling in AGS, SAVI and IMD subjects I have computed the IFN-score. Figure 18E reports the distribution of the IFN-score in all samples analyzed. I have observed a significant increase in the IFN-score distribution in AGS, SAVI and IMD groups with respect to healthy controls, with the IMD group characterized by the highest IFN-score observed among all groups analyzed. This result confirms the presence of an increased IFN response in all the interferonopathies analyzed confirming the pivotal role of IFN signaling in those pathologies. The identification of a heightened IFN response in those samples suggests the presence of commonalities among those pathologies and propose that their concomitant study can help the identification of novel molecular players at the basis of those diseases.

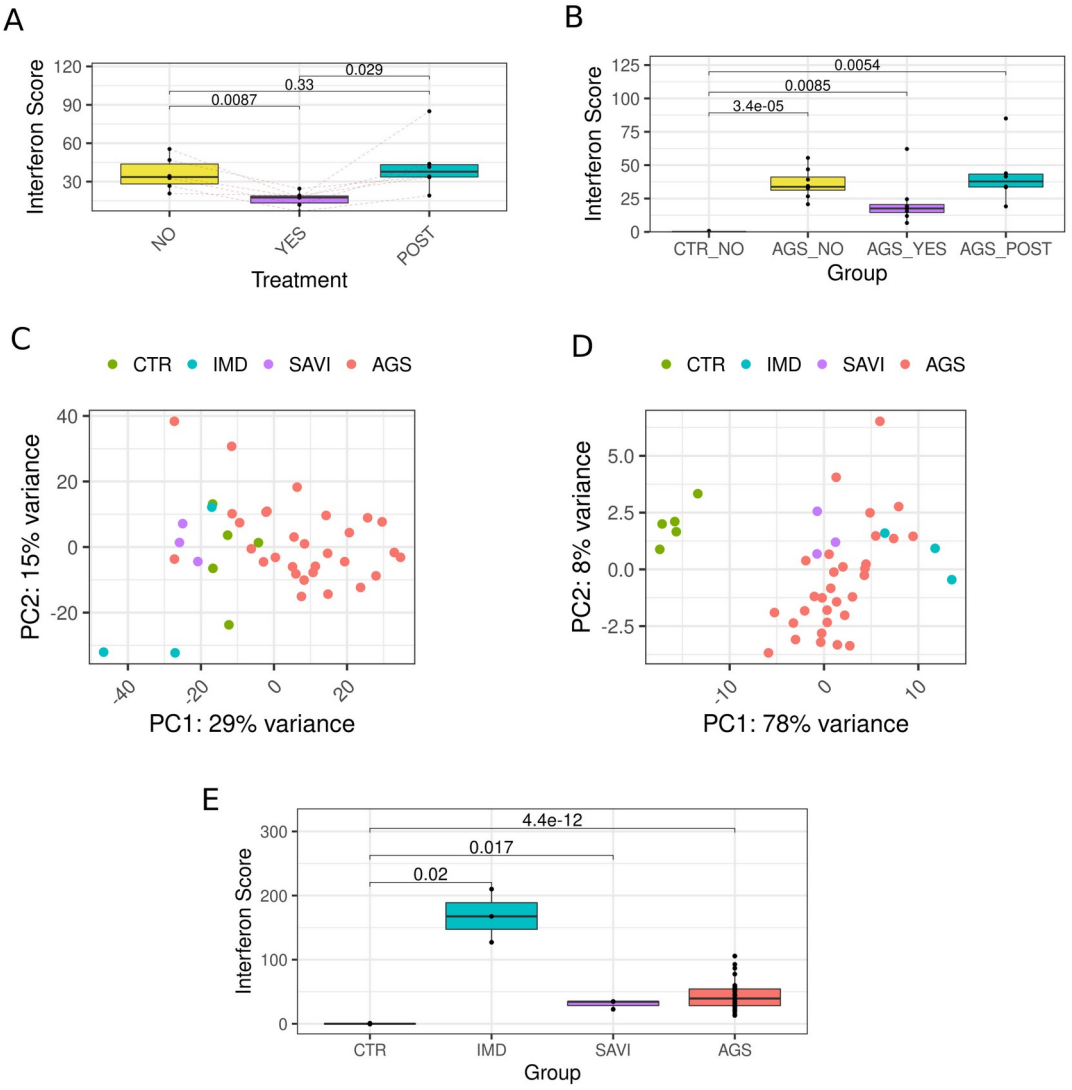


Figure 18: AGS, SAVI and IMD are accomunated by a positive IFN signature. (A) IFN-score distribution in AGS subjects untreated, treated and 6 months after RT inhibitor treatment. The brown dashed lines connect the same subject at the 3 time points. (B) IFN-score distribution in AGS subjects at 3 different time points and healthy controls. (C) PCA on vst corrected gene expression data in control samples, AGS, SAVI and IMD samples (F) PCA on vst correceted gene expression data of 32 selected IFN stumulated genes in controls, AGS, SAVI and IMD subjects. (E) IFN-score distribution in AGS, SAVI, IMD and CTR subjects.

5.1.4 SLE B-cells subtypes samples are characterized by increased IFN and ISGs expression

To explore the IFN gene expression signature in a context similar to AGS, I have decided to analyze RNA-seq data from FACS-sorted B-cell subtypes of SLE subjects. SLE shares many common tracts with AGS, in particular for the autoimmune component of the disease, like an aberrant intracellular nucleic acids metabolism. In fact recent findings have revealed that the induction of autoimmunity can be initiated, also in SLE subjects, by immune recognition of endogenous nucleic acids originating from processes such as DNA replication, DNA repair or endogenous retroelements activity (M A Lee-Kirsch, Wolf, and Günther 2014). This evidence candidate SLE as a good model to deepen and extend the dysfunctional molecular mechanisms of AGS. Moreover, this dataset is particularly important given the presence of FACS-sorted isolated B-cell subtypes that can help to identify, in the mixture of blood cells, some important cellular players. The FACS-isolated B-cell subtypes are peripheral blood progenitors cells (PbPc), resting naive (rN), Transitional B cells (T3), activated naive (aN), switched memory (SM) and double negative (DN). PbPc comprise hematopoietic stem cells that can differentiate into all types of blood cells, including lymphoid cells (J. Y. Lee and Hong 2019). rN cells are naive B-cells that have matured and entered the bloodstream but have not yet been exposed to the antigen-presenting cells (APCs) specific to their B cell receptor (Sanz et al. 2019). T3 B-cells represent the state between resting and activated B-cells. Even if classified as the state immediately before B-cells activation the precise role of T3 cells remains yet unknown. It has been proposed that they can represent the immediate precursor of mature B-cells (Sanz et al. 2019). Interestingly T3 cells seem to represent early activated naïve B cells in active SLE (Scharer et al. 2019). aN B-cells are cells that have been activated, in the secondary lymphoid organs, upon APCs interaction. They are capable of producing antigen-specific immunoglobulin (Sanz et al. 2019). SM B-cells have undergone the class-switch recombination, a process that changes the immunoglobulin type without affecting the antigen specificity thus allowing the cells to modulate the response upon antigen binding (Sanz et al. 2019). Finally DN B-cells constitute a rare population lacking the expression of immunoglobulin D and CD27 memory markers. These cells play important roles in autoimmune and infectious diseases and have been reported to be particularly abundant in SLE and other autoimmune diseases (Scharer et al. 2019;

Sanz et al. 2019). By computing the IFN-score I have observed a significant increase in all the different B-cell subtypes except for PbPc cells (adjusted p-value < 0.05) (figure 19A). This result confirms that B-cells are a source of IFN in SLE subjects with the strongest IFN activation in the DN cells, which represent the last stage of differentiation available in this dataset. The molecular commonalities and the presence of an increased IFN response suggest the possible presence of molecular similarities among AGS and SLE that will be further explored in the following chapters.

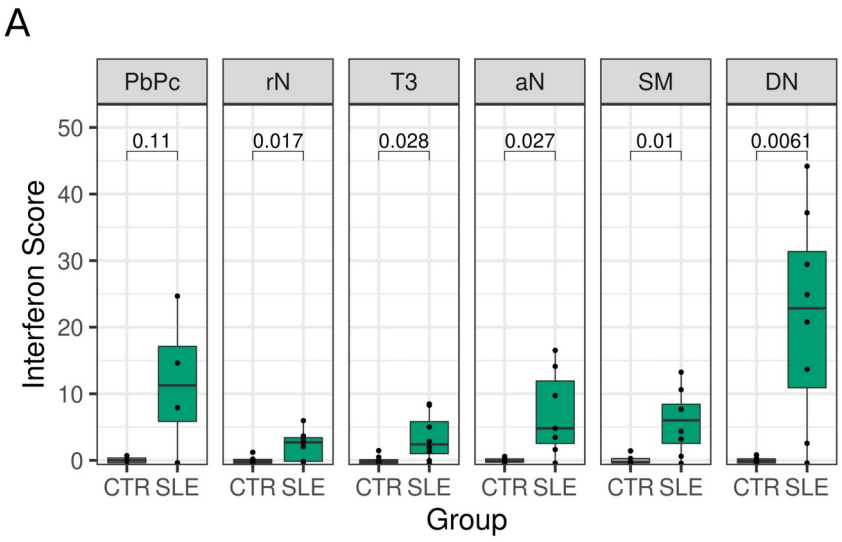


Figure 19: SLE subjects are characterized by increased IFN signaling. (A) IFN-score distribution in SLE and healthy controls (CTR) for the different B-cell subtypes analyzed.

5.1.5 IFN signaling is mainly supported by astrocytes in AGS neuronal samples

Given the known role of immune system cells in supporting IFN response I wondered whether, in tissue derived by AGS subjects, other cell types are able to sustain IFN production. Recently it has been proposed that the aberrant IFN signaling in AGS is supported by specific cell types that can, in turn, influence the development of other cells present in the same cellular milieu (Cuadrado et al. 2015). In fact IFN plays a central role in driving an antiviral state also in non-immune cells and it has been reported that IFN stimulation has profound effects on all mammalian cells (E. Roy and Cao 2022; Murira and Lamarre 2016). This effect is particularly important and studied

during neuron development and have been hypothesized to mediate neuronal loss and synaptic damage. For example in Alzheimer disease IFN stimulation can increase amyloid precursor protein cleavage, through transmembrane protein 3 (IFITM3) induction, resulting in increasing amyloid pathology (E. R. Roy et al. 2022). Interestingly, the selective elimination of the IFN receptor from neuronal cells reduces amyloid beta plaques, linking the increased IFN signaling to neurodegeneration (Hur et al. 2020). To quantify the IFN activation in a disease-relevant neuronal context and to identify the cell types capable of sustaining the IFN production in neuronal tissue derived from AGS subjects I have retrieved bulk RNA-seq data from the study of Giordano *et al.* (Giordano et al. 2022). The authors provide polyA+ RNA-seq data from healthy donor and subject-derived (with RNASEH2B and TREX1 mutations) induced pluripotent stem cells (iPSC) *in vitro* differentiated into neural stem cells (NSC), astrocytes and neurons. To identify the possible gene expression dysregulation in NSC, astrocytes and neurons I have performed differential gene expression analysis by comparing the AGS samples *vs* healthy controls for each cell type. The PCA analysis performed on gene expression for all the samples analyzed reveal that the cell type is the most important factor in clustering. In fact the PCA reveal 3 clearly separated clusters defined by the 3 cell types analyzed (figure 20A). Differential gene expression analysis reveals a general gene expression deregulation in all the 3 cell types analyzed with astrocytes and neurons having a higher number of deregulated genes with respect to NSC cells. I have indentified 251 upregulated and 255 downregulated genes in AGS astrocytes, 573 upregulated and 904 downregulated genes in AGS neurons and 29 upregulated and 7 downregulated genes in NSC (figure 20B and 20C and 20D respectively). These results may suggest that AGS NSC does not harbour strong gene deregulation in AGS samples suggesting that the diseases cellular phenotypes are probably acquired at a later developmental stage, even if it is necessary to analyze a larger dataset to mitigate the effect of possible differences in expression variability of the different cell types (Giordano et al. 2022). Interestingly astrocyte is the only cell type, among those analyzed here, characterized by a significant upregulation of IFN-responsive genes. In fact pathway enrichment analysis of genes significantly upregulated in AGS astrocytes reveal “Interferon alpha/beta signaling” and “Cytokine Signaling in Immune system” as the most significantly enriched pathways (adjusted p.-value = $5.53*10^{-13}$ and $1.85*10^{-09}$ respectively) (figure

21A) while in AGS neurons and AGS NSC there are no significant enriched pathways related to IFN (figure 21A-B). To specifically assess the level of IFN signaling in AGS NSC, astrocytes and neurons I have computed the IFN-score. The results confirm that the strongest IFN-score increase is observed in astrocytes cells (p.value = 0.03) while in NSC cells, even if significant changes are observed, the magnitude of the variation is much lower compared to what was observed in astrocytes (figure 21C). Interestingly, AGS neurons reports a significant decrease in the level of ISGs gene expression compared with their healthy counterparts (figure 21C). Like other cell types of the central nervous system, neurons are capable of expressing IFN and ISGs, for example after a viral infection, even if with lower level compared to astrocytes (Lang et al. 2022). In this scenario the reduced IFN-score observed in AGS neurons remains yet to be understood. Interestingly, treating neurons with conditioned media from AGS astrocytes caused neuronal toxicity that recapitulates the molecular changes observed in AGS samples (Sase et al. 2018). Taken together these results suggest that IFN signaling is sustained by specific cell types, particularly astrocytes, in AGS-derived neuronal tissue.

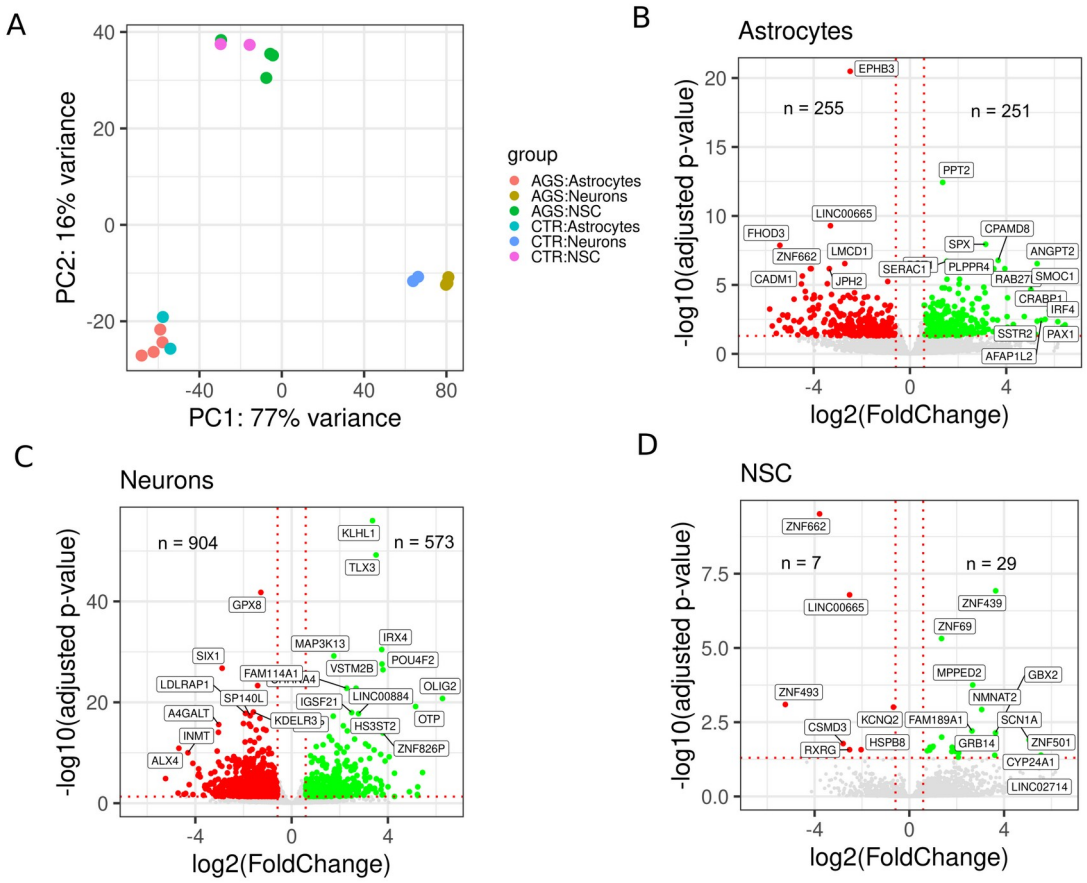


Figure 20: AGS gene expression alterations are cell-type specific. (A) PCA on vst corrected genes expression of neurons, astrocytes and NSC from AGS subjects-derived iPSC cells. Volcano plot of deregulated genes in (B) AGS astrocytes, (C) AGS neurons and (D) AGS NSC. The green points represent the significantly upregulated genes while the red points the significantly downregulated one.

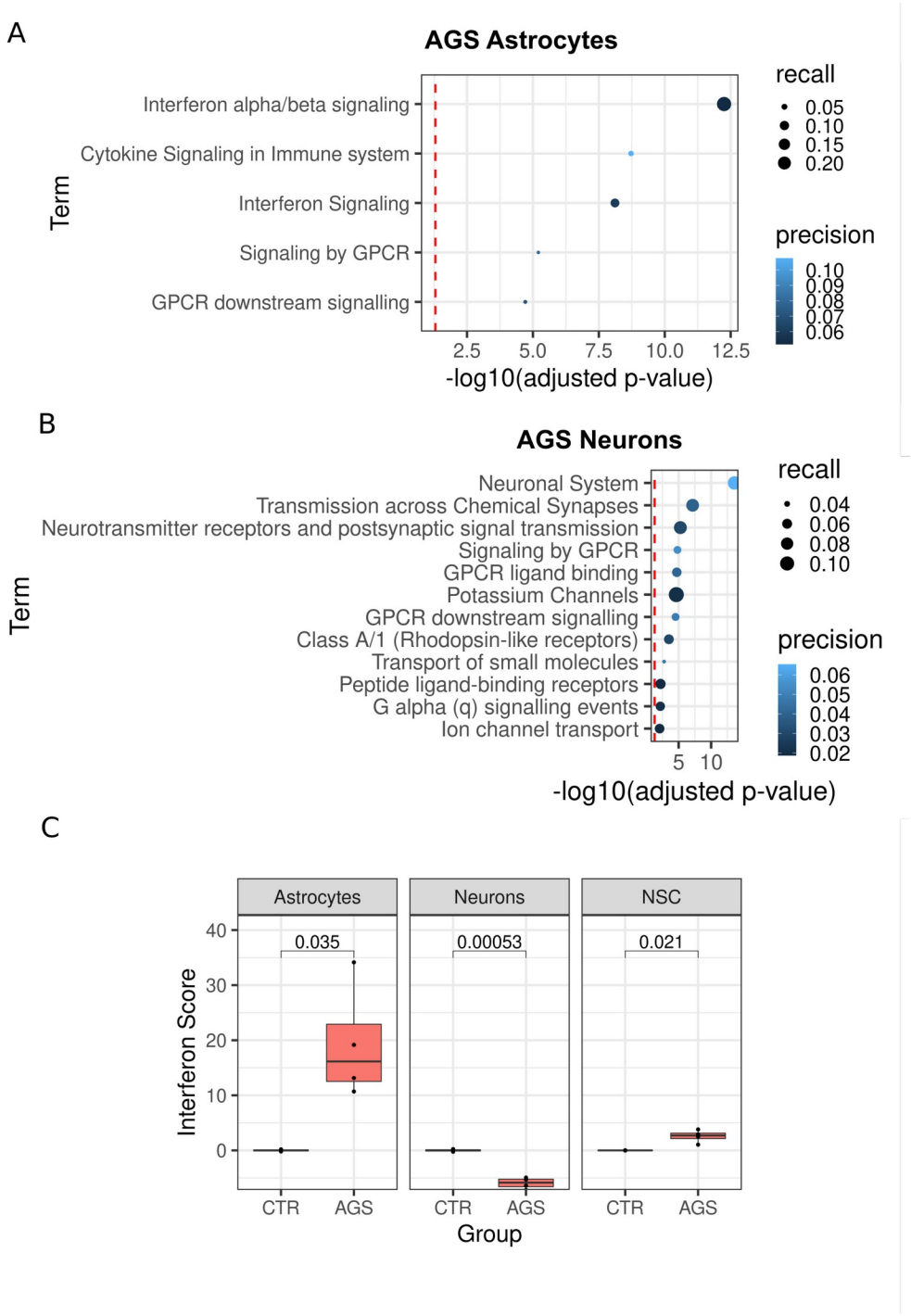


Figure 21: AGS gene expression alterations are cell-type specific. (A) Enrichment analysis of upregulated genes in AGS astrocytes. (B) Enrichment analysis of upregulated genes in AGS neurons. The x-axis reports the -log10 of the adjusted p-value. The point dimension represents the recall while the precision is represented by the color of the point. The recall indicates the proportion of functionally annotated genes that the query recovers while the precision represents the proportion of genes in the input list that are annotated to the function. (C) IFN-score distribution in AGS and control samples from astrocytes neurons and NSC cells.

As also observed by other studies, my results confirm that AGS, SAVI, IMD and SLE blood samples are characterized by an increased expression of IFN and ISGs. Interestingly, by extending the analysis to AGS astrocytes, neurons and NSC, I have observed that the enhanced IFN signaling is sustained by particular cell types. IFN signaling is especially elevated in cells from whole blood and astrocytes while in neurons and NSC is more similar to healthy controls, or even lower, as in the case of neurons. These findings confirm that the IFN pathway is a key player in AGS, SAVI, IMD and SLE and that the ISGs upregulation follows a cell-type specific pattern suggesting that the inflammation may be sustained by particular cell types in AGS and other related diseases (Giordano et al. 2022; Cuadrado et al. 2015; Rice et al. 2018).

5.2 Alternative splicing is impaired in AGS samples

Alternative splicing (AS) is a fundamental mechanism in Eukaryotic cells capable to increase transcriptome complexity through the production of different transcripts from a single gene (Y. Wang et al. 2015). AS is primarily regulated by the activity of different proteins collectively called splicing factors that are able to finely tune the 5 different AS events till now described (Y. Wang et al. 2015). Besides this, many other factors like chromatin status, RNA secondary structure and transcription rate can contribute to modulate AS (L.-Y. Zhu et al. 2018; Muniz, Nicolas, and Trouche 2021). Interestingly AS is involved in immune system development and activation (Su and Huang 2021) and recently it has been reported that AS alterations are involved in the onset of different autoimmune diseases making it an interesting topic to be studied in the context of AGS and SLE (P. Ren et al. 2021; Lenzken, Loffreda, and Barabino 2013). The continuous IFN signaling observed in AGS, SAVI, IMD and SLE subjects is an important stimulator of immune system activation, that is, in turn, intimately connected with AS changes (Su and Huang 2021; Murira and Lamarre 2016). Since a comprehensive characterization of AS in AGS has not been produced yet, and keeping into account the involvement, in the splicing mechanism, of specific genes I have observed downregulated in AGS, I decided to characterize AS in AGS samples to identify possible alterations related to disease onset and progression.

To obtain a global picture of the AS landscape from RNA-seq data I have used the rMATS software (Shen et al. 2014). rMATS is able to quantify all the 5 different AS events from RNA-seq data and to perform a differential AS events abundance analysis by comparing AGS samples *vs* healthy controls. As observed in Figure 22A-E AGS samples are characterized by a global differential AS program with both increase or decrease in specific AS events with respect to healthy controls. Interestingly IR is the AS event with the strongest increase in AGS with respect to control samples (figure 22F). IR has been characterized as a mechanism for controlling gene and protein expression during immune system and neuron differentiation, which represent crucial processes impaired AGS and SLE (Jacob and Smith 2017; Wong et al. 2013; Ullrich and Guigó 2020; Green et al. 2020; Yap et al. 2012). These results led me to further investigate and characterize the IR changes in order to try to better understand the molecular consequences of IR dysregulation in autoimmune diseases like AGS and SLE.

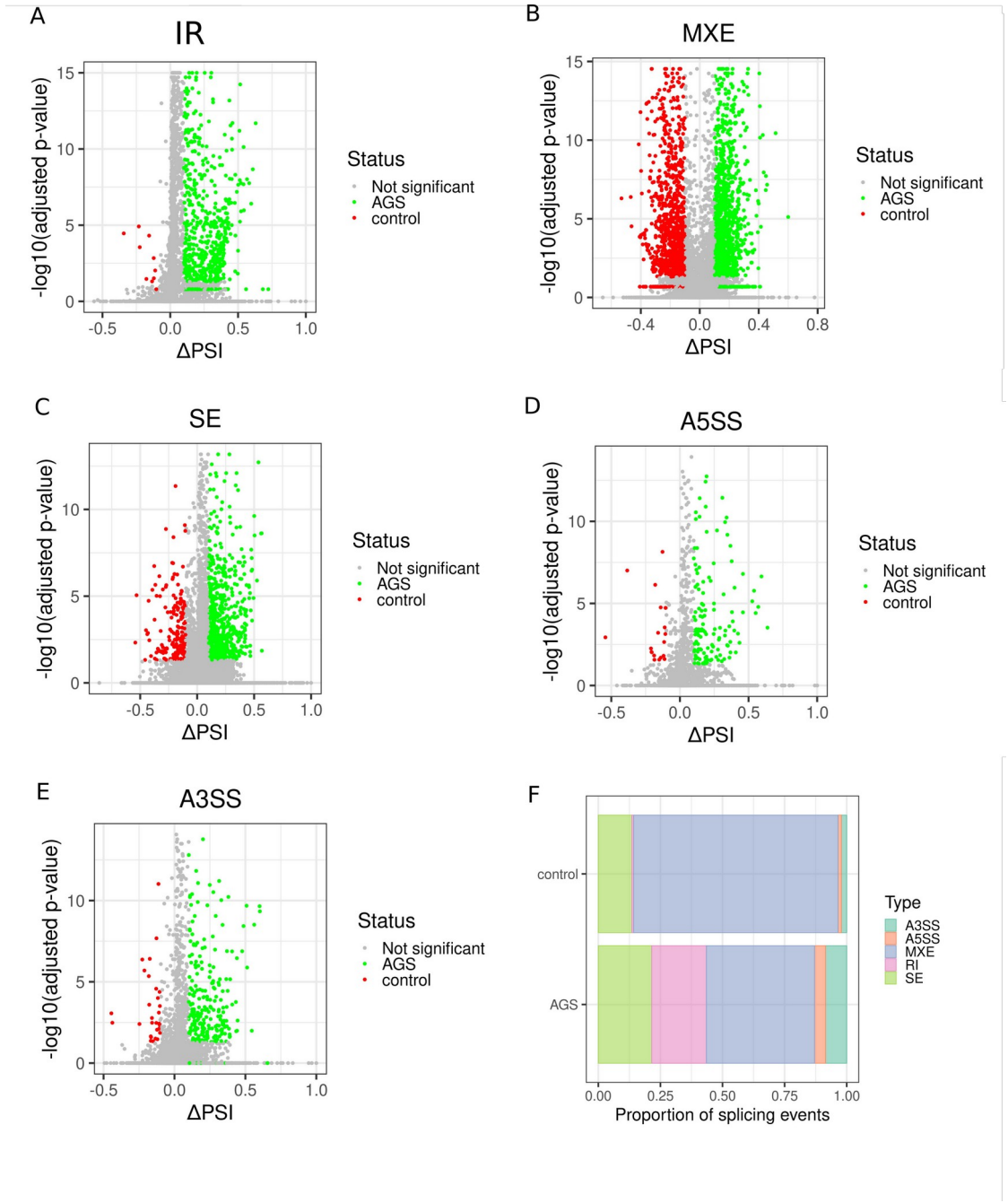


Figure 22: Alternative splicing is impaired in AGS samples and IR changes are the most abundant among all AS events. Volcano plot of (A) Differentially retained introns, (B) Mutually exclusive exons, (C) Skipped exons, (D) Alternative 3' splice site, (E) Alternative 5' splice site. In red are reported the AS events with significantly decreased abundance in AGS while in green the events with significantly increased abundance in AGS (adjusted p.value < 0.05, minimum coverage >= 3 reads). (F) Percentage of significant de-regulated AS events with respect to total de-regulated AS events in AGS and control samples.

5.2.1 AGS and SLE samples are characterized by increased IR which is not directly associated with IFN signaling

IR is nowadays recognized as an important AS mechanism due to its reported role in gene expression control and protein homeostasis among others (Zheng et al. 2020). The study of IR is complicated by different confounding factors that can prevent a precise IR quantification (Broseus and Ritchie 2020). To precisely quantify IR I have used the IRFinder software (Middleton et al. 2017). IRFinder is able to quantify the level of retention of every intron in every annotated and expressed gene. Introns overlapping other exons on the same strand were annotated and removed from the analysis as their quantification can be inflated by reads coming from expressed overlapping exons. IRFinder is able to reliably quantify 262,040 out of 268,574 introns in the AGS whole blood dataset. To quantitatively explore the extent of IR in AGS samples I have quantified the number of retained introns per sample and normalized it on the total number of mapped reads as the number of detectable introns significantly depends on the number of mapped reads (figure 23A). This value will be referred to as the IRscore and represent an indication of the global level of IR in a given sample regardless of the identity of retained introns. As we can observe in figure 23B AGS samples are characterized by a significant (adjusted p.value = $8.1 * 10^{-8}$) global increase in IR with respect to healthy controls. I have not observed a global IR increase in IMD and SAVI samples suggesting that aberrant IR can be a specific hallmark of AGS among the different interferonopathies analyzed (figure 23B). Interestingly, an increased IR has been also reported in other neurodegenerative and autoimmune diseases, like amyotrophic lateral sclerosis and Alzheimer, confirming the possible involvement of this process in the onset and progression of other related pathologies (Adusumalli et al. 2019; H.-D. Li et al. 2021; Luisier et al. 2018; Q. Wang et al. 2020). AGS is characterized by both an immunological and a neurological component, implying that the study of IR should be expanded to include other AGS-relevant cell types (Cuadrado et al. 2015; Giordano et al. 2022). To further extend the analysis to other AGS-relevant cell types I have decided to analyze RNA-seq data derived from astrocytes, neurons and NSC *in vitro* differentiated from AGS subjects cells and compare them to their healthy counterparts. Interestingly I have observed different IRscore in the 3 cell types analyzed. In particular, AGS astrocytes are

characterized by a significantly increased IRscore with respect to control astrocytes (p-value = 0.04). Concerning AGS neurons and NSC I have not observed a significant increase in the IRscore with respect to their healthy counterparts although the data suggest a trend toward an increase. More samples are needed to asses the upregulation of IR also in AGS neurons and NSC beyond any reasonable doubt. These results suggest that the increased IR observed in AGS whole blood samples may be further extended to other relevant cell types (figure 23C). Finally, given the phenotypic and genetic overlap between AGS and SLE I have further included in the analysis 5 different subject-derived FACS-sorted SLE B-cell subtypes. IRscore quantification in healthy subjects reveals that PbPc cells retain a lower number of introns compared to the other stages of B-cell differentiation suggesting that IR changes may be modulated during healthy B-cells maturation. Following differentiation from PbPc into rN cells the IRscore strongly increase and then start to decrease till DN cell in healthy controls, as also previously observed by Ullrich and Guigò (Ullrich and Guigó 2020). In SLE samples I have not observed the same decrease in IRscore which suggests that the level of IR remain higher during all differentiation stages compared to their healthy counterparts (figure 23D). In fact, by comparing the IRscore of SLE subject *vs* healthy controls, at each developmental stage, I have observed a significant increase in the IRscore at the T3, SM and DN stages, with the DN stage being the most affected by increased IR in SLE (p-values: T3 = 0.03, aN = 0.06, SM = 0.01, DN = 0.002) (figure 23E). This result suggests that SLE B-cells are probably characterized by an increased IR with higher differences in the final differentiation steps. Taken together these results reveal a global increase of IR in AGS and SLE samples both from whole blood samples and from neuronal samples, in particular in astrocytes, suggesting that increased IR may be associated with AGS and SLE pathologies. It is worth noting that IRscore variations do not follow those of the IFN-score suggesting that the increased IR observed in AGS and SLE samples cannot be explained by IFN signaling alone. In fact, SAVI and IMD samples, both characterized by an aberrant IFN response, do not show a significant increase in the IRscore when compared to healthy controls (figure 23B). To confirm this result and to understand if IR is associated with increased IFN and ISGs expression I have quantified the IRscore in AGS samples treated with RT inhibitors that, as previously reported, are capable to restore IFN gene expression to near wild-type levels. The result indicates that treatment with RT inhibitors has no

effect on the level of intron retention further suggesting that the increased IR is not a direct consequence of the increased IFN signaling observed in AGS samples (figure 23F). Given the previous observations I can conclude that both AGS and SLE whole blood samples are characterized by an increased IR compared to controls and that these results may be extended also to AGS neuronal samples. Finally, the increased IR observed in AGS and SLE samples it is not directly due to the increased IFN signaling and further analyses are required to understand the molecular basis of increased IR in AGS and SLE samples.

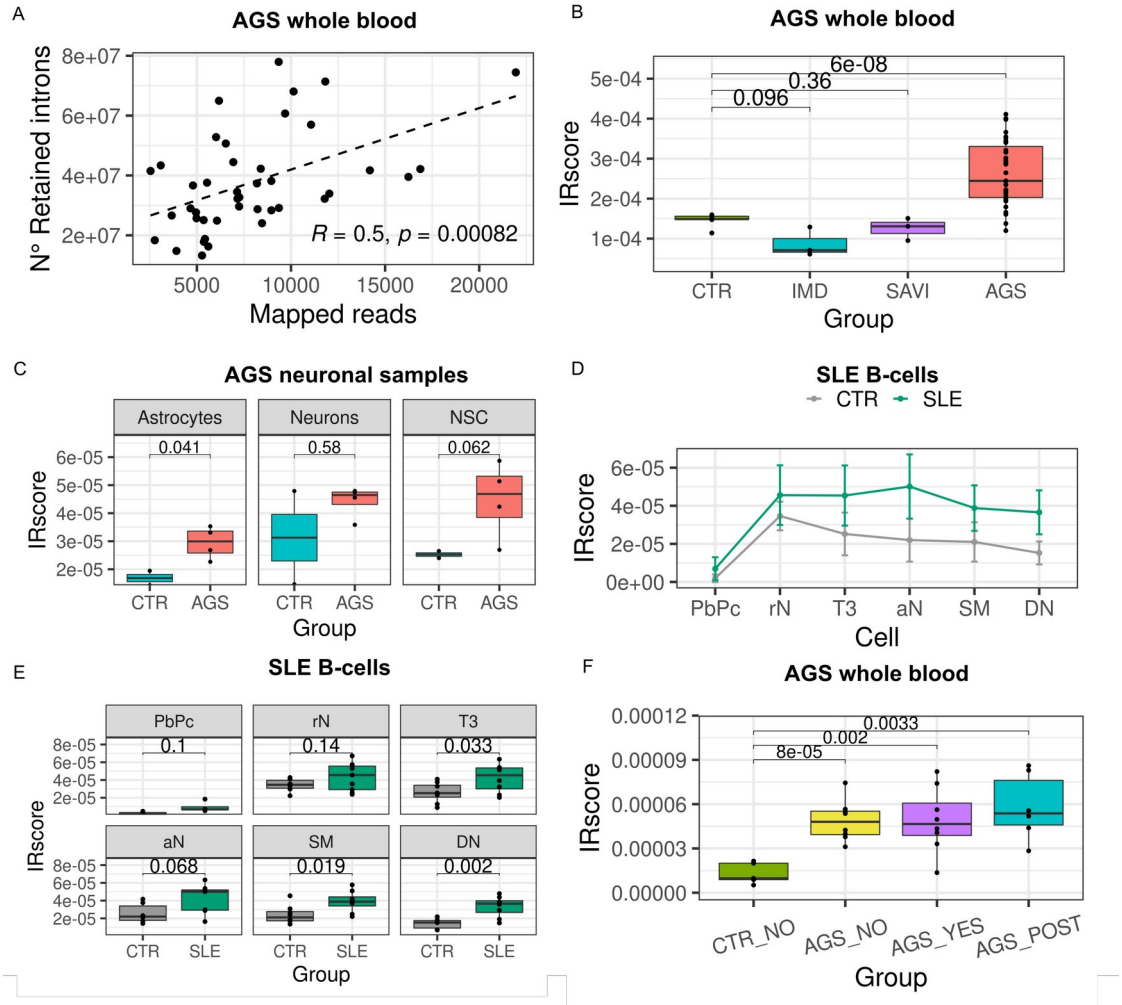


Figure 23: AGS and SLE samples are characterized by increased IR that is not rescued by RT inhibitors treatment. (A) Correlation between mapped reads and retained introns in the AGS whole blood dataset. (B) IRscore distribution in CTR, AGS, SAVI and IMD samples. (C) IRscore distribution in AGS neuronal samples. (D) Median IRscore during B-cells development in CTR and SLE samples. The errorbars represent the standard deviation. (E) IRscore distribution in controls and SLE B-cell subtypes. (F) IRscore distribution in controls, untreated AGS subjects, AGS subjects treated with RT inhibitors and AGS subjects 6 months after treatment suspension.

5.2.2 Inspection of differential IR in AGS and SLE

To determine whether IR dysregulation is a random or a gene-specific event I have performed PCA analysis on the IRratio measures of the introns in the RNA-seq dataset collected from the whole blood of subjects diagnosed with AGS, SAVI and IMD. I have observed that the 2 first principal components are capable of separating AGS subjects from SAVI, IMD and controls suggesting that the level of IRratio of introns in the AGS samples differs from that of introns in SAVI, IMD and controls (figure 24A). Moreover, the PCA analysis is not able to separate AGS untreated subjects from RT inhibitors treated AGS subjects, again suggesting that the level of IR is not directly related to the IFN gene expression as previously hypothesized (figure 24B). At this regard I have also observed that both IMD and SAVI samples, characterized by an active IFN response, cluster together near the control samples further confirming the previous result (figure 24A). Moreover, I have observed that the increased IR is a feature common to all the AGS subjects with different mutations suggesting that the effect is not mutation specific and so it could be associated to a pathway common to all AGS subjects. The clustering of AGS subjects based on the IRratio suggests that IR can targets the same set of introns. To identify introns showing differential retention in AGS samples with respect to controls I have compared the distribution of IRratio for all comparable introns (see materials and methods) using a t-test. Differential retention analysis highlight 5,734 introns with increased retention and 137 with decreased retention in AGS samples with respect to controls confirming the increased IR rate in AGS whole blood samples (figure 24C). The figures 25A and 25B represent the IGV screenshot of 2 differentially retained introns between AGS samples, in red, and control samples, in green. I have observed an increased intronic coverage in AGS samples further confirming the increased retention of this 2 introns in AGS samples with respect to controls. The results obtained in AGS whole blood samples can be further extended to AGS neuronal samples. In figure 26 A, B and C I have observed an higher number of retained introns in neurons, astrocytes and NSC AGS cell with respect to their healthy counterparts. These results confirm the presence of differential IR also in AGS neuronal samples and therefore that increased IR is a characteristic of different AGS cell types. One of the main mechanism in which a tightly regulated IR is involved is the differentiation and activation of the immune system (Ullrich and Guigó 2020; Song et al. 2022; Green et al. 2020). The hypothesis that IR can be a key player

in immune system cell differentiation strongly request the use of samples of isolated specific cell-types to precisely understand the behavior of IR. At this regard, I have used the 5 different B-cell subtypes from SLE and healthy subjects. PCA analysis on IRratio identifies a separation between SLE and control DN cells suggesting that introns affected by highly variable retention level may be able to seprate SLE DN samples to their helthy counterpart and that IR dysregulation may affect a specific subset of introns (figure 26D). To further identify the differentially retained introns I have compared the SLE DN cells vs the control DN cells, as this is the cell type in which IRscore difference is the highest among all tested cell types (figure 23E). Differentially retained introns analysis highlight 433 introns showing increased retention in SLE DN cells and 35 introns showing decreased retention with respect to controls confirming an increased IR in SLE DN cells (figure 26E). To identify common genes affected by increased retention in all the datasets analyzed I have performed an overlap analysis by counting the number of common genes containing at least one intron characterized by a significant increase in IR level. The analysis identify a little degree of overlap among the different datasets with only 2 genes characterized by significantly increased IR in all the datasets analyzed. Anyway, the level of overlap is higher when looking at group pairs. For example the AGS whole blood dataset shares 189, 166, 142 and 88 common genes affected by increased IR with AGS neurons, AGS astrocytes, SLE DN cells and AGS NSC respectively. This may suggest that genes affected by increased IR may be part of cell type-specific pathways and that IR can participate in the regulation of different cellular processes in different cellular types (figure 27).

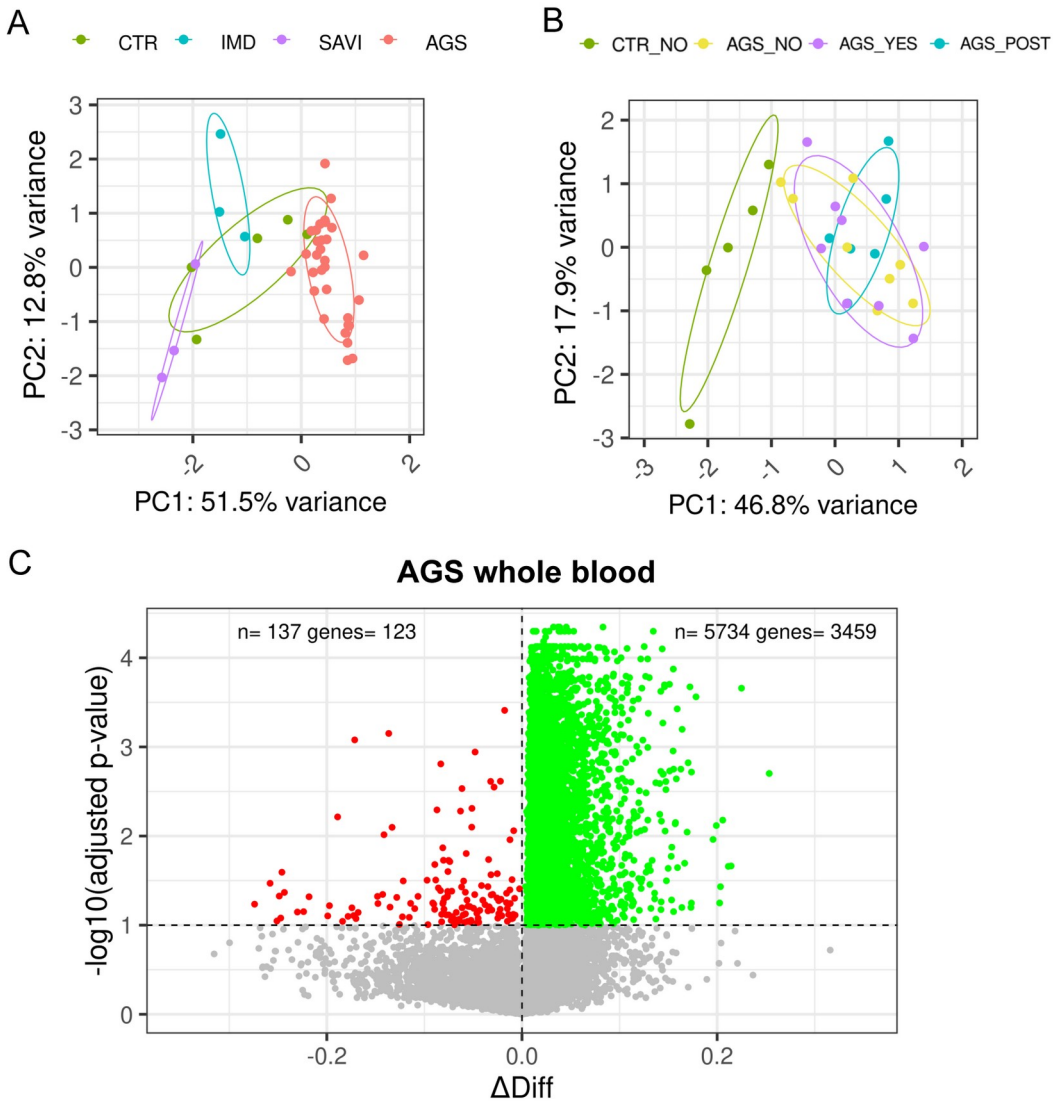
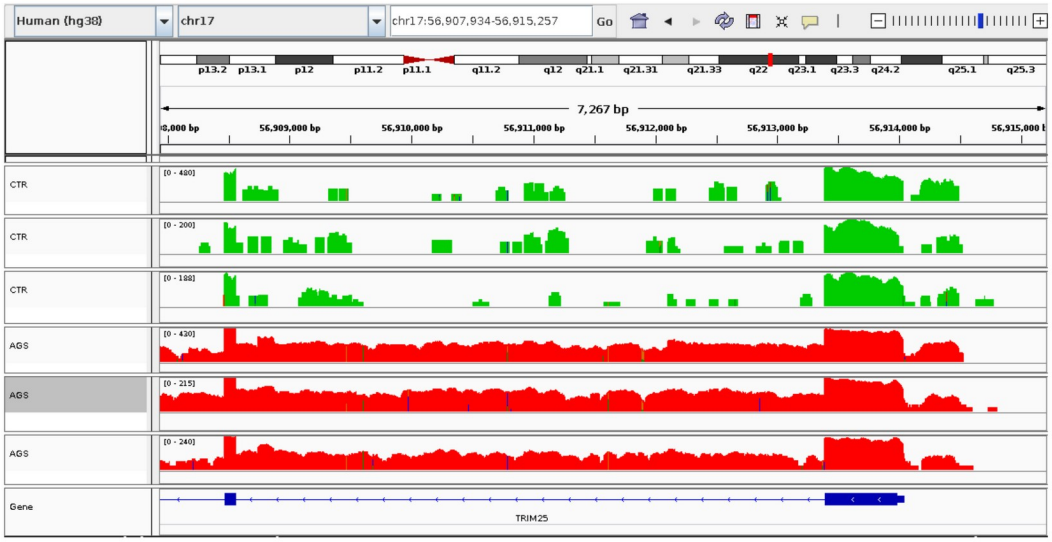


Figure 24: AGS whole blood samples are characterized by differential IR. (A) PCA analysis on IRratio in control, AGS, SAVI and IMD samples. (B) PCA analysis on IRratio of controls, untreated, treated and after treatment AGS samples. (C) Volcano plot of differentially retained introns in AGS samples vs healthy controls. "n" represents the number of differentially retained introns while "genes" the number of genes containing at least one differentially retained intron. ΔDiff represents the difference between the mean IRratio of the intron in AGS samples and the mean IRratio in CTR samples. Introns with significantly increased retention are represented by the green points while introns with significantly reduced retention are represented by the red points.

A



B

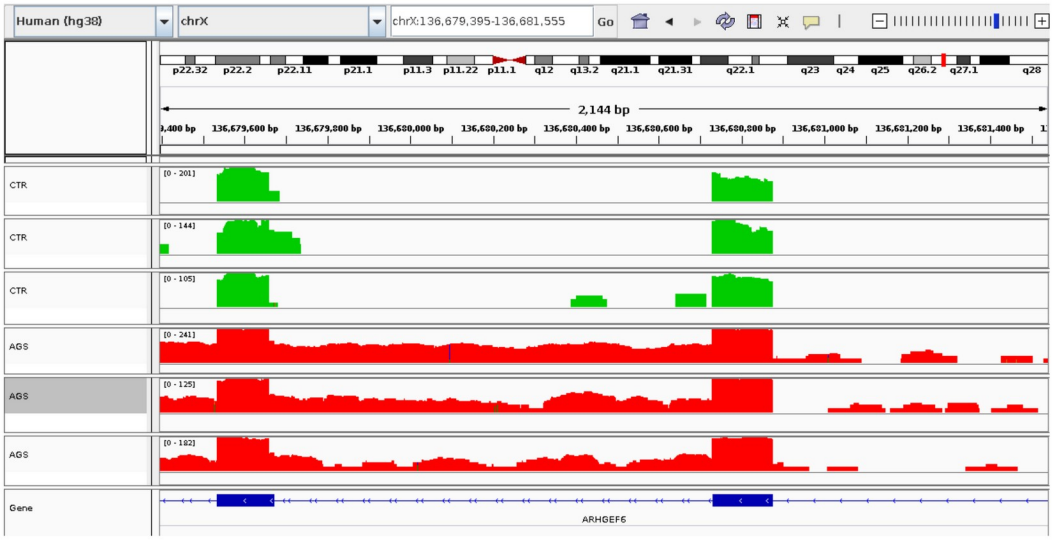


Figure 25: Example of differential IR between controls and AGS samples. The figure reports the IGV screenshot of introns undergoing increased IR in AGS samples (red) with respect to healthy controls (green). Coverage is reported in logaritmic scale. The gene model is reported on the "Gene" track. The blue boxes represents exons while the blue lines represents introns (A) TRIM25 gene, (B) ARHGEF6 gene.

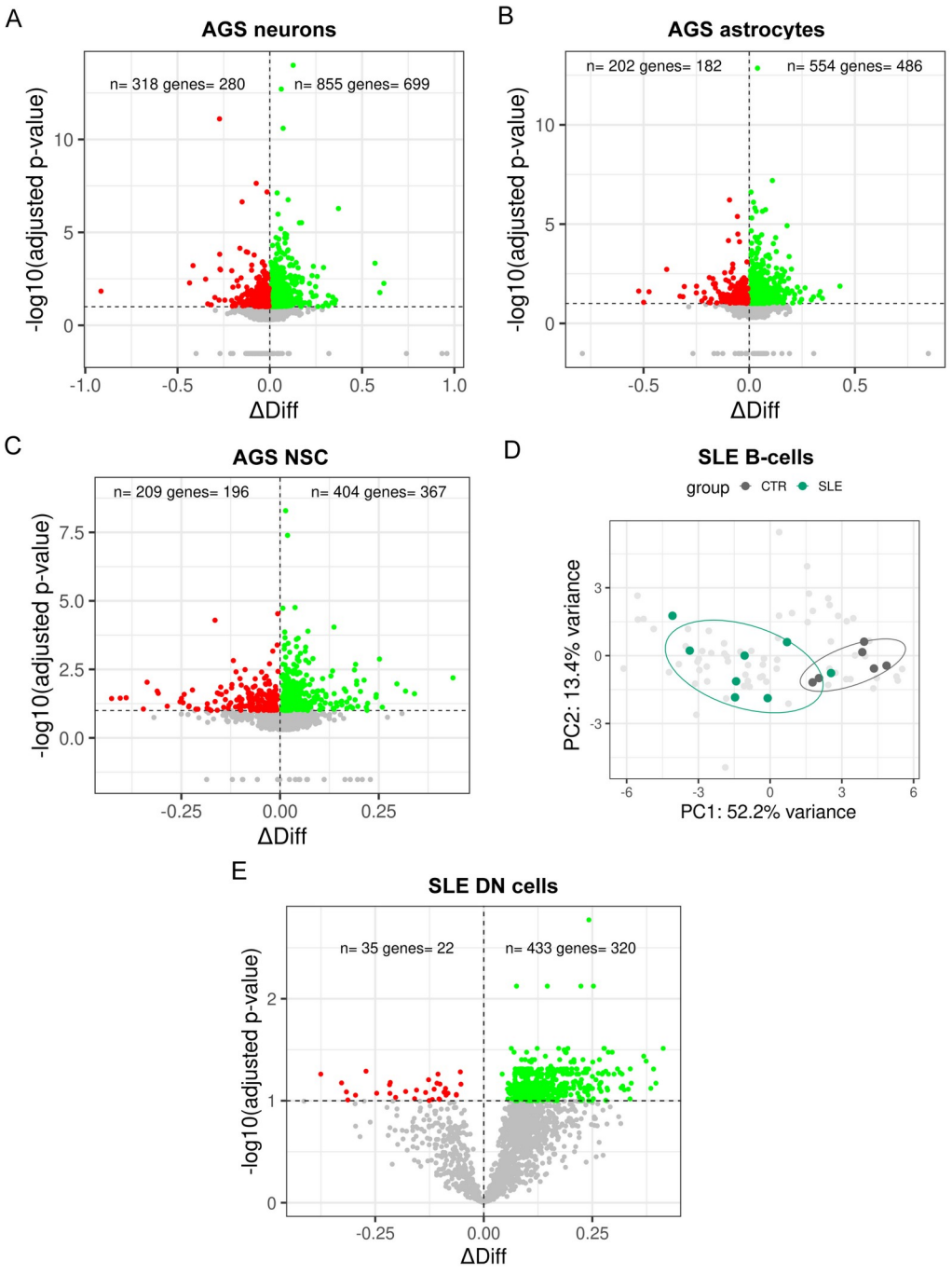


Figure 26: AGS neuronal samples and SLE DN cells are characterized by differential intron retention. (A) Volcano plot of differentially retained introns in AGS neurons. (B) Volcano plot of differentially retained introns in AGS astrocytes. (C) Volcano plot of differentially retained introns in AGS NSC. "n" represents the number of differentially retained introns while "genes" the number of genes containing at least one differentially retained intron. ΔDiff represents the difference between the mean IRratio of the intron in AGS samples and the mean IRratio in CTR samples. (D) PCA analysis on intron IRratio in SLE B-cell subtypes dataset. (E) Volcano plot of differentially retained introns in SLE DN cells.

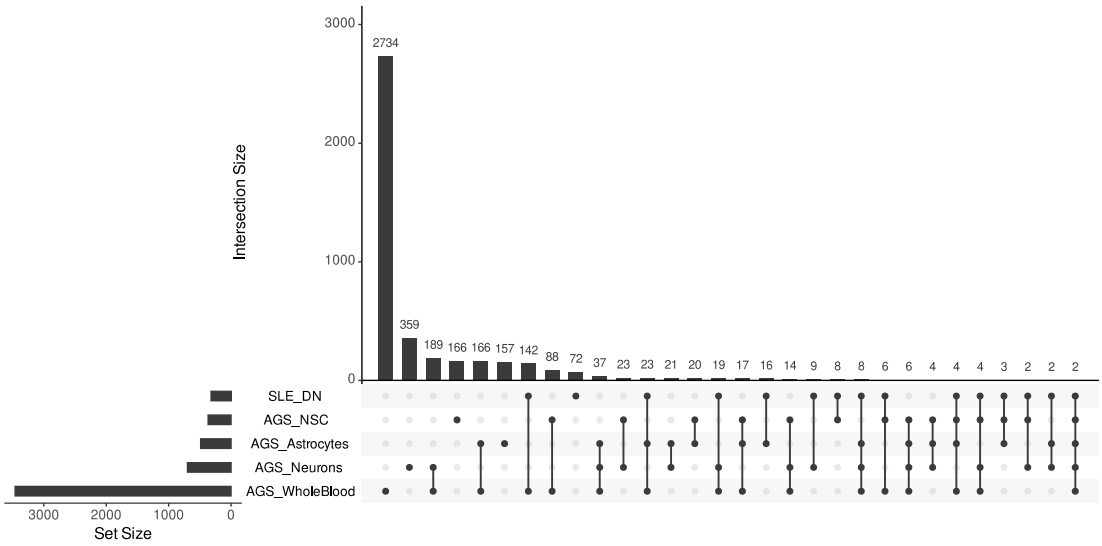


Figure 27: Overlap between genes affected by significantly increased IR in all the dataset analyzed. Upset plot summarizing the overlap between genes affected by increased IR in the dataset analyzed. The intersections matrix represents the different sets generated by the intersection of the elements in the groups analyzed, with the rows corresponding to the sets, and the columns to the intersections between these sets. The size of the sets and of the intersections are shown as barplots.

5.2.3 Genes containing differentially retained introns belong to specific pathways

To further understand the role of IR in the context of autoimmune and neurodevelopmental disease I have performed gene pathways enrichment analysis on genes with at least one intron showing increased retention in AGS whole blood samples with respect to controls. Pathways enrichment analysis reveals that the pathways of “metabolism of RNA”, “immune system” and “mRNA splicing” are the most significantly affected suggesting a possible involvement of IR in those processes (figure 28A). The presence of increased IR in genes involved in mRNA processing and splicing can suggest the presence of feedback mechanism in which increased IR leads to the dowregulation of specific splicing factors that, in turn, cause the increase in the level of IR as also proposed by Laureau and colleagues (Lareau et al. 2007). Interestingly, Ullrich and Guigò, during their studies on normal B-cell development have reported that many genes, involved in splicing regulation, are among the most affected by IR (Ullrich and Guigó 2020). Another pathway significantly enriched for

genes undergoing increased IR in AGS subjects-derived whole blood samples is the "cell cycle" pathway suggesting that IR can be involved in cell cycle regulation probably by affecting the gene expression or protein stability of cell cycle regulators (Middleton et al. 2017). Importantly the "innate immune system" and "adaptive immune system" pathways are significantly enriched for genes undergoing increased intron retention suggesting that IR can be also involved in the regulation of the immune response in both healthy and disease conditions (figure 28A). Regarding this, Ullrich and Guigò, have identified that retained introns during normal B-cell development are enriched in genes involved in NF-kB signaling, an important mediator of the immune response (Ullrich and Guigó 2020). Moreover, they have identified that the level of IR in those genes positively correlates with their expression suggesting that IR can participate in modulating the gene expression levels of genes involved in inflammation (Ullrich and Guigó 2020). Studying macrophages development and activation, Green and colleagues, have extended these results demonstrating that IR events affect the expression of key genes and proteins involved in macrophage development and activation (Green et al. 2020). In particular, they have highlighted a set of genes, among which CXCL2 and NFKBIZ, that are characterized by increased IR. In the presence of pro-inflammatory stimuli, IR transcripts are rapidly spliced, exported to cell cytosol and translated, thus enabling a fine control on the expression of key inflammatory regulators to coordinate the immune response (Green et al. 2020). To understand the role of IR in other cell types I have performed pathway enrichment analysis on genes with increased IR in astrocyte and neuron samples from AGS subjects. I have observed that astrocytes derived from AGS subjects are characterized by a significantly increased IR in genes belonging to the "metabolism of RNA" and "RNA splicing" pathways suggesting a possible parallelism with the results observed in whole blood samples of AGS subjects (figure 28B). Astrocytes are key players in the regulation of neuroinflammation in the brain (Colombo and Farina 2016; Dong and Benveniste 2001). Indeed they have been reported to be one of the major source of IFN production in AGS brains (Cuadrado et al. 2013). Among the pathways significantly enriched for genes characterized by increased IR in astrocytes derived from AGS subjects I have also found "Immune system" and "Cytochine signaling in immune system" further extending the possible role of IR in immune response regulation also in neuronal samples. To identify the pathways affected by IR in AGS

neurons I have performed pathways enrichment analysis on genes characterized by increased IR in neurons derived from AGS subjects with respect to controls. Here the most significantly enriched pathways are related to nervous system development, axon guidance and extracellular matrix organization suggesting that, in this cell type, the role of IR can be related to the development of neurons and their capability to form synapse or respond to external stimuli (figure 28C) (Mazille et al. 2022; Mauger, Lemoine, and Scheiffele 2016). This may suggests that dysregulation of IR in neuronal cells may impair neuronal differentiation and plasticity thus relating IR with neurodevelopmental diseases. Finally, regarding the B-cell SLE dataset, I have observed similar results to those of AGS whole blood samples with genes undergoing increased IR being part of “RNA splicing” and “RNA metabolism” pathways confirming the results previously obtained (figure 28D). These results suggest that IR have the capability of targeting different classes of functionally related genes possibly in a cell-type specific manner and to have an impact on different biological process depending on the targeted genes. I have observed that differential IR affects specific genes in pathways mainly related to RNA metabolism, immune system response, cell cycle control and neuronal development. In this scenario, due to the important pathways targeted by IR, and their involvement in autoimmune and neurodevelopmental diseases, it is critical to deepen the knowledge about this important cellular process and further explore the causes and consequences of IR in AGS and other related pathologies.

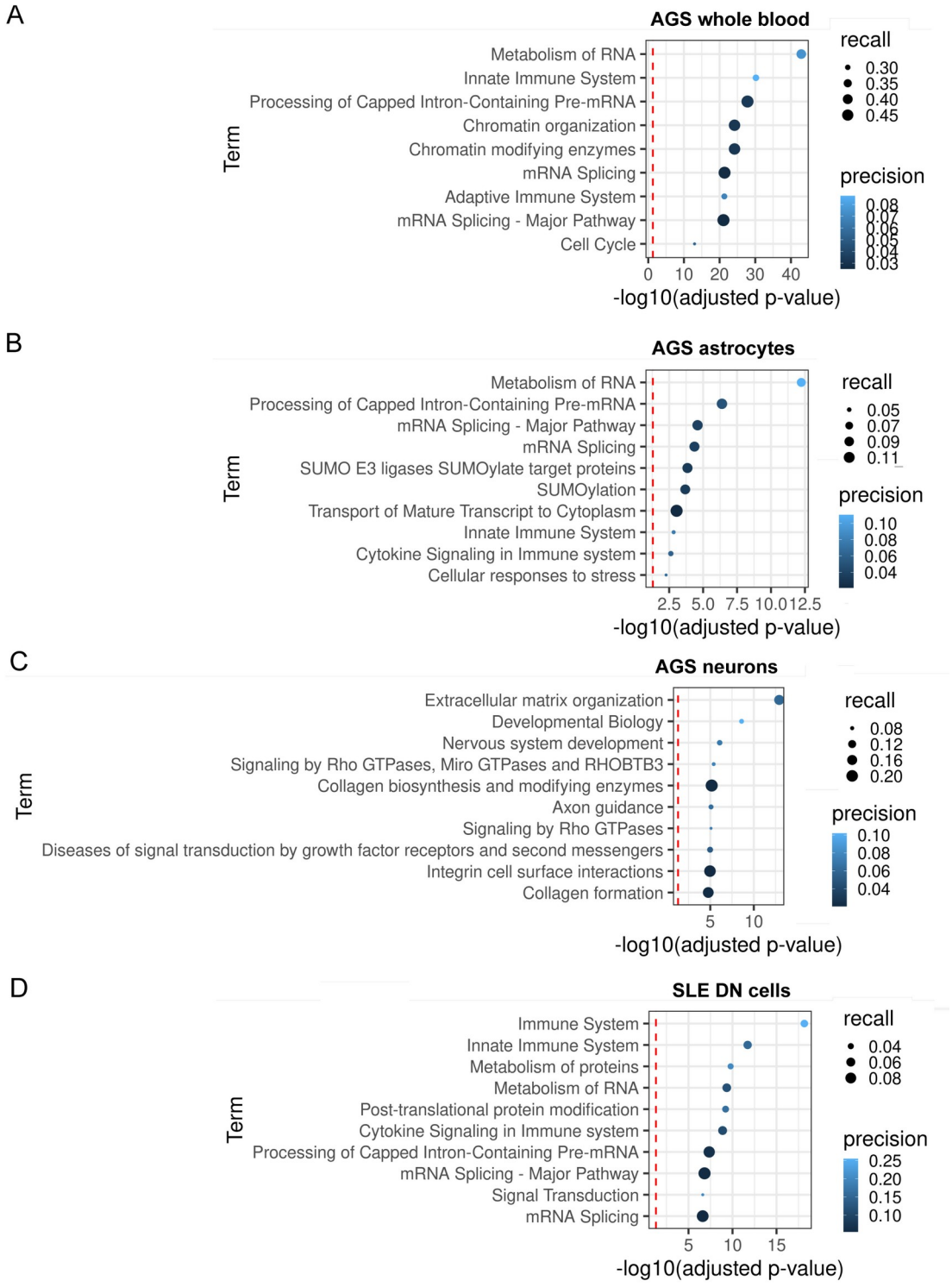


Figure 28: Genes undergoing increased IR are enriched in RNA metabolism, IFN signaling and Neuronal development pathways. The panel reports the most significantly enriched pathways of genes with at least one intron showing increased retention in (A) AGS whole blood samples, (B) AGS astrocytes, (C) AGS neurons and (D) SLE DN B-cells.

5.2.4 IR may be associated with changes in gene expression of hosting genes in AGS whole blood samples

One of the main functions associated with IR is the control of transcripts and/or protein abudance (Grabski et al. 2021). Generally, an increase IR has been associated with a reduced transcript and/or protein abundance even if associations between increased IR and transcript upregulation are also reported in literature (Schmitz et al. 2017; Ullrich and Guigó 2020; Yao et al. 2020; Middleton et al. 2017; Song et al. 2022). To identify a possible association between IR and transcript abundance I have performed a randomization analysis and computed the Z-score. This can help to understand if introns characterized by increased retention in AGS whole blood samples are significantly enriched in differentially expressed genes, with respect to a random state, possibly suggesting an association between IR and transcript abundance. The possible association of IR with transcript abundance was tested by counting the real number of introns characterized by increased retention in AGS subjects (n = 5,734) located in deregulated genes (n = 2,531) and by comparing it with the average of 1000 randomizations of the same number of introns taken from an unbiased background of introns located in expressed genes (n = 212,561). Figure 29A reports the distribution of the 1000 randomizations. The black dot represents the real value. I have observed that 900 retained introns are located in deregulated genes. By comparing this value with the mean of 1000 randomizations (mean = 786, sd = 25) I have obtained a Z-score of 4.5 suggesting that introns with increased retention in AGS subjects are significantly associated with deregulated genes (figure 29A). Figure 29B reports the relation between the ΔIR (AGS - CTR) on the x-axis and the Log2FoldChange of gene expression (AGS / CTR) on the y-axis. With different colors are reported the significant results (adjusted p-value < 0.05 for gene expression and adjusted p-value < 0.1 for IR analysis). The majority of introns (grey point), seems to be not associated with the deregulation of the hosting gene while 451 and 449 introns appear to co-occur in upregulated (yellow point) and downregulated (red point) genes, respectively. Interestingly in literature is reported that IR can be associated both with upregulated and downregulated genes usually with a gene class-specific effect. For example, in their studies upon megakaryocyte and erythrocyte development, Edwards and colleagues have identified a group of introns showing increased retention associated with a reduced expression level of the harbouring gene suggesting that IR serves to

broadly regulate mRNA levels (Edwards et al. 2016). Those genes were mainly enriched in ontologies related to splicing and cell cycle suggesting that IR is a mode of mRNA regulation that functions in a gene-specific manner (Edwards et al. 2016). Interestingly other studies have reported an opposite result in which increased IR is associated with increased transcript abundance. For example, during normal B-cell development, Ullrich and Guigò have identified that some IR events are associated with increased transcript abundance of genes enriched in the NF-kB pathway, suggesting that the effect of IR on transcript abundance may depend, at least in part, on gene function (Ullrich and Guigó 2020). One mechanism that can explain these apparently unexpected results is the discovery of detained introns (DIs). DIs mark transcripts that remain in the nucleus and are not subject to NMD degradation. Interestingly, studies in different cellular contexts, have revealed that DIs are enriched for gene ontologies related to functions specific of the cell type in which they occur (Boutz, Bhutkar, and Sharp 2015). As also proposed by Fu and Bouts, IR transcripts can be divided into 2 main classes: the first one corresponds to transcripts degraded by NMD upon cytoplasmic exports, and the second one corresponds to transcripts retained in the nucleus and insensitive to NMD (Fu 2017; Boutz, Bhutkar, and Sharp 2015). Due to the possible different outcome of IR on transcripts abundance, I have defined 3 classes of genes in which IR occur: non-deregulated, upregulated or downregulated. To determine the function of genes in which IR is associated with a dysregulation of gene expression I have performed genes pathway enrichment analysis using as background the set of upregulated or downregulated genes respectively (adjsuted p-value < 0.05). Interestingly I have observed that the association of IR with gene expression largely depends on gene functions as previously observed by other groups (Ullrich and Guigó 2020; Pimentel et al. 2016). For example genes with increased IR belonging to the group of significant downregulated genes are involved in mRNA processing and splicing while significantly upregulated genes are involved in IFN signaling and immune response in accordance with the literature (figure 29C-D) (Liu et al. 2017; Bhatt et al. 2012; Ullrich and Guigó 2020). Taken together these results suggest that IR is associated with gene expression in different ways with gene-specific outcomes. Moreover, the relation between IR and gene expression seems to rely on gene function. Taken together these results may suggest that IR could be associated with changes in transcripts abundance in whole blood samples from AGS

subjects and that the relationship between IR and gene expression may be gene type specific.

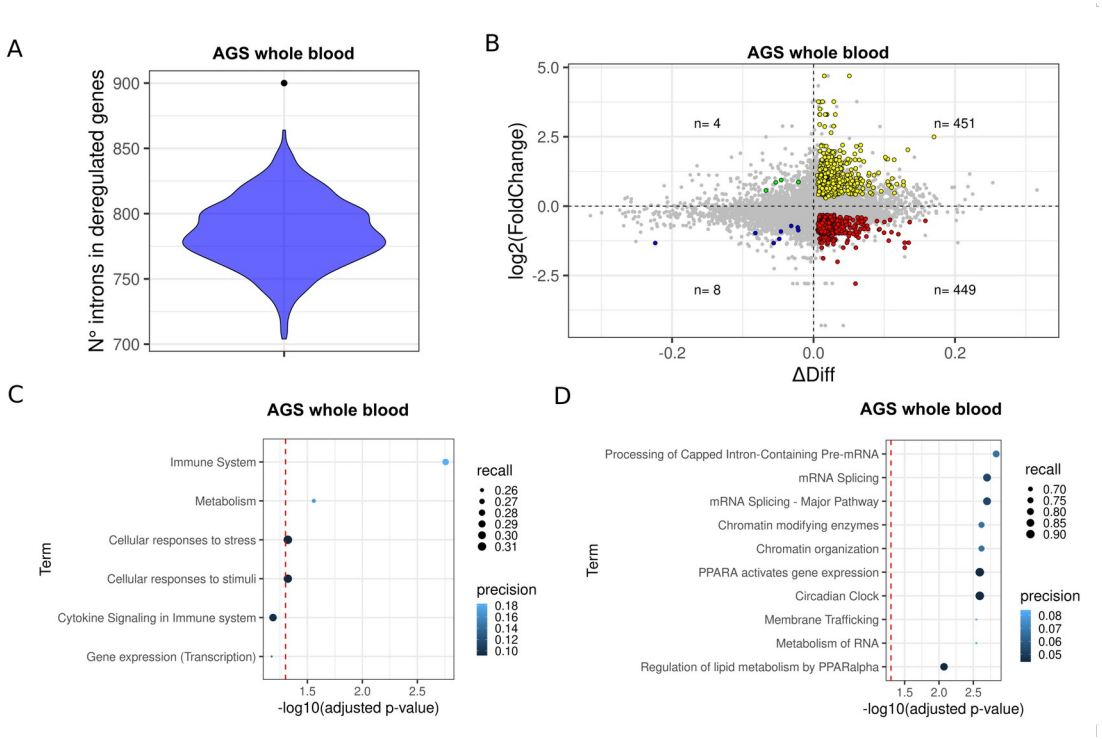


Figure 29: IR may be associated with changes in transcript abundance. (A) The blu violin plot represents the distribution of the number of randomly chosen introns located in deregulated genes from 1000 randomizations. The black dot represent the real number of differentially retained introns located in deregulated genes. Z-score = 4.5. (B) Comparison between log2FoldChange of gene expression changes in AGS samples on the y-axis and the Difference in IRratio between AGS and control samples on x-axis. In yellow are reported upregulated genes characterized by increased IR, in red downregulated genes characterized by increased IR, in blu downregulated genes characterized by decreased IR while in green upregulated genes characterized by decreased IR. "n" represents the number of introns (C) Enrichment analysis on upregulated genes with at least one intron showing increased retention. Gene background was represented by upregulated genes (adjusted p.value < 0.05 and log2(FoldChange) > 0) (D) Enrichment analysis on downregulated genes with at least one intron showing increased retention. Gene background was represented by downregulated genes (adjusted p.value < 0.05 and log2 (FoldChange) < 0.

5.2.5 Retained introns are characterized by specific features

Intron retention is a tightly regulated mechanism that is influenced by a variety of different factors (Monteuuis et al. 2019; Sakabe and de Souza 2007). Interestingly different works have tried to shed light on the complicated IR code that controls IR and have highlighted some intrinsic characteristics that are associated with increased

IR. In addition to this also cis-acting players, like specific RNA-binding proteins and splicing factors, are involved in the control of IR (Monteuuis et al. 2019; Middleton et al. 2017). To decipher the possible underlying "IR code" and understand the intrinsic factors influencing IR in AGS whole blood samples I have decided to annotate each retained intron with different characteristics to shed light on features that differentiate them with respect to randomly chosen introns. In this analysis I have taken into account the mean intronic GC content, the intron length and the splice site types and the splice sites strength and compared it to the mean of 100 randomizations of randomly chosen introns. Retained introns have been divided into 3 groups according to their association with gene expression (downregulated, non-deregulated and upregulated genes) as introns associated with deregulated genes may have different features and participates in different pathways (Jacob and Smith 2017; Boutz, Bhutkar, and Sharp 2015; Mazille et al. 2022; Ullrich and Guigó 2020). Interestingly all the 3 groups of retained introns are characterized by canonical splice sites suggesting that retained introns may be removed by the canonical splicing machinery and probably additional factors are involved in the modulation of IR (table 5). To understand if the increased IR may be due to weaker splice sites associated with lower recognition by the canonical splicing machinery I have predicted both the 5' and the 3' site splice strength using the MaxEntScan software (Yeo and Burge 2004). The splice site strength is a quantitative measure of the propensity of the splice site of being recognized and used by the splicing machinery (N. Choi et al. 2021). Here higher values correspond to a stronger splice site. Figure 30A-B reports the random distribution of 5' splice site strength and 3' splice site strength respectively for the 3 classes of retained introns. Retained introns associated with downregulated genes are characterized by slightly stronger 5' and 3' splice sites with respect to the random state (figure 30A-B, z-score for 5' = 2.20 and z-score for 3' = 2.39). Retained introns associated with non-deregulated genes do not show significant changes in neither 5' nor 3' splice site strength with respect to the random state (figure 30A-B, z-score for 5' = -0.2 and z-score for 3' = -0.4). Finally, retained introns associated with upregulated genes are characterized by a sligthly weaker splice 5' splice site strength (figure 30A-B, z-score for 5' = -2.5 and z-score for 3' = -1.3). Another important factor influencing IR is the intron length (Braunschweig et al. 2014; Dvinge and Bradley 2015; Jacob and Smith 2017). I have observed that introns showing increased retention in AGS whole

blood samples are characterized by a significantly shorter length compared to randomly chosen introns, irrespectively to their associations with gene expression (figure 30C). The last investigated intron characteristic is the GC content. A differential GC content has been associated with IR in the vast majority of the studies conducted. Figure 30D reports the distribution of GC content of randomly chosen introns compared with retained introns associated with downregulated, non-deregulated and upregulated genes represented by the black dots. I have observed that retained introns associated with downregulated genes are characterized by a significant increase in the GC content while retained introns associated with upregulated genes are characterized by a significantly lower GC content when compared to randomly chosen introns possibly in accordance with the results of Mazille at al. (Mazille et al. 2022). Mazille et al., studying the subcellular localization of IR transcripts, have identified 2 main groups of retained introns: those associated with cytosolic transcripts, probably undergoing NMD or other forms of degradation, and those nuclear localized, protected from degradation, that are used as a nuclear reservoir of transcripts, spliced and exported upon external stimuli, to increase protein expression. In this scenario they have reported that retained introns of nuclear localized IR transcripts are characterized by a decreased GC content while retained introns associated with cytosolic localized IR transcripts show an inverse trend (Mazille et al. 2022). Taken together these results suggest that introns showing increased retention are characterized by a shorter length, a differential GC content and a different 3' and 5' splice site strength with respect to randomly chosen introns. Interestingly these features seem to differentiate not only retained introns from non-retained introns but also the different classes of retained introns associated with upregulated or downregulated genes.

Group	Type	Site	Counts	Total	Percentage
Downregulated	GT	5’	447	449	99.5%
Downregulated	AG	3’	449	449	100%
Non-deregulated	GT	5’	4712	4772	98.7%
Non-deregulated	AG	3’	4756	4772	99.6%
Upregulated	GT	5’	439	451	97.3%
Upregulated	AG	3’	450	451	99.7%

Table 5: Number of different splice sites sequences in differentially retained introns in AGS whole blood samples associated with downregulated, non-deregulated or upregulated genes. The table reports the number and type of splice sites sequences associated with introns showing increased retention in AGS whole blood samples. Introns have been divided into 3 classes according to their association with the expression of the harbouring gene.

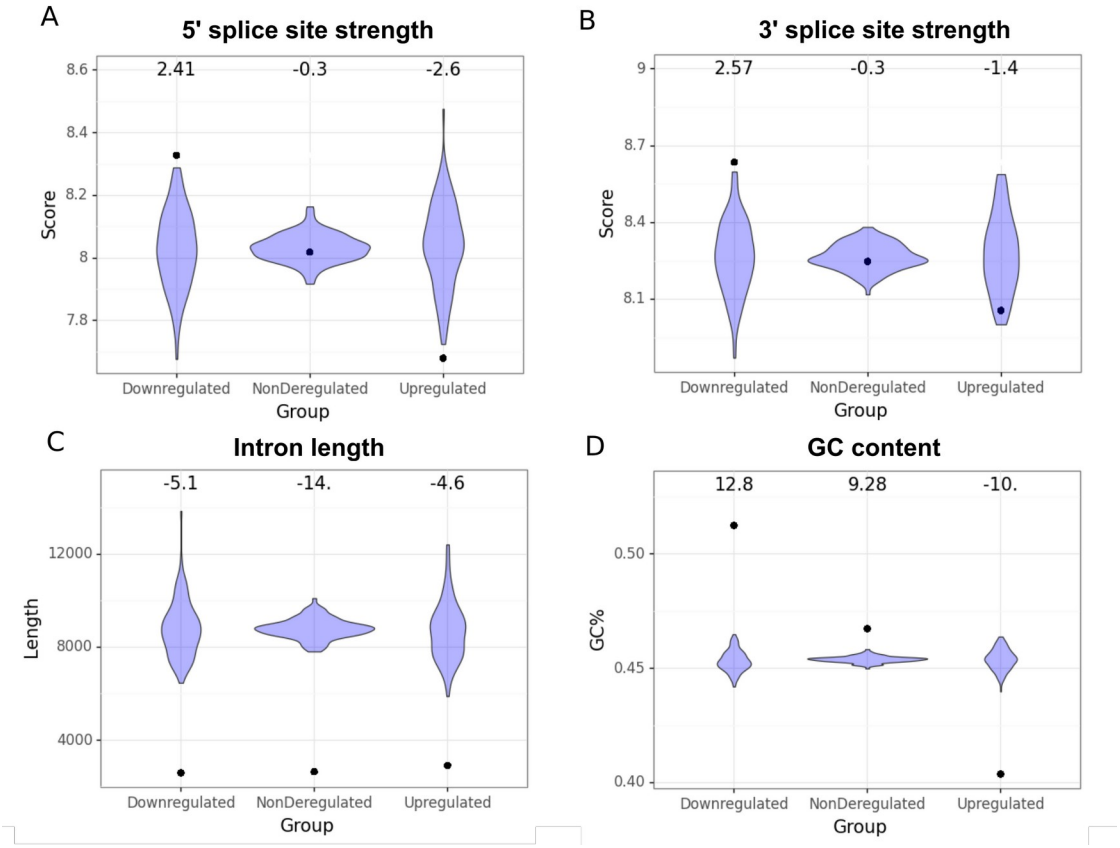


Figure 30: Retained introns are characterized by specific features. (A) Distribution of 5’ splice sites strength for retained introns (black dot, mean) and for randomly chosen introns (violin plot, distributions of 100 randomizations). Z-score are reported on the top of the plot. (B) Distribution of 3’ splice sites strength for retained introns (black dot, mean) and for randomly chosen introns (violin plot, distributions of 100 randomizations). Z-score are reported on the top of the plot. (C) Distribution of intron length for retained introns (black dot, mean) and for randomly chosen introns (violin plot, distributions of 100 randomizations). Z-score are reported on the top of the plot. (D) Distribution of GC% for retained introns (black dot, mean) and for randomly chosen introns (violin plot, distributions of 100 randomizations). Z-score are reported on the top of the plot.

5.2.6 IR may be modulated by specific RNA binding protein

In addition to the previously described *cis*-acting features, IR is probably regulated in *trans* by a plethora of different RNA-binding proteins and splicing factors proteins (Monteuuis et al. 2019; Middleton et al. 2017). In this scenario, it is possible that a variety of different mechanisms lead to different IR programs, and that the type of retained introns may influence the fates of intron-retaining transcripts (Jacob and Smith 2017). The different features, in particular the different GC content, that differentiate the groups of retained and non-retained introns in AGS whole blood samples suggest the possibility that both the level of IR and the fate of IR transcripts may be regulated by specific RNA binding proteins (Monteuuis et al. 2019). To identify possible enrichment of RNA binding motifs in retained introns I have counted the number of specific RNA binding motifs in the 3 different groups of retained introns in AGS whole blood samples and compared it to the mean of 1000 randomizations by computing the Z-score. The RNA motifs have been retrieved from the CISBP-RNA database which contains the binding sequences of more the 300 RNA binding protein (Ray et al. 2013). Due to the fact that retained introns are significantly shorter than randomly chosen introns the number of binding sites per each intron has been normalized on the intron length as longer introns have higher chances to contain more binding sites (Monteuuis et al. 2019). Interestingly I have observed that retained introns associated with downregulated, non-deregulated and upregulated genes are significantly enriched for the biding of 9, 4 and 22 RNA motifs with respect to the mean of 1000 randomizations (according to a Z-score > 3) (figure 31A-B-C). This result may suggest that different classes of retained introns are regulated by specific RNA binding proteins. By looking at common RNA binding motifs in the 3 different groups of differentially retained introns I have observed that the downregulated and non-deregulated groups have the highest number of common RNA binding motifs with respect to the upregulated group that is characterized only by private RNA binding motifs (figure 31D). In particular there are 2 RNA binding protein significantly enriched in differentially retained introns associated with downregulated and non-deregulated genes (FUS and SRP54). Interestingly, by performing RNA motif enrichment in introns showing increased retention in AGS astrocytes and neurons I have observed similar results (Figure 31E-F). In fact I have observed significant enrichment for FUS, KHSRP, SRP54, and RBMX which are in common with retained

introns in AGS whole blood samples. This confirms that IR regulation by RNA binding protein may be conserved between different cell types. One of the most interesting results is the putative binding by the FUS protein. Indeed, FUS has been reported to bind retained introns during B-cell development and to be itself affected by IR and also to be involved in ALS insurgence (Ullrich and Guigó 2020; Humphrey et al. 2020). Interestingly, FUS is an autoregulated protein that participates in splicing and RNA metabolism and its impairment could influence the IR levels. Another interesting group of RNA binding proteins that are significantly enriched in retained introns of AGS subjects-derived whole blood samples and astrocytes is the SR proteins. In particular the SRSF1 and SRSF9 proteins. Other works have reported enrichment of SRSF1 binding sites in retained introns with respect to non retained further supporting the validity of the observation (Jacob and Smith 2017; Ullrich and Guigó 2020). Those proteins can mediate the mRNA export from nucleus to cytoplasm thus increasing the NMD rate of intron-retaining transcripts. Moreover, the SRSF1 protein can directly induce NMD by recruiting UPF1, one of the most important NMD regulators (Aznarez et al. 2018, 1). Another protein whose binding site sequence is significantly enriched in differentially retained introns, belonging to the downregulated group in AGS whole blood dataset and AGS astrocytes, is the KHSRP protein. This protein is involved in the degradation of mRNAs that contain AU-rich elements in their 3'-UTR, possibly by recruiting degradation machinery to AU-rich elements containing mRNAs. Moreover this protein is involved in monocytic differentiation by controlling AS and in particular IR (Xu et al. 2022). For what regards the binding sites enriched only in retained introns of upregulated genes in the AGS whole blood dataset I have reported that PTBP1 protein has been associated with mRNAs protection from recognition by the nonsense-mediated mRNA decay pathway (Ge et al. 2016). In this way the bind of PTBP1 to the retained intron can shield the transcript from NMD leading to its cellular accumulation. The IGF2BP1 and other proteins belonging to the IGFBP protein family, specifically enriched in retained introns of upregulated genes in AGS whole blood samples, are involved in protecting m^6A-containing mRNAs from degradation and subsequently promote their translation (X. Huang et al. 2018; H. Huang et al. 2018). Figure 31G reports the mean binding frequency (number of binding sites/intron length), in the 3 different classes of retained introns in the AGS whole blood dataset, for the FUS RNA binding site motif. I have observed a higher

mean binding frequency in retained introns associated with downregulated and non-deregulated genes with respect to retained introns associated with upregulated genes. Analyzing the mean binding frequency of PTPB1 and IGF2BP1, whose binding motifs are significantly enriched only in retained introns associated with upregulated genes, I have observed an opposite result, with higher mean binding frequency in retained introns associated with upregulated genes with respect to other groups (figure 31H). Taken together those results identify a potential set of RNA binding proteins capable of influencing the level of IR and/or the fates of intron retaining transcripts. Interestingly, the set of RNA binding motifs significantly enriched in retained introns associated with upregulated genes shows no overlap with the set of RNA binding motif significantly enriched in retained introns associated with downregulated or non-deregulated genes suggesting a possible different post-transcriptional regulation program for the different class of IR transcripts. Finally, it is important to point out here that the specific function of the putative binding proteins can explain, at least in part, the direction of the differential expression for intron retaining transcripts.

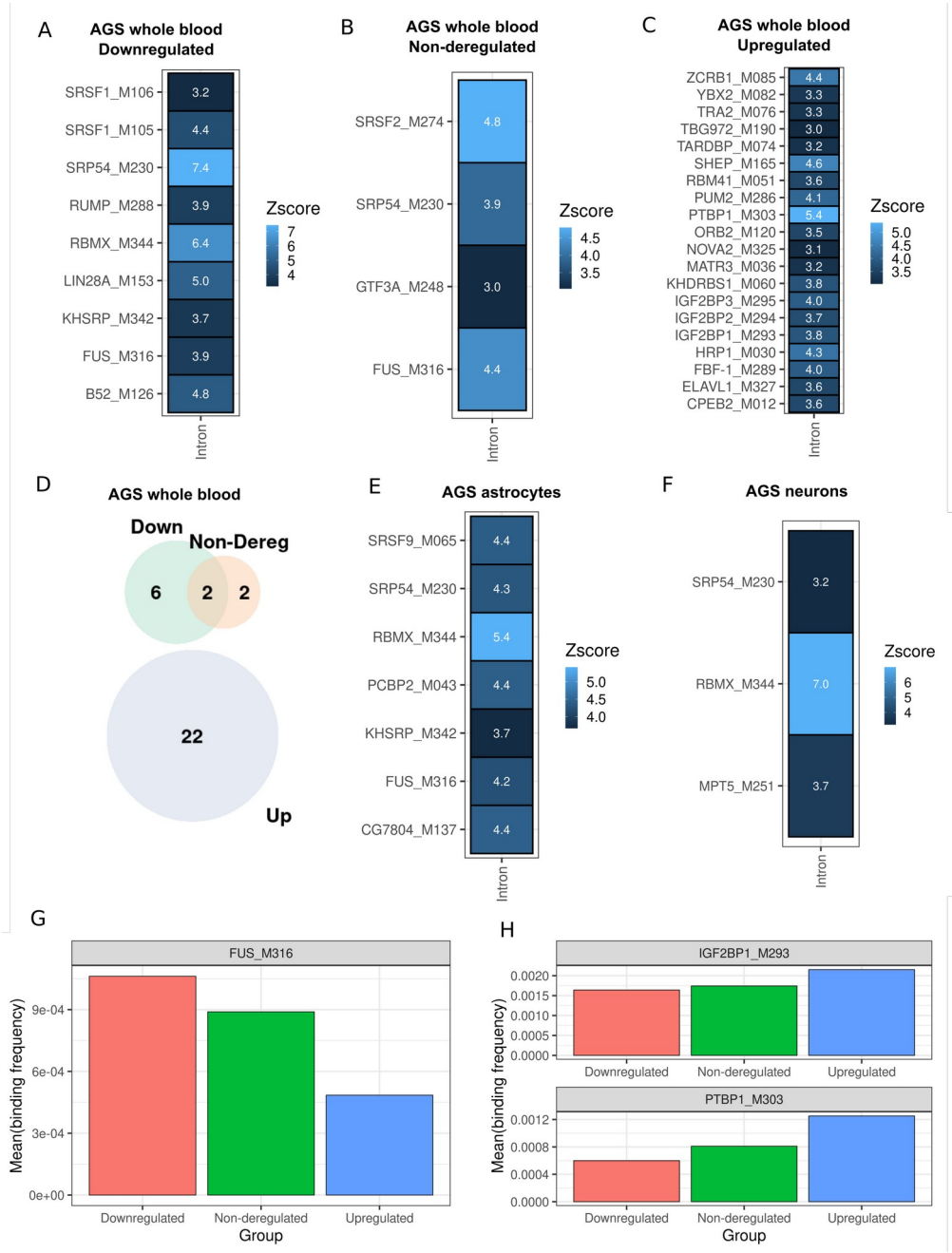


Figure 31: IR may be modulated by specific RNA binding proteins. Heatmap of Z-score enrichment for different RNA binding motifs in intronic regions for retained introns associated with downregulated (A), non-deregulated (B) or upregulated (C) genes in AGS whole blood sample dataset. Only Z-score >= 3 and only top 20 results are reported. (D) Overlap between significant RNA binding motifs (Z-score > 3) in differentailly ratained itnrons associated with downregulated (Down), non-deregulated (Non-Dereg) or upregulated (Up) genes. (E-F) Heatmap of Z-score enrichment for different RNA binding motifs in retained introns regions for differentially retained introns in AGS astrocytes (E) and AGS neurons (F). (G-H) Barplot of binding motifs frequency for FUS protein (G), IGF2BP1 and PTBP1 (H) in differentially retained introns in AGS whole blood samples associated with downregulated, non-deregulated or upregulated genes. The mean binding frequency, reported on the y-axis, is computed, per each intron, as the number of binding sites/intron length.

5.2.7 AGS whole blood samples are characterized by expansion of immune system cells associated with increased IR

Whole blood samples are a complex mixture of different cell types that, on one hand, can complicate results interpretation but, on the other hand, allow for the quantification of the relative abundance of different cell types. Recently, in scientific literature, it has been reported that IR can regulate the physiological differentiation and activation of different immune cell types (Song et al. 2022; Ullrich and Guigó 2020; Wong et al. 2013). Moreover, given the highlighted over-activation of the IFN signaling in AGS and its stimulatory effect on the activation of immune cells it is worth hypothesizing a differential activation of immune cells in AGS whole blood samples (Le Page et al. 2000). The activation of immune system cells is usually accomplished by their maturation, differentiation and activation and this process is often coupled with their expansion in terms of cell number that can be reflected with different reciprocal cell-type abundance (Kan and Hodgkin 2014). To quantify the relative abundance of different cell-types in the samples analyzed I have performed cell-types enrichment analysis on the bulk RNA-seq data using as reference the gene expression signature from 64 different immune and stromal cell types taking advantage of the xCell deconvolution software (Aran, Hu, and Butte 2017). Results revealed a significant enrichment of B-cells in AGS samples with respect to healthy controls (p.value = 0.02 , figure 32A). Interestingly, samples not characterized by a significant increase in IR, like SAVI and IMD samples, do not show a significant enrichment of B-cells with respect to healthy controls suggesting a potential relation between IR and B-cells expansion. This result reveals that B-cells are over-represented in whole blood samples from AGS subjects and suggests an over-activation of this cell type. To further explore a possible relationship between increased IR and B-cells expansion I have performed correlation analysis between the IRscore and the B-cells enrichment score. The result reveal a positive significant correlation between IR and B-cells content (p.value = $4.6 * 10^{-7}$) suggesting that an increased IR may be associated with an increase in the relative B-cells content (Figure 32B). In 2012 Gall and colleagues provided important clues on the contribution to autoimmunity of different cell types in TREX1 mutated cells. In particular, they have demonstrated that crossing a *Trex1*-null mouse with a B-cell deficient mouse (*Ighm* $^{-/-}$) rescues the cellular toxicity but does not reduce IFN levels suggesting that probably IFN alone is not able to sustain disease

progression (Gall et al. 2012). Taken together these results highlight a B-cells expansion in AGS samples associated with an increase in IR events. Given these results it is possible to hypothesize that the increased IR observed in AGS and SLE samples may be associated with the proliferation/differentiation decision of immune system cells and in particular of B-cells (Werner and Jumaa 2015).

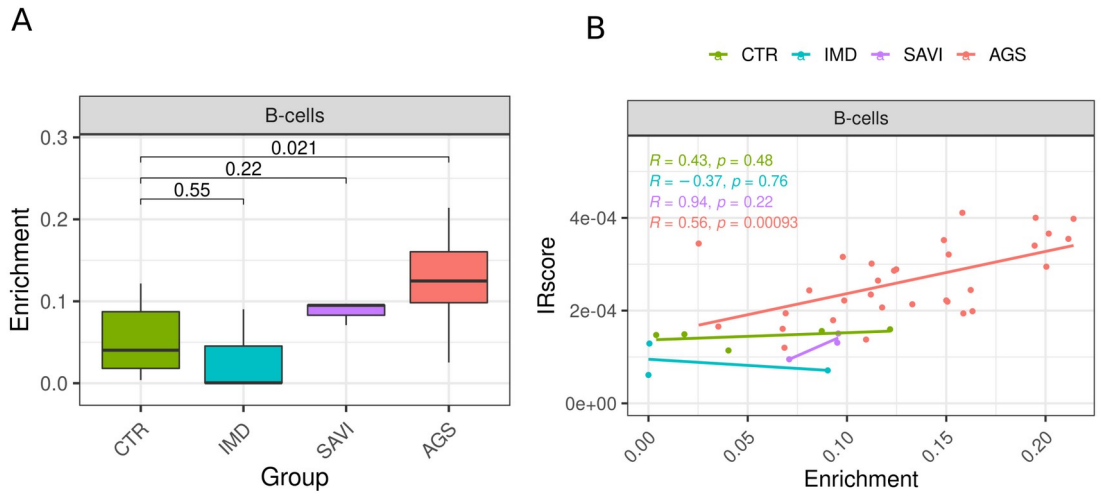


Figure 32: Increased IR is associated with increased B-cell content in AGS whole blood samples. (A) Boxplot representing the distribution of B-cells enrichment in AGS, IMD, SAVI and Control samples. (B) Correlation between IRscore and B-cells enrichment score. In the top left of the plot are reported the correlation coefficients (R) and the p-value (p) for the different groups in the dataset.

5.2.8 IR may be associated with the proliferation / differentiation decision

Given the evidence that IR is associated with genes involved in mRNA splicing and cell cycle, and the fact that an increased IR is associated with an expansion of B-cells in AGS whole blood samples, one interesting hypothesis is the possibility that IR can serve as a "proliferation/differentiation" marker (Petasny et al. 2021; Bowman et al. 2006; Ullrich and Guigó 2020). In particular the increased IR observed in AGS and SLE samples may be associated with the increased proliferation/differentiation and activation of immune system cells, in particular B-cells, that can be further involved in disease progression. To decipher the possible association between IR and the proliferation/differentiation decision I have correlated the global level of IR, represented by the IRscore, with the proliferation score which represents the propensity of cell to proliferate rather than differentiate. This score is computed from

the expression of different genes involved in the control of cells proliferation. It represents thus a proposed strategy for obtaining a quantitative cell proliferation measurement from gene expression data. This can help in identifying differences in cell proliferation rates between conditions and identify other variables correlated with proliferation representing a quantitative measurement of the state of a cell. Figure 33A reported the correlation between IRscore and proliferation score in the whole blood dataset composed by AGS, SAVI, IMD and control subjects. I have observed a significant negative correlation (R = - 0.47, p-value = 0.007) between the IRscore and the proliferation score in AGS whole blood samples while no significant correlations have been reported for the other groups analyzed. Even if the result can be influenced by the different samples sizes, this result suggest that, in AGS whole blood samples, the observed increased IR may be related to the proliferative state of blood cells. Taken together with the previous result, highlighting an enrichment of B-cell in AGS whole blood samples with respect to healthy controls, this result may be in agreement with the activation and differentiation of B-cell in AGS. To study the relation between IR and the proliferative activity in the context of B-cells development I have performed the same correlation analysis in the SLE B-cell dataset, composed of 5 different populations of FACS-isolated B-cell subtypes. As observed in figure 33B the IRscore negatively correlates with the proliferation score in both SLE B-cell and healthy controls (SLE R = -0.57 and p-value = $4.7 * 10^{-5}$, CTR R = -0.35 and p-value = 0.03). This result confirs that an increased IR is associated with a reduced proliferation both in healthy and in disease samples. Extending the correlation analysis to the AGS neuronal dataset, composed by NSC, astrocytes and neurons from both AGS subjects and healthy controls I have reported a a significant negative correlation between the IRscore and the proliferation score in the AGS group (R = -0.59 and p-value = 0.04) while in the control group the correlation is not significant (R = -0.28 and p-value = 0.59) (figure 33C). Interestingly, in the AGS neuronal dataset, I have observed that neurons are the cell type with the lowest proliferation score as post-mitotic neurons have low or absent proliferation potential (Aranda-Anzaldo 2012). In this regards, in neurons, the principal role of IR has been described in controlling the neuronal plasticity after neuronal stimulation suggesting that possibly IR can have a cell-type specific role (Mauger, Lemoine, and Scheiffele 2016; Mazille et al. 2022). In this scenario, I have hypothesized that an increased IR is associated with reduced cell

proliferation and, possibly a tendency of cells to differentiate. In fact usually cell proliferation and differentiation follow an opposite direction with cells capable of proliferating and dividing till the acquisition of a fully differentiated state characterized by proliferation arrest (Ruijtenberg and van den Heuvel 2016). In this scenario, IR appears to be possibly associated with a reduction in the proliferative potential of cells. Taken together these results suggest that a global increase in IR level may be associated with a general reduction in the proliferation, possibly in favour of differentiation and activation, in cells with immune-related functions.

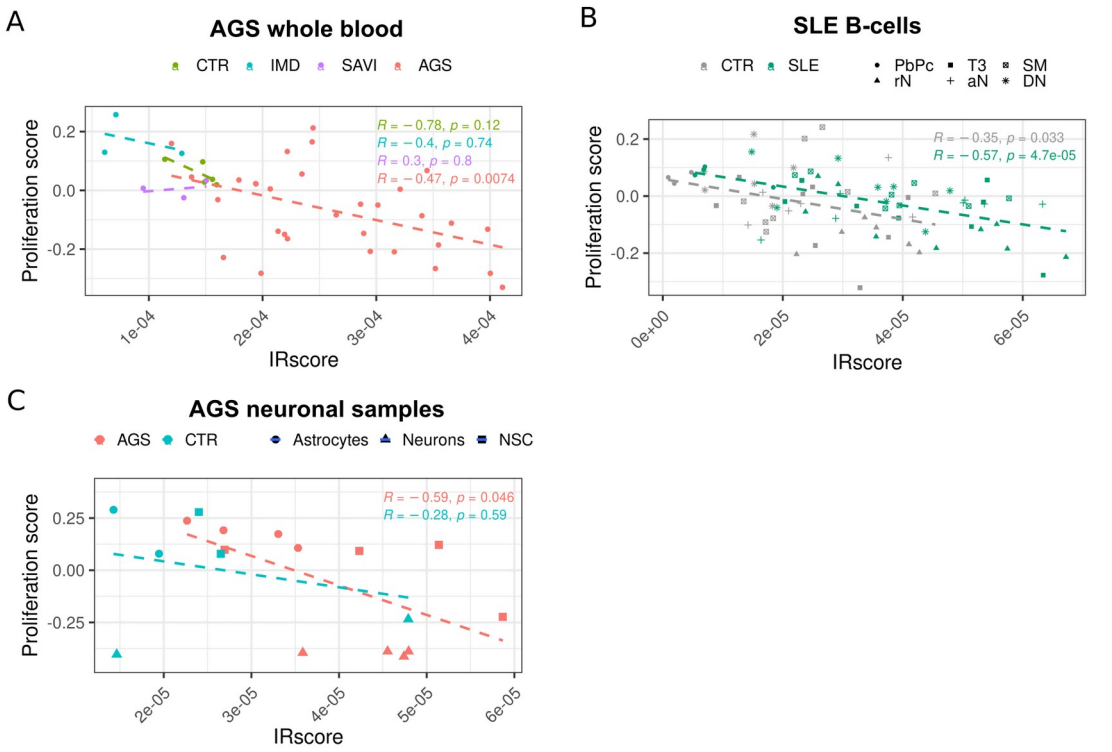


Figure 33: Increased IR is associated with reduced cell proliferation. (A) Correlation between IRscore and proliferation score in AGS whole blood samples. (B) Correlation between IRscore and proliferation score in B-cells SLE samples. (C) Correlation between IRscore and proliferation score in AGS neuronal samples.

6 Conclusions

AGS is an autoimmune and neurodegenerative disorder that is thought to arise from aberrant and continuous IFN signaling. Nowadays the mechanisms underlying the onset and progression of AGS are only partially understood and more efforts are needed to comprehend the molecular basis of this disease for which currently there is no cure. In this work, I have observed that whole blood samples from AGS subjects are characterized by increased IR in key genes involved in IFN response, mRNA metabolism and cell cycle. Interestingly literature research have revealed that increased IR in genes involved in mRNA splicing and inflammation has been observed during normal B-cell development. This observation, given the aberrant immune response observed in AGS, has inspired the hypothesis that the increased IR observed in AGS subjects may be associated with an aberrant B-cells response.

To investigate this hypothesis I have performed bulk RNA-seq data cell-type enrichment analysis that has revealed an expansion of B-cells in AGS samples compared to healthy controls that can be further associated with an overactivation of this cell type and the increased IR observed in AGS samples. To confirm the involvement of IR in B-cells development and activation in another relevant and similar disease to AGS, I have analyzed RNA-seq data from SLE B-cell samples isolated at 5 different differentiation stages confirming the result previously obtained. Finally, to study IR in other AGS relevant cell types, I have retrieved RNA-seq data from AGS subjects-derived iPSC *in vitro* differentiated into neuronal stem cells, astrocytes and neurons. I have observed that neurons, astrocytes and neural stem cells derived from AGS subjects are characterized by differential IR with respect to their healthy counterparts further extending the possible roles of IR in neuronal tissue. Further analysis of pathways enriched for genes affected by increased IR has revealed that IR dysregulation is mainly involved in the control of immune response and cell development. Associations analysis between IR levels and the expression of genes hosting the retained introns have revealed that IR is associated with both upregulated, downregulated and non deregulated genes possibly in a gene group function-specific manner.

Genes showing increased IR in AGS whole blood samples can be divided into 3 categories based on their expression level: downregulated, non deregulated and upregulated with respect to healthy controls. Interestingly, upregulated genes associated with increased IR are enriched in pathways related to IFN signaling and immune system response while downregulated genes associated with increased IR are part of the RNA metabolism and splicing pathways. This result suggests that aberrant IR in AGS may participate in the dysregulation of gene expression in key genes involved in mRNA processing and IFN response. In AGS neuronal samples, I have reported an association between IR and genes involved in neuronal development and synapse formation that can further account for some neurodegenerative aspects observed in AGS. This suggests that IR may have a cell type-specific effect. The observation of IR in different related datasets and its association with specific pathways suggest that IR may be a tightly regulated process. In fact retained introns are characterized by common hallmarks including a reduced length and a differential GC content with respect to randomly chosen introns. I have observed that different classes of retained introns are enriched for specific RNA binding protein motifs suggesting that transcripts containing retained introns are post-transcriptionally regulated. Given the pathways enriched by genes characterized by differential IR and given its associations with the immune system expansion and activation in AGS whole blood samples this AS event has been proposed as a marker of cell proliferation/differentiation decision. I have observed a significant negative correlation between the IRscore and the proliferation score in all the dataset analyzed further confirming that an increased IR is negatively correlated with the proliferative activity and possibly points to a more differentiated and activated cell state. I can conclude that increased IR may be associated with the aberrant over-activation of the immune response observed in AGS whole blood samples, astrocytes and SLE B-cells. Moreover IR can play important roles in the aberrant neuronal development observed in AGS subjects. In fact, in AGS neurons, aberrant IR has been associated with genes involved in neuronal development and axon guidance. Even if additionally analyses are needed to confirm end extend these results, to my knowledge, this is the first work suggesting the presence of a differential IR programme in whole blood and neuronal samples derived from AGS subjects.

Whether IR participates to AGS onset or is the result of AGS remains yet to be confirmed. I propose that IR, in AGS and SLE, is associated with the aberrant activation of the immune system that is involved in the onset and progression of these diseases. This evidence can lead to the development of treatments for alleviating disease sympthoms and reducing the autoimmune response in AGS and SLE subjects.

7 Bibliography

Aditi, Susanna M. Downing, Patrick A. Schreiner, Young Don Kwak, Yang Li, Timothy I. Shaw, Helen R. Russell, and Peter J. McKinnon. 2021. "Genome Instability Independent of Type I Interferon Signaling Drives Neuropathology Caused by Impaired Ribonucleotide Excision Repair." *Neuron*, October. https://doi.org/10.1016/j.neuron.2021.09.040.

Adusumalli, Swarnaseetha, Zhen-Kai Ngian, Wei-Qi Lin, Touati Benoukraf, and Chin-Tong Ong. 2019. "Increased Intron Retention Is a Post-Transcriptional Signature Associated with Progressive Aging and Alzheimer's Disease." *Aging Cell* 18 (3): e12928. https://doi.org/10.1111/acel.12928.

Aicardi, J., and F. Goutières. 1984. "A Progressive Familial Encephalopathy in Infancy with Calcifications of the Basal Ganglia and Chronic Cerebrospinal Fluid Lymphocytosis." *Annals of Neurology* 15 (1): 49–54. https://doi.org/10.1002/ana.410150109.

Alberts, Bruce, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. 2002. "How Genomes Evolve." *Molecular Biology of the Cell. 4th Edition*. https://www.ncbi.nlm.nih.gov/books/NBK26836/.

Ali, Arsala, Kyudong Han, and Ping Liang. 2021. "Role of Transposable Elements in Gene Regulation in the Human Genome." *Life* 11 (2): 118. https://doi.org/10.3390/life11020118.

Anders, Simon, Paul Theodor Pyl, and Wolfgang Huber. 2015. "HTSeq--a Python Framework to Work with High-Throughput Sequencing Data." *Bioinformatics (Oxford, England)* 31 (2): 166–69. https://doi.org/10.1093/bioinformatics/btu638.

Aran, Dvir, Zicheng Hu, and Atul J. Butte. 2017. "XCell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape." *Genome Biology* 18 (1): 220. https://doi.org/10.1186/s13059-017-1349-1.

Aranda-Anzaldo, Armando. 2012. "The Post-Mitotic State in Neurons Correlates with a Stable Nuclear Higher-Order Structure." *Communicative & Integrative Biology* 5 (2): 134–39. https://doi.org/10.4161/cib.18761.

Asefa, Benyam, Kimberly D. Klarmann, Neal G. Copeland, Debra J. Gilbert, Nancy A. Jenkins, and Jonathan R. Keller. 2004. "The Interferon-Inducible P200 Family of Proteins: A Perspective on Their Roles in Cell Cycle Regulation and Differentiation." *Blood Cells, Molecules & Diseases* 32 (1): 155–67. https://doi.org/10.1016/j.bcmd.2003.10.002.

Ayarpadikannan, Selvam, and Heui-Soo Kim. 2014. "The Impact of Transposable Elements in Genome Evolution and Genetic Instability and Their Implications in Various Diseases." *Genomics & Informatics* 12 (3): 98–104. https://doi.org/10.5808/GI.2014.12.3.98.

Aznarez, Isabel, Tomoki T. Nomakuchi, Jaclyn Tetenbaum-Novatt, Mohammad Alinoor Rahman, Oliver Fregoso, Holly Rees, and Adrian R. Krainer. 2018. "Mechanism of Nonsense-Mediated MRNA Decay Stimulation by Splicing Factor SRSF1." *Cell Reports* 23 (7): 2186–98. https://doi.org/10.1016/j.celrep.2018.04.039.

Baechler, Emily C., Franak M. Batliwalla, George Karypis, Patrick M. Gaffney, Ward A. Ortmann, Karl J. Espe, Katherine B. Shark, et al. 2003. "Interferon-Inducible Gene Expression Signature in Peripheral Blood Cells of Patients with

Severe Lupus." *Proceedings of the National Academy of Sciences* 100 (5): 2610–15. https://doi.org/10.1073/pnas.0337679100.
Bartsch, Kareen, Katharina Knittler, Christopher Borowski, Sönke Rudnik, Markus Damme, Konrad Aden, Martina E. Spehlmann, et al. 2017. "Absence of RNase H2 Triggers Generation of Immunogenic Micronuclei Removed by Autophagy." *Human Molecular Genetics* 26 (20): 3960–72. https://doi.org/10.1093/hmg/ddx283.
Bass, Brenda L. 2002. "RNA Editing by Adenosine Deaminases That Act on RNA." *Annual Review of Biochemistry* 71: 817–46. https://doi.org/10.1146/annurev.biochem.71.110601.135501.
Batalis, Stephanie, LeAnn C. Rogers, Wayne O. Hemphill, Christopher H. Mauney, David A. Ornelles, and Thomas Hollis. 2021. "SAMHD1 Phosphorylation at T592 Regulates Cellular Localization and S-Phase Progression." *Frontiers in Molecular Biosciences* 0. https://doi.org/10.3389/fmolb.2021.724870.
Beck, Christine R., José Luis Garcia-Perez, Richard M. Badge, and John V. Moran. 2011. "LINE-1 Elements in Structural Variation and Disease." *Annual Review of Genomics and Human Genetics* 12: 187–215. https://doi.org/10.1146/annurev-genom-082509-141802.
Belancio, Victoria P., Astrid M. Roy-Engel, Radhika R. Pochampally, and Prescott Deininger. 2010. "Somatic Expression of LINE-1 Elements in Human Tissues." *Nucleic Acids Research* 38 (12): 3909–22. https://doi.org/10.1093/nar/gkq132.
Belgnaoui, S Mehdi, Roger G Gosden, O John Semmes, and Abdelali Haoudi. 2006. "Human LINE-1 Retrotransposon Induces DNA Damage and Apoptosis in Cancer Cells." *Cancer Cell International* 6 (May): 13. https://doi.org/10.1186/1475-2867-6-13.
Beloglazova, Natalia, Robert Flick, Anatoli Tchigvintsev, Greg Brown, Ana Popovic, Boguslaw Nocek, and Alexander F. Yakunin. 2013. "Nuclease Activity of the Human SAMHD1 Protein Implicated in the Aicardi-Goutières Syndrome and HIV-1 Restriction." *The Journal of Biological Chemistry* 288 (12): 8101–10. https://doi.org/10.1074/jbc.M112.431148.
Benitez-Guijarro, Maria, Cesar Lopez-Ruiz, Žygimantė Tarnauskaitė, Olga Murina, Mahwish Mian Mohammad, Thomas C. Williams, Adeline Fluteau, et al. 2018. "RNase H2, Mutated in Aicardi-Goutières Syndrome, Promotes LINE-1 Retrotransposition." *The EMBO Journal* 37 (15). https://doi.org/10.15252/embj.201798506.
Bhatt, Dev M., Amy Pandya-Jones, Ann-Jay Tong, Iros Barozzi, Michelle M. Lissner, Gioacchino Natoli, Douglas L. Black, and Stephen T. Smale. 2012. "Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions." *Cell* 150 (2): 279–90. https://doi.org/10.1016/j.cell.2012.05.043.
Blaudin de Thé, François-Xavier, Hocine Rekaik, Eugenie Peze-Heidsieck, Olivia Massiani-Beaudoin, Rajiv L Joshi, Julia Fuchs, and Alain Prochiantz. 2018. "Engrailed Homeoprotein Blocks Degeneration in Adult Dopaminergic Neurons through LINE-1 Repression." *The EMBO Journal* 37 (15): e97374. https://doi.org/10.15252/embj.201797374.
Bodak, Maxime, Jian Yu, and Constance Ciaudo. 2014. "Regulation of LINE-1 in Mammals." *Biomolecular Concepts* 5 (5): 409–28. https://doi.org/10.1515/bmc-2014-0018.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics (Oxford, England)* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Boutz, Paul L., Arjun Bhutkar, and Phillip A. Sharp. 2015. “Detained Introns Are a Novel, Widespread Class of Post-Transcriptionally Spliced Introns.” *Genes & Development* 29 (1): 63–80. https://doi.org/10.1101/gad.247361.114.

Bowman, Teresa V., Andrew J. McCooey, Akil A. Merchant, Carlos A. Ramos, Patricia Fonseca, Alan Poindexter, Steven B. Bradfute, et al. 2006. “Differential MRNA Processing in Hematopoietic Stem Cells.” *Stem Cells* 24 (3): 662–70. https://doi.org/10.1634/stemcells.2005-0552.

Brandsma, Inger, and Dik C. van Gent. 2012. “Pathway Choice in DNA Double Strand Break Repair: Observations of a Balancing Act.” *Genome Integrity* 3 (1): 9. https://doi.org/10.1186/2041-9414-3-9.

Braunschweig, Ulrich, Nuno L. Barbosa-Morais, Qun Pan, Emil N. Nachman, Babak Alipanahi, Thomas Gonatopoulos-Pournatzis, Brendan Frey, Manuel Irimia, and Benjamin J. Blencowe. 2014. “Widespread Intron Retention in Mammals Functionally Tunes Transcriptomes.” *Genome Research* 24 (11): 1774–86. https://doi.org/10.1101/gr.177790.114.

Broseus, Lucile, and William Ritchie. 2020. “Challenges in Detecting and Quantifying Intron Retention from next Generation Sequencing Data.” *Computational and Structural Biotechnology Journal* 18: 501–8. https://doi.org/10.1016/j.csbj.2020.02.010.

Buckley, Peter T., Mugdha Khaladkar, Junhyong Kim, and James Eberwine. 2014. “Cytoplasmic Intron Retention, Function, Splicing, and the Sentinel RNA Hypothesis.” *Wiley Interdisciplinary Reviews. RNA* 5 (2): 223–30. https://doi.org/10.1002/wrna.1203.

Busse, D. C., D. Habgood-Coote, S. Clare, C. Brandt, I. Bassano, M. Kaforou, J. Herberg, et al. 2020. “Interferon-Induced Protein 44 and Interferon-Induced Protein 44-Like Restrict Replication of Respiratory Syncytial Virus.” *Journal of Virology* 94 (18): e00297-20. https://doi.org/10.1128/JVI.00297-20.

Caceres, Javier F., and Tom Misteli. 2007. “Division of Labor: Minor Splicing in the Cytoplasm.” *Cell* 131 (4): 645–47. https://doi.org/10.1016/j.cell.2007.11.001.

Chang, Howard H. Y., Nicholas R. Pannunzio, Noritaka Adachi, and Michael R. Lieber. 2017. “Non-Homologous DNA End Joining and Alternative Pathways to Double-Strand Break Repair.” *Nature Reviews. Molecular Cell Biology* 18 (8): 495–506. https://doi.org/10.1038/nrm.2017.48.

Chatterjee, Nimrat, and Graham C. Walker. 2017. “Mechanisms of DNA Damage, Repair and Mutagenesis.” *Environmental and Molecular Mutagenesis* 58 (5): 235–63. https://doi.org/10.1002/em.22087.

Choi, Jongsu, Sung-Yeon Hwang, and Kwangseog Ahn. 2018. “Interplay between RNASEH2 and MOV10 Controls LINE-1 Retrotransposition.” *Nucleic Acids Research* 46 (4): 1912–26. https://doi.org/10.1093/nar/gkx1312.

Choi, Namjeong, Yongchao Liu, Jagyeong Oh, Jiyeon Ha, Claudia Ghigna, Xuexiu Zheng, and Haihong Shen. 2021. “Relative Strength of 5’ Splice-Site Strength Defines Functions of SRSF2 and SRSF6 in Alternative Splicing of Bcl-x Pre-MRNA.” *BMB Reports* 54 (3): 176–81. https://doi.org/10.5483/BMBRep.2021.54.3.170.

Choubey, Divaker, Xin Duan, Eric Dickerson, Larissa Ponomareva, Ravichandran Panchanathan, Hui Shen, and Ratika Srivastava. 2010. “Interferon-Inducible P200-Family Proteins as Novel Sensors of Cytoplasmic DNA: Role in

Inflammation and Autoimmunity." *Journal of Interferon & Cytokine Research* 30 (6): 371–80. https://doi.org/10.1089/jir.2009.0096.
Chuong, Edward B., Nels C. Elde, and Cédric Feschotte. 2017. "Regulatory Activities of Transposable Elements: From Conflicts to Benefits." *Nature Reviews Genetics* 18 (2): 71–86. https://doi.org/10.1038/nrg.2016.139.
Cock, Peter J. A., Tiago Antao, Jeffrey T. Chang, Brad A. Chapman, Cymon J. Cox, Andrew Dalke, Iddo Friedberg, et al. 2009. "Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics." *Bioinformatics* 25 (11): 1422–23. https://doi.org/10.1093/bioinformatics/btp163.
Colombo, Emanuela, and Cinthia Farina. 2016. "Astrocytes: Key Regulators of Neuroinflammation." *Trends in Immunology* 37 (9): 608–20. https://doi.org/10.1016/j.it.2016.06.006.
Cooper, Geoffrey M. 2000. "The Complexity of Eukaryotic Genomes." *The Cell: A Molecular Approach. 2nd Edition*. https://www.ncbi.nlm.nih.gov/books/NBK9846/.
Cosby, Rachel L., Ni-Chen Chang, and Cédric Feschotte. 2019. "Host–Transposon Interactions: Conflict, Cooperation, and Cooption." *Genes & Development* 33 (17–18): 1098–1116. https://doi.org/10.1101/gad.327312.119.
Cristini, Agnese, Michael Tellier, Flavia Constantinescu, Clelia Accalai, Laura Oana Albulescu, Robin Heiringhoff, Nicolas Bery, Olivier Sordet, Shona Murphy, and Natalia Gromak. 2022. "RNase H2, Mutated in Aicardi-Goutières Syndrome, Resolves Co-Transcriptional R-Loops to Prevent DNA Breaks and Inflammation." *Nature Communications* 13 (1): 2961. https://doi.org/10.1038/s41467-022-30604-0.
Crow, Y J, A Vanderver, S Orcesi, T W Kuijpers, and G I Rice. 2014. "Therapies in Aicardi–Goutières Syndrome." *Clinical and Experimental Immunology* 175 (1): 1–8. https://doi.org/10.1111/cei.12115.
Crow, Yanick J. 1993. "Aicardi-Goutières Syndrome." In *GeneReviews®*, edited by Margaret P. Adam, Holly H. Ardinger, Roberta A. Pagon, Stephanie E. Wallace, Lora JH Bean, Ghayda Mirzaa, and Anne Amemiya. Seattle (WA): University of Washington, Seattle. http://www.ncbi.nlm.nih.gov/books/NBK1475/.
Crow, Yanick J., and Nicolas Manel. 2015. "Aicardi–Goutières Syndrome and the Type I Interferonopathies." *Nature Reviews Immunology* 15 (7): 429–40. https://doi.org/10.1038/nri3850.
Crow, Yanick J., and Jan Rehwinkel. 2009. "Aicardi-Goutières Syndrome and Related Phenotypes: Linking Nucleic Acid Metabolism with Autoimmunity." *Human Molecular Genetics* 18 (R2): R130–36. https://doi.org/10.1093/hmg/ddp293.
Cuadrado, Eloy, Machiel H. Jansen, Jasper Anink, Lidia De Filippis, Angelo L. Vescovi, Colin Watts, Eleonora Aronica, Elly M. Hol, and Taco W. Kuijpers. 2013. "Chronic Exposure of Astrocytes to Interferon-α Reveals Molecular Changes Related to Aicardi-Goutieres Syndrome." *Brain: A Journal of Neurology* 136 (Pt 1): 245–58. https://doi.org/10.1093/brain/aws321.
Cuadrado, Eloy, Iliana Michailidou, Emma J. van Bodegraven, Machiel H. Jansen, Jacqueline A. Sluijs, Dirk Geerts, Pierre-Olivier Couraud, et al. 2015. "Phenotypic Variation in Aicardi-Goutières Syndrome Explained by Cell-Specific IFN-Stimulated Gene Response and Cytokine Release." *Journal of Immunology (Baltimore, Md.: 1950)* 194 (8): 3623–33. https://doi.org/10.4049/jimmunol.1401334.

Czech, Benjamin, Marzia Munafò, Filippo Ciabrelli, Evelyn L. Eastwood, Martin H. Fabry, Emma Kneuss, and Gregory J. Hannon. 2018. “PiRNA-Guided Genome Defense: From Biogenesis to Silencing.” *Annual Review of Genetics* 52 (November): 131–57. https://doi.org/10.1146/annurev-genet-120417-031441.

Dai, YunFan, XiuYun Liu, ZhiPeng Zhao, JianXin He, and QingQin Yin. 2020. “Stimulator of Interferon Genes-Associated Vasculopathy With Onset in Infancy: A Systematic Review of Case Reports.” *Frontiers in Pediatrics* 8 (December): 577918. https://doi.org/10.3389/fped.2020.577918.

De Cecco, Marco, Takahiro Ito, Anna P. Petrashen, Amy E. Elias, Nicholas J. Skvir, Steven W. Criscione, Alberto Caligiana, et al. 2019. “L1 Drives IFN in Senescent Cells and Promotes Age-Associated Inflammation.” *Nature* 566 (7742): 73–78. https://doi.org/10.1038/s41586-018-0784-9.

Demaria, Sandra, and Claire Vanpouille-Box. 2017. “TREX1 Is a Checkpoint for Innate Immune Sensing of DNA Damage That Fosters Cancer Immune Resistance.” *Emerging Topics in Life Sciences* 1 (5): 509–15. https://doi.org/10.1042/ETLS20170063.

Denis, Melvin M., Neal D. Tolley, Michaeline Bunting, Hansjörg Schwertz, Huimiao Jiang, Stephan Lindemann, Christian C. Yost, et al. 2005. “Escaping the Nuclear Confines: Signal-Dependent Pre-MRNA Splicing in Anucleate Platelets.” *Cell* 122 (3): 379–91. https://doi.org/10.1016/j.cell.2005.06.015.

Dewannieux, Marie, Cécile Esnault, and Thierry Heidmann. 2003. “LINE-Mediated Retrotransposition of Marked Alu Sequences.” *Nature Genetics* 35 (1): 41–48. https://doi.org/10.1038/ng1223.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dong, Y., and E. N. Benveniste. 2001. “Immune Function of Astrocytes.” *Glia* 36 (2): 180–90. https://doi.org/10.1002/glia.1107.

Dvinge, Heidi. 2018. “Regulation of Alternative MRNA Splicing: Old Players and New Perspectives.” *FEBS Letters* 592 (17): 2987–3006. https://doi.org/10.1002/1873-3468.13119.

Dvinge, Heidi, and Robert K. Bradley. 2015. “Widespread Intron Retention Diversifies Most Cancer Transcriptomes.” *Genome Medicine* 7 (1): 45. https://doi.org/10.1186/s13073-015-0168-9.

Edwards, Christopher R., William Ritchie, Justin J.-L. Wong, Ulf Schmitz, Robert Middleton, Xiuli An, Narla Mohandas, John E. J. Rasko, and Gerd A. Blobel. 2016. “A Dynamic Intron Retention Program in the Mammalian Megakaryocyte and Erythrocyte Lineages.” *Blood* 127 (17): e24–34. https://doi.org/10.1182/blood-2016-01-692764.

Erwin, Jennifer A., Apuã C. M. Paquola, Tatjana Singer, Iryna Gallina, Mark Novotny, Carolina Quayle, Tracy A. Bedrosian, et al. 2016. “L1-Associated Genomic Regions Are Deleted in Somatic Cells of the Healthy Human Brain.” *Nature Neuroscience* 19 (12): 1583–91. https://doi.org/10.1038/nn.4388.

Fabregat, Antonio, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, et al. 2018. “The Reactome Pathway Knowledgebase.” *Nucleic Acids Research* 46 (D1): D649–55. https://doi.org/10.1093/nar/gkx1132.

Farkash, Evan A., Gary D. Kao, Shane R. Horman, and Eline T. Luning Prak. 2006. “Gamma Radiation Increases Endonuclease-Dependent L1 Retrotransposition

in a Cultured Cell Assay." *Nucleic Acids Research* 34 (4): 1196–1204. https://doi.org/10.1093/nar/gkj522.
Farkash, Evan A., and Eline T. Luning Prak. 2006. "DNA Damage and L1 Retrotransposition." *Journal of Biomedicine and Biotechnology* 2006: 37285. https://doi.org/10.1155/JBB/2006/37285.
Feng, Q., J. V. Moran, H. H. Kazazian, and J. D. Boeke. 1996. "Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition." *Cell* 87 (5): 905–16. https://doi.org/10.1016/s0092-8674(00)81997-2.
Finnegan, D. J. 1997. "Transposable Elements: How Non-LTR Retrotransposons Do It." *Current Biology* 7 (4): R245–48. https://doi.org/10.1016/S0960-9822(06)00112-6.
Fitzgerald, Katherine A. 2011. "The Interferon Inducible Gene: Viperin." *Journal of Interferon & Cytokine Research* 31 (1): 131–35. https://doi.org/10.1089/jir.2010.0127.
Fong, Nova, Hyunmin Kim, Yu Zhou, Xiong Ji, Jinsong Qiu, Tassa Saldi, Katrina Diener, Ken Jones, Xiang-Dong Fu, and David L. Bentley. 2014. "Pre-MRNA Splicing Is Facilitated by an Optimal RNA Polymerase II Elongation Rate." *Genes & Development* 28 (23): 2663–76. https://doi.org/10.1101/gad.252106.114.
Forrest, Scott T., Kurt G. Barringhaus, Demetra Perlegas, Marie-Louise Hammarskjold, and Coleen A. McNamara. 2004. "Intron Retention Generates a Novel Id3 Isoform That Inhibits Vascular Lesion Formation*." *Journal of Biological Chemistry* 279 (31): 32897–903. https://doi.org/10.1074/jbc.M404882200.
Frankiw, Luke, Devdoot Majumdar, Christian Burns, Logan Vlach, Annie Moradian, Michael J. Sweredoski, and David Baltimore. 2019. "BUD13 Promotes a Type I Interferon Response by Countering Intron Retention in Irf7." *Molecular Cell* 73 (4): 803-814.e6. https://doi.org/10.1016/j.molcel.2018.11.038.
Fredi, M., M. Bianchi, L. Andreoli, G. Greco, I. Olivieri, S. Orcesi, E. Fazzi, C. Cereda, and A. Tincani. 2015. "Typing TREX1 Gene in Patients with Systemic Lupus Erythematosus." *Reumatismo* 67 (1): 1–7. https://doi.org/10.4081/reumatismo.2015.782.
Fu, Xiang-Dong. 2017. "Exploiting the Hidden Treasure of Detained Introns." *Cancer Cell* 32 (4): 393–95. https://doi.org/10.1016/j.ccell.2017.09.005.
Fu, Xiang-Dong, and Manuel Ares. 2014. "Context-Dependent Control of Alternative Splicing by RNA-Binding Proteins." *Nature Reviews Genetics* 15 (10): 689–701. https://doi.org/10.1038/nrg3778.
Fukuda, Shinichi, Akhil Varshney, Benjamin J. Fowler, Shao-bin Wang, Siddharth Narendran, Kameshwari Ambati, Tetsuhiro Yasuma, et al. 2021. "Cytoplasmic Synthesis of Endogenous Alu Complementary DNA via Reverse Transcription and Implications in Age-Related Macular Degeneration." *Proceedings of the National Academy of Sciences of the United States of America* 118 (6): e2022751118. https://doi.org/10.1073/pnas.2022751118.
Galante, Pedro Alexandre Favoretto, Noboru Jo Sakabe, Natanja Kirschbaum-Slager, and Sandro José De Souza. 2004. "Detection and Evaluation of Intron Retention Events in the Human Transcriptome." *RNA* 10 (5): 757–65. https://doi.org/10.1261/rna.5123504.
Gall, Alevtina, Piper Treuting, Keith B. Elkon, Yueh-Ming Loo, Michael Gale, Glen N. Barber, and Daniel B. Stetson. 2012. "Autoimmunity Initiates in

Nonhematopoietic Cells and Progresses via Lymphocytes in an Interferon-Dependent Autoimmune Disease." *Immunity* 36 (1): 120–31. https://doi.org/10.1016/j.immuni.2011.11.018.
Gasior, Stephen L., Timothy P. Wakeman, Bo Xu, and Prescott L. Deininger. 2006. "The Human LINE-1 Retrotransposon Creates DNA Double-Strand Breaks." *Journal of Molecular Biology* 357 (5): 1383–93. https://doi.org/10.1016/j.jmb.2006.01.089.
Ge, Zhiyun, Bao Lin Quek, Karen L. Beemon, and J. Robert Hogg. 2016. "Polypyrimidine Tract Binding Protein 1 Protects MRNAs from Recognition by the Nonsense-Mediated MRNA Decay Pathway." *ELife* 5 (January): e11155. https://doi.org/10.7554/eLife.11155.
Gene Ontology Consortium. 2004. "The Gene Ontology (GO) Database and Informatics Resource." *Nucleic Acids Research* 32 (suppl_1): D258–61. https://doi.org/10.1093/nar/gkh036.
Genovesio, Auguste, Yong-Jun Kwon, Marc P. Windisch, Nam Youl Kim, Seo Yeon Choi, Hi Chul Kim, Sungyong Jung, et al. 2011. "Automated Genome-Wide Visual Profiling of Cellular Proteins Involved in HIV Infection." *Journal of Biomolecular Screening* 16 (9): 945–58. https://doi.org/10.1177/1087057111415521.
Giglia-Mari, Giuseppina, Angelika Zotter, and Wim Vermeulen. 2011. "DNA Damage Response." *Cold Spring Harbor Perspectives in Biology* 3 (1): a000745. https://doi.org/10.1101/cshperspect.a000745.
Giordano, Anna Maria Sole, Marco Luciani, Francesca Gatto, Monah Abou Alezz, Chiara Beghè, Lucrezia Della Volpe, Alessandro Migliara, et al. 2022. "DNA Damage Contributes to Neurotoxic Inflammation in Aicardi-Goutières Syndrome Astrocytes." *The Journal of Experimental Medicine* 219 (4): e20211121. https://doi.org/10.1084/jem.20211121.
Goldstone, David C., Valerie Ennis-Adeniran, Joseph J. Hedden, Harriet C. T. Groom, Gillian I. Rice, Evangelos Christodoulou, Philip A. Walker, et al. 2011. "HIV-1 Restriction Factor SAMHD1 Is a Deoxynucleoside Triphosphate Triphosphohydrolase." *Nature* 480 (7377): 379–82. https://doi.org/10.1038/nature10623.
Goodier, John L. 2016. "Restricting Retrotransposons: A Review." *Mobile DNA* 7 (August): 16. https://doi.org/10.1186/s13100-016-0070-z.
Goodsell, David S. 2001. "The Molecular Perspective: Ultraviolet Light and Pyrimidine Dimers." *The Oncologist* 6 (3): 298–99. https://doi.org/10.1634/theoncologist.6-3-298.
Goulielmos, George N., Maria I. Zervou, Vassilis M. Vazgiourakis, Yogita Ghodke-Puranik, Alexandros Garyfallos, and Timothy B. Niewold. 2018. "The Genetics and Molecular Pathogenesis of Systemic Lupus Erythematosus (SLE) in Populations of Different Ancestry." *Gene* 668 (August): 59–72. https://doi.org/10.1016/j.gene.2018.05.041.
Grabski, David F., Lucile Broseus, Bandana Kumari, David Rekosh, Marie-Louise Hammarskjold, and William Ritchie. 2021. "Intron Retention and Its Impact on Gene Expression and Protein Diversity: A Review and a Practical Guide." *WIREs RNA* 12 (1): e1631. https://doi.org/10.1002/wrna.1631.
Gramberg, Thomas, Tanja Kahle, Nicolin Bloch, Sabine Wittmann, Erik Müllers, Waaqo Daddacha, Henning Hofmann, Baek Kim, Dirk Lindemann, and Nathaniel R. Landau. 2013. "Restriction of Diverse Retroviruses by

SAMHD1." *Retrovirology* 10 (1): 26. https://doi.org/10.1186/1742-4690-10-26.
Green, Immanuel D., Natalia Pinello, Renhua Song, Quintin Lee, James M. Halstead, Chau-To Kwok, Alex C. H. Wong, et al. 2020. "Macrophage Development and Activation Involve Coordinated Intron Retention in Key Inflammatory Regulators." *Nucleic Acids Research* 48 (12): 6513–29. https://doi.org/10.1093/nar/gkaa435.
Han, Jeffrey S. 2010. "Non-Long Terminal Repeat (Non-LTR) Retrotransposons: Mechanisms, Recent Developments, and Unanswered Questions." *Mobile DNA* 1 (1): 15. https://doi.org/10.1186/1759-8753-1-15.
Hänzelmann, Sonja, Robert Castelo, and Justin Guinney. 2013. "GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data." *BMC Bioinformatics* 14 (January): 7. https://doi.org/10.1186/1471-2105-14-7.
Harper, J. Wade, and Stephen J. Elledge. 2007. "The DNA Damage Response: Ten Years After." *Molecular Cell* 28 (5): 739–45. https://doi.org/10.1016/j.molcel.2007.11.015.
Hasan, Maroof, James Koch, Dinesh Rakheja, Asit K. Pattnaik, James Brugarolas, Igor Dozmorov, Beth Levine, Edward K. Wakeland, Min Ae Lee-kirsch, and Nan Yan. 2013. "Trex1 Regulates Lysosomal Biogenesis and Interferon-Independent Activation of Antiviral Genes." *Nature Immunology* 14 (1): 61–71. https://doi.org/10.1038/ni.2475.
Havecker, Ericka R., Xiang Gao, and Daniel F. Voytas. 2004. "The Diversity of LTR Retrotransposons." *Genome Biology* 5 (6): 225. https://doi.org/10.1186/gb-2004-5-6-225.
Hayakawa, Sumio, Souichi Shiratori, Hiroaki Yamato, Takeshi Kameyama, Chihiro Kitatsuji, Fumi Kashigi, Showhey Goto, et al. 2011. "ZAPS Is a Potent Stimulator of Signaling Mediated by the RNA Helicase RIG-I during Antiviral Responses." *Nature Immunology* 12 (1): 37–44. https://doi.org/10.1038/ni.1963.
Hiller, Bjoern, Martin Achleitner, Silke Glage, Ronald Naumann, Rayk Behrendt, and Axel Roers. 2012. "Mammalian RNase H2 Removes Ribonucleotides from DNA to Maintain Genome Integrity." *The Journal of Experimental Medicine* 209 (8): 1419–26. https://doi.org/10.1084/jem.20120876.
Hopfner, Karl-Peter, and Veit Hornung. 2020. "Molecular Mechanisms and Cellular Functions of CGAS–STING Signalling." *Nature Reviews Molecular Cell Biology* 21 (9): 501–21. https://doi.org/10.1038/s41580-020-0244-x.
Hou, Shuai, Dajun Qu, Yue Li, Baohui Zhu, Dapeng Liang, Xinyue Wei, Wei Tang, et al. 2019. "XAB2 Depletion Induces Intron Retention in POLR2A to Impair Global Transcription and Promote Cellular Senescence." *Nucleic Acids Research* 47 (15): 8239–54. https://doi.org/10.1093/nar/gkz532.
Huang, Huilin, Hengyou Weng, Wenju Sun, Xi Qin, Hailing Shi, Huizhe Wu, Boxuan Simen Zhao, et al. 2018. "Recognition of RNA N6-Methyladenosine by IGF2BP Proteins Enhances MRNA Stability and Translation." *Nature Cell Biology* 20 (3): 285–95. https://doi.org/10.1038/s41556-018-0045-z.
Huang, Xinwei, Hong Zhang, Xiaoran Guo, Zongxin Zhu, Haibo Cai, and Xiangyang Kong. 2018. "Insulin-like Growth Factor 2 MRNA-Binding Protein 1 (IGF2BP1) in Cancer." *Journal of Hematology & Oncology* 11 (June): 88. https://doi.org/10.1186/s13045-018-0628-y.
Humphrey, Jack, Nicol Birsa, Carmelo Milioto, Martha McLaughlin, Agnieszka M. Ule, David Robaldo, Andrea B. Eberle, et al. 2020. "FUS ALS-Causative

Mutations Impair FUS Autoregulation and Splicing Factor Networks through Intron Retention." *Nucleic Acids Research* 48 (12): 6889–6905. https://doi.org/10.1093/nar/gkaa410.
Hur, Ji-Yeun, Georgia R. Frost, Xianzhong Wu, Christina Crump, Si Jia Pan, Eitan Wong, Marilia Barros, et al. 2020. "The Innate Immunity Protein IFITM3 Modulates γ-Secretase in Alzheimer's Disease." *Nature* 586 (7831): 735–40. https://doi.org/10.1038/s41586-020-2681-2.
Jackson, Stephen P., and Jiri Bartek. 2009. "The DNA-Damage Response in Human Biology and Disease." *Nature* 461 (7267): 1071–78. https://doi.org/10.1038/nature08467.
Jacob, Aishwarya G., and Christopher W. J. Smith. 2017. "Intron Retention as a Component of Regulated Gene Expression Programs." *Human Genetics* 136 (9): 1043–57. https://doi.org/10.1007/s00439-017-1791-x.
Jansz, Natasha. 2019. "DNA Methylation Dynamics at Transposable Elements in Mammals." *Essays in Biochemistry* 63 (6): 677–89. https://doi.org/10.1042/EBC20190039.
Jönsson, Marie E., Raquel Garza, Pia A. Johansson, and Johan Jakobsson. 2020. "Transposable Elements: A Common Feature of Neurodevelopmental and Neurodegenerative Disorders." *Trends in Genetics* 36 (8): 610–23. https://doi.org/10.1016/j.tig.2020.05.004.
Jönsson, Marie E., Per Ludvik Brattås, Charlotte Gustafsson, Rebecca Petri, David Yudovich, Karolina Pircs, Shana Verschuere, et al. 2019. "Activation of Neuronal Genes via LINE-1 Elements upon Global DNA Demethylation in Human Neural Progenitors." *Nature Communications* 10 (1): 3182. https://doi.org/10.1038/s41467-019-11150-8.
Justiz Vaillant, Angel A., Amandeep Goyal, and Matthew Varacallo. 2022. "Systemic Lupus Erythematosus." In *StatPearls*. Treasure Island (FL): StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK535405/.
Kadyrov, Farid A., Leonid Dzantiev, Nicoleta Constantin, and Paul Modrich. 2006. "Endonucleolytic Function of MutLalpha in Human Mismatch Repair." *Cell* 126 (2): 297–308. https://doi.org/10.1016/j.cell.2006.05.039.
Kaer, Kristel, Jelena Branovets, Anni Hallikma, Pilvi Nigumann, and Mart Speek. 2011. "Intronic L1 Retrotransposons and Nested Genes Cause Transcriptional Interference by Inducing Intron Retention, Exonization and Cryptic Polyadenylation." *PloS One* 6 (10): e26099. https://doi.org/10.1371/journal.pone.0026099.
Kaer, Kristel, and Mart Speek. 2012. "Intronic Retroelements." *Mobile Genetic Elements* 2 (3): 154–57. https://doi.org/10.4161/mge.20774.
Kan, Andrey, and Philip D. Hodgkin. 2014. "Mechanisms of Cell Division as Regulators of Acute Immune Response." *Systems and Synthetic Biology* 8 (3): 215–21. https://doi.org/10.1007/s11693-014-9149-3.
Kapitonov, Vladimir V., Sébastien Tempel, and Jerzy Jurka. 2009. "Simple and Fast Classification of Non-LTR Retrotransposons Based on Phylogeny of Their RT Domain Protein Sequences." *Gene* 448 (2): 207–13. https://doi.org/10.1016/j.gene.2009.07.019.
Kaul, Arvind, Caroline Gordon, Mary K. Crow, Zahi Touma, Murray B. Urowitz, Ronald van Vollenhoven, Guillermo Ruiz-Irastorza, and Graham Hughes. 2016. "Systemic Lupus Erythematosus." *Nature Reviews Disease Primers* 2 (1): 1–21. https://doi.org/10.1038/nrdp.2016.39.

Kidwell, Margaret G. 2002. “Transposable Elements and the Evolution of Genome Size in Eukaryotes.” *Genetica* 115 (1): 49–63. https://doi.org/10.1023/A:1016072014259.
Kienzle, N., D. B. Young, D. Liaskou, M. Buck, S. Greco, and T. B. Sculley. 1999. “Intron Retention May Regulate Expression of Epstein-Barr Virus Nuclear Antigen 3 Family Genes.” *Journal of Virology* 73 (2): 1195–1204. https://doi.org/10.1128/JVI.73.2.1195-1204.1999.
Kim, Eddo, Alon Magen, and Gil Ast. 2007. “Different Levels of Alternative Splicing among Eukaryotes.” *Nucleic Acids Research* 35 (1): 125–31. https://doi.org/10.1093/nar/gkl924.
Kim, Hanna, Gina A. Montealegre Sanchez, and Raphaela Goldbach-Mansky. 2016. “Insights from Mendelian Interferonopathies: Comparison of CANDLE, SAVI with AGS, Monogenic Lupus.” *Journal of Molecular Medicine (Berlin, Germany)* 94 (10): 1111–27. https://doi.org/10.1007/s00109-016-1465-5.
Kind, Barbara, Britta Muster, Wolfgang Staroske, Henry D. Herce, René Sachse, Alexander Rapp, Franziska Schmidt, Sarah Koss, M. Cristina Cardoso, and Min Ae Lee-Kirsch. 2014. “Altered Spatio-Temporal Dynamics of RNase H2 Complex Assembly at Replication and Repair Sites in Aicardi-Goutières Syndrome.” *Human Molecular Genetics* 23 (22): 5950–60. https://doi.org/10.1093/hmg/ddu319.
Kines, Kristine J., Mark Sokolowski, Dawn L. deHaro, Claiborne M. Christian, Melody Baddoo, Madison E. Smither, and Victoria P. Belancio. 2016. “The Endonuclease Domain of the LINE-1 ORF2 Protein Can Tolerate Multiple Mutations.” *Mobile DNA* 7 (1): 8. https://doi.org/10.1186/s13100-016-0064-x.
Kitkumthorn, Nakarin, and Apiwat Mutirangura. 2011. “Long Interspersed Nuclear Element-1 Hypomethylation in Cancer: Biology and Clinical Applications.” *Clinical Epigenetics* 2 (2): 315–30. https://doi.org/10.1007/s13148-011-0032-8.
Kojima, Kenji, Misato Baba, Motoki Tsukiashi, Takuto Nishimura, and Kiyoshi Yasukawa. 2018. “RNA/DNA Structures Recognized by RNase H2.” *Briefings in Functional Genomics* 18 (3): 169–73. https://doi.org/10.1093/bfgp/ely024.
Kojima, Kenji K. 2018. “Human Transposable Elements in Repbase: Genomic Footprints from Fish to Humans.” *Mobile DNA* 9 (1): 2. https://doi.org/10.1186/s13100-017-0107-y.
Kolberg, Liis, Uku Raudvere, Ivan Kuzmin, Jaak Vilo, and Hedi Peterson. 2020. “Gprofiler2 -- an R Package for Gene List Functional Enrichment Analysis and Namespace Conversion Toolset g:Profiler.” *F1000Research* 9: ELIXIR-709. https://doi.org/10.12688/f1000research.24956.2.
Kondo, Yutaka, and Jean-Pierre J. Issa. 2003. “Enrichment for Histone H3 Lysine 9 Methylation at Alu Repeats in Human Cells.” *The Journal of Biological Chemistry* 278 (30): 27658–62. https://doi.org/10.1074/jbc.M304072200.
Kramerov, D. A., and N. S. Vassetzky. 2011. “Origin and Evolution of SINEs in Eukaryotic Genomes.” *Heredity* 107 (6): 487–95. https://doi.org/10.1038/hdy.2011.43.
Krestel, Heinz, and Jochen C. Meier. 2018. “RNA Editing and Retrotransposons in Neurology.” *Frontiers in Molecular Neuroscience* 11: 163. https://doi.org/10.3389/fnmol.2018.00163.
Kretschmer, Stefanie, Christine Wolf, Nadja König, Wolfgang Staroske, Jochen Guck, Martin Häusler, Hella Luksch, et al. 2015. “SAMHD1 Prevents Autoimmunity

by Maintaining Genome Stability." *Annals of the Rheumatic Diseases* 74 (3): e17. https://doi.org/10.1136/annrheumdis-2013-204845.
Krokan, Hans E., and Magnar Bjørås. 2013. "Base Excision Repair." *Cold Spring Harbor Perspectives in Biology* 5 (4): a012583. https://doi.org/10.1101/cshperspect.a012583.
Kuhfittig-Kulle, Steffi, Elke Feldmann, Andrea Odersky, Aneta Kuliczkowska, Wolfgang Goedecke, Angelika Eggert, and Petra Pfeiffer. 2007. "The Mutagenic Potential of Non-Homologous End Joining in the Absence of the NHEJ Core Factors Ku70/80, DNA-PKcs and XRCC4-LigIV." *Mutagenesis* 22 (3): 217–33. https://doi.org/10.1093/mutage/gem007.
Lahouassa, Hichem, Waaqo Daddacha, Henning Hofmann, Diana Ayinde, Eric C. Logue, Loïc Dragin, Nicolin Bloch, et al. 2012. "SAMHD1 Restricts the Replication of Human Immunodeficiency Virus Type 1 by Depleting the Intracellular Pool of Deoxynucleoside Triphosphates." *Nature Immunology* 13 (3): 223–28. https://doi.org/10.1038/ni.2236.
Lamers, Mart M., Bernadette G. van den Hoogen, and Bart L. Haagmans. 2019. "ADAR1: 'Editor-in-Chief' of Cytoplasmic Innate Immunity." *Frontiers in Immunology* 10: 1763. https://doi.org/10.3389/fimmu.2019.01763.
Lanciano, Sophie, and Gael Cristofari. 2020. "Measuring and Interpreting Transposable Element Expression." *Nature Reviews Genetics* 21 (12): 721–36. https://doi.org/10.1038/s41576-020-0251-y.
Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, et al. 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature* 409 (6822): 860–921. https://doi.org/10.1038/35057062.
Landolfo, Santo, Marisa Gariglio, Giorgio Gribaudo, and David Lembo. 1998. "The Ifi 200 Genes: An Emerging Family of IFN-Inducible Genes." *Biochimie* 80 (8): 721–28. https://doi.org/10.1016/S0300-9084(99)80025-X.
Lang, Rui, Huiting Li, Xiaoqin Luo, Cencen Liu, Yiwen Zhang, ShunYu Guo, Jingyi Xu, Changshun Bao, Wei Dong, and Yang Yu. 2022. "Expression and Mechanisms of Interferon-Stimulated Genes in Viral Infection of the Central Nervous System (CNS) and Neurological Diseases." *Frontiers in Immunology* 13 (October): 1008072. https://doi.org/10.3389/fimmu.2022.1008072.
Lareau, Liana F., Maki Inada, Richard E. Green, Jordan C. Wengrod, and Steven E. Brenner. 2007. "Unproductive Splicing of SR Genes Associated with Highly Conserved and Ultraconserved DNA Elements." *Nature* 446 (7138): 926–29. https://doi.org/10.1038/nature05676.
Le Page, C., P. Génin, M. G. Baines, and J. Hiscott. 2000. "Interferon Activation and Innate Immunity." *Reviews in Immunogenetics* 2 (3): 374–86.
Lebon, P., J. Badoual, G. Ponsot, F. Goutières, F. Hémeury-Cukier, and J. Aicardi. 1988. "Intrathecal Synthesis of Interferon-Alpha in Infants with Progressive Familial Encephalopathy." *Journal of the Neurological Sciences* 84 (2–3): 201–8. https://doi.org/10.1016/0022-510x(88)90125-6.
Lee, Amanda J., and Ali A. Ashkar. 2018. "The Dual Nature of Type I and Type II Interferons." *Frontiers in Immunology* 9. https://www.frontiersin.org/articles/10.3389/fimmu.2018.02061.
Lee, Ji Yoon, and Seok-Ho Hong. 2019. "Hematopoietic Stem Cells and Their Roles in Tissue Regeneration." *International Journal of Stem Cells* 13 (1): 1–12. https://doi.org/10.15283/ijsc19127.

Lee, Yeon, and Donald C. Rio. 2015. “Mechanisms and Regulation of Alternative Pre-MRNA Splicing.” *Annual Review of Biochemistry* 84: 291–323. https://doi.org/10.1146/annurev-biochem-060614-034316.

Lee-Kirsch, M A, C Wolf, and C Günther. 2014. “Aicardi–Goutières Syndrome: A Model Disease for Systemic Autoimmunity.” *Clinical and Experimental Immunology* 175 (1): 17–24. https://doi.org/10.1111/cei.12160.

Lee-Kirsch, Min Ae, Maolian Gong, Dipanjan Chowdhury, Lydia Senenko, Kerstin Engel, Young-Ae Lee, Udesh de Silva, et al. 2007. “Mutations in the Gene Encoding the 3’-5’ DNA Exonuclease TREX1 Are Associated with Systemic Lupus Erythematosus.” *Nature Genetics* 39 (9): 1065–67. https://doi.org/10.1038/ng2091.

Lenzken, Silvia C., Alessia Loffreda, and Silvia M. L. Barabino. 2013. “RNA Splicing: A New Player in the DNA Damage Response.” *International Journal of Cell Biology* 2013 (September): e153634. https://doi.org/10.1155/2013/153634.

Li, Guo-Min. 2008. “Mechanisms and Functions of DNA Mismatch Repair.” *Cell Research* 18 (1): 85–98. https://doi.org/10.1038/cr.2007.115.

Li, Hong-Dong, Cory C. Funk, Karen McFarland, Eric B. Dammer, Mariet Allen, Minerva M. Carrasquillo, Yona Levites, et al. 2021. “Integrative Functional Genomic Analysis of Intron Retention in Human and Mouse Brain with Alzheimer’s Disease.” *Alzheimer’s & Dementia* 17 (6): 984–1004. https://doi.org/10.1002/alz.12254.

Li, Hui, Padmaja Gade, Weihua Xiao, and Dhan V. Kalvakolanu. 2007. “The Interferon Signaling Network and Transcription Factor C/EBP-β.” *Cellular & Molecular Immunology* 4 (6): 407–18.

Li, Peng, Juan Du, John L. Goodier, Jingwei Hou, Jian Kang, Haig H. Kazazian, Ke Zhao, and Xiao-Fang Yu. 2017. “Aicardi-Goutières Syndrome Protein TREX1 Suppresses L1 and Maintains Genome Integrity through Exonuclease-Independent ORF1p Depletion.” *Nucleic Acids Research* 45 (8): 4619–31. https://doi.org/10.1093/nar/gkx178.

Li, Tuo, and Zhijian J. Chen. 2018. “The CGAS–CGAMP–STING Pathway Connects DNA Damage to Inflammation, Senescence, and Cancer.” *The Journal of Experimental Medicine* 215 (5): 1287–99. https://doi.org/10.1084/jem.20180139.

Li, Xuan, and Wolf-Dietrich Heyer. 2008. “Homologous Recombination in DNA Repair and DNA Damage Tolerance.” *Cell Research* 18 (1): 99–113. https://doi.org/10.1038/cr.2008.1.

Lim, Yoong Wearn, Lionel A Sanz, Xiaoqin Xu, Stella R Hartono, and Frédéric Chédin. 2015. “Genome-Wide DNA Hypomethylation and RNA:DNA Hybrid Accumulation in Aicardi–Goutières Syndrome.” Edited by Bing Ren. *ELife* 4 (July): e08007. https://doi.org/10.7554/eLife.08007.

Lin, Chien-Liang Glenn, Lynn A. Bristol, Lin Jin, Margaret Dykes-Hoberg, Thomas Crawford, Lora Clawson, and Jeffrey D. Rothstein. 1998. “Aberrant RNA Processing in a Neurodegenerative Disease: The Cause for Absent EAAT2, a Glutamate Transporter, in Amyotrophic Lateral Sclerosis.” *Neuron* 20 (3): 589–602. https://doi.org/10.1016/S0896-6273(00)80997-6.

Lindahl, Tomas, Deborah E. Barnes, Yun-Gui Yang, and Peter Robins. 2009a. “Biochemical Properties of Mammalian TREX1 and Its Association with DNA Replication and Inherited Inflammatory Disease.” *Biochemical Society Transactions* 37 (Pt 3): 535–38. https://doi.org/10.1042/BST0370535.

———. 2009b. “Biochemical Properties of Mammalian TREX1 and Its Association with DNA Replication and Inherited Inflammatory Disease.” *Biochemical Society Transactions* 37 (Pt 3): 535–38. https://doi.org/10.1042/BST0370535.
Liu, Ting, Lingyun Zhang, Donghyun Joo, and Shao-Cong Sun. 2017. “NF-KB Signaling in Inflammation.” *Signal Transduction and Targeted Therapy* 2 (July): 17023. https://doi.org/10.1038/sigtrans.2017.23.
Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.
Luisier, Raphaelle, Giulia E. Tyzack, Claire E. Hall, Jamie S. Mitchell, Helen Devine, Doaa M. Taha, Bilal Malik, et al. 2018. “Intron Retention and Nuclear Loss of SFPQ Are Molecular Hallmarks of ALS.” *Nature Communications* 9 (1): 2010. https://doi.org/10.1038/s41467-018-04373-8.
Mackenzie, Karen J., Paula Carroll, Laura Lettice, Žygimantė Tarnauskaitė, Kaalak Reddy, Flora Dix, Ailsa Revuelta, et al. 2016. “Ribonuclease H2 Mutations Induce a CGAS/STING-Dependent Innate Immune Response.” *The EMBO Journal* 35 (8): 831–44. https://doi.org/10.15252/embj.201593339.
Mackenzie, Karen J., Paula Carroll, Carol-Anne Martin, Olga Murina, Adeline Fluteau, Daniel J. Simpson, Nelly Olova, et al. 2017. “CGAS Surveillance of Micronuclei Links Genome Instability to Innate Immunity.” *Nature* 548 (7668): 461–65. https://doi.org/10.1038/nature23449.
Maes, Olivier C., Suying Xu, Bo Yu, Howard M. Chertkow, Eugenia Wang, and Hyman M. Schipper. 2007. “Transcriptional Profiling of Alzheimer Blood Mononuclear Cells by Microarray.” *Neurobiology of Aging* 28 (12): 1795–1809. https://doi.org/10.1016/j.neurobiolaging.2006.08.004.
Mannion, Niamh M., Sam M. Greenwood, Robert Young, Sarah Cox, James Brindle, David Read, Christoffer Nellåker, et al. 2014. “The RNA-Editing Enzyme ADAR1 Controls Innate Immune Responses to RNA.” *Cell Reports* 9 (4): 1482–94. https://doi.org/10.1016/j.celrep.2014.10.041.
Maquat, Lynne E. 2020. “Short Interspersed Nuclear Element (SINE)-Mediated Post-Transcriptional Effects on Human and Mouse Gene Expression: SINE-UP for Active Duty.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1795): 20190344. https://doi.org/10.1098/rstb.2019.0344.
Martin, Sandra L. 2006. “The ORF1 Protein Encoded by LINE-1: Structure and Function During L1 Retrotransposition.” *Journal of Biomedicine and Biotechnology* 2006: 45621. https://doi.org/10.1155/JBB/2006/45621.
Mauger, Oriane, Frédéric Lemoine, and Peter Scheiffele. 2016. “Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity.” *Neuron* 92 (6): 1266–78. https://doi.org/10.1016/j.neuron.2016.11.032.
Mazille, Maxime, Katarzyna Buczak, Peter Scheiffele, and Oriane Mauger. 2022. “Stimulus-Specific Remodeling of the Neuronal Transcriptome through Nuclear Intron-Retaining Transcripts.” *The EMBO Journal* n/a (n/a): e110192. https://doi.org/10.15252/embj.2021110192.
McClintock, Barbara. 1965. “Components of Action of the Regulators Spm and Ac.” In , 527–36. Washington, D.C.: Carnegie Institution of Washington. https://archive.org/stream/yearbookcarne64196465carn#page/526/mode/2up/search/Mcclintock.

McCoy, Matthew J., and Andrew Z. Fire. 2020. “Intron and Gene Size Expansion during Nervous System Evolution.” *BMC Genomics* 21 (1): 360. https://doi.org/10.1186/s12864-020-6760-4.
Middleton, Robert, Dadi Gao, Aubin Thomas, Babita Singh, Amy Au, Justin J-L Wong, Alexandra Bomane, et al. 2017. “IRFinder: Assessing the Impact of Intron Retention on Mammalian Gene Expression.” *Genome Biology* 18 (1): 51. https://doi.org/10.1186/s13059-017-1184-4.
Mills, Ryan E., E. Andrew Bennett, Rebecca C. Iskow, and Scott E. Devine. 2007. “Which Transposable Elements Are Active in the Human Genome?” *Trends in Genetics* 23 (4): 183–91. https://doi.org/10.1016/j.tig.2007.02.006.
Mistry, Jaina, Sara Chuguransky, Lowri Williams, Matloob Qureshi, Gustavo A. Salazar, Erik L. L. Sonnhammer, Silvio C. E. Tosatto, et al. 2021. “Pfam: The Protein Families Database in 2021.” *Nucleic Acids Research* 49 (D1): D412–19. https://doi.org/10.1093/nar/gkaa913.
Moldovan, John B., and John V. Moran. 2015. “The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition.” *PLOS Genetics* 11 (5): e1005121. https://doi.org/10.1371/journal.pgen.1005121.
Monteuuis, Geoffray, Justin J L Wong, Charles G Bailey, Ulf Schmitz, and John E J Rasko. 2019. “The Changing Paradigm of Intron Retention: Regulation, Ramifications and Recipes.” *Nucleic Acids Research* 47 (22): 11497–513. https://doi.org/10.1093/nar/gkz1068.
Motwani, Mona, Scott Pesiridis, and Katherine A. Fitzgerald. 2019. “DNA Sensing by the CGAS–STING Pathway in Health and Disease.” *Nature Reviews Genetics* 20 (11): 657–74. https://doi.org/10.1038/s41576-019-0151-1.
Muniz, Lisa, Estelle Nicolas, and Didier Trouche. 2021. “RNA Polymerase II Speed: A Key Player in Controlling and Adapting Transcriptome Composition.” *The EMBO Journal* 40 (15): e105740. https://doi.org/10.15252/embj.2020105740.
Murira, Armstrong, and Alain Lamarre. 2016. “Type-I Interferon Responses: From Friend to Foe in the Battle against Chronic Viral Infection.” *Frontiers in Immunology* 7. https://www.frontiersin.org/articles/10.3389/fimmu.2016.00609.
Naughton, Bartholomew J., F. Jason Duncan, Darren A. Murrey, Aaron S. Meadows, David E. Newsom, Nicoleta Stoicea, Peter White, Douglas W. Scharre, Douglas M. Mccarty, and Haiyan Fu. 2015. “Blood Genome-Wide Transcriptional Profiles Reflect Broad Molecular Impairments and Strong Blood-Brain Links in Alzheimer’s Disease.” *Journal of Alzheimer’s Disease : JAD* 43 (1): 93–108. https://doi.org/10.3233/JAD-140606.
Ni, Ting, Wenjing Yang, Miao Han, Yubo Zhang, Ting Shen, Hongbo Nie, Zhihui Zhou, et al. 2016. “Global Intron Retention Mediated Gene Regulation during CD4+ T Cell Activation.” *Nucleic Acids Research* 44 (14): 6817–29. https://doi.org/10.1093/nar/gkw591.
Nigrovic, Peter A., Pui Y. Lee, and Hal M. Hoffman. 2020. “Monogenic Autoinflammatory Disorders: Conceptual Overview, Phenotype, and Clinical Approach.” *The Journal of Allergy and Clinical Immunology* 146 (5): 925–37. https://doi.org/10.1016/j.jaci.2020.08.017.
Nishikura, Kazuko. 2010. “Functions and Regulation of RNA Editing by ADAR Deaminases.” *Annual Review of Biochemistry* 79: 321–49. https://doi.org/10.1146/annurev-biochem-060208-105251.
O’Connor, Mark J. 2015. “Targeting the DNA Damage Response in Cancer.” *Molecular Cell* 60 (4): 547–60. https://doi.org/10.1016/j.molcel.2015.10.040.

Oda, Hirotsugu, Kenji Nakagawa, Junya Abe, Tomonari Awaya, Masahide Funabiki, Atsushi Hijikata, Ryuta Nishikomori, et al. 2014. “Aicardi-Goutières Syndrome Is Caused by IFIH1 Mutations.” *American Journal of Human Genetics* 95 (1): 121–25. https://doi.org/10.1016/j.ajhg.2014.06.007.

O’Grady, Tina M., Melody Baddoo, Samuel A. Flemington, Eman Y. Ishaq, Nathan A. Ungerleider, and Erik K. Flemington. 2022. “Reversal of Splicing Infidelity Is a Pre-Activation Step in B Cell Differentiation.” *Frontiers in Immunology* 13. https://www.frontiersin.org/articles/10.3389/fimmu.2022.1060114.

Ohtani, Hitoshi, and Yuka W. Iwasaki. 2021. “Rewiring of Chromatin State and Gene Expression by Transposable Elements.” *Development, Growth & Differentiation* 63 (4–5): 262–73. https://doi.org/10.1111/dgd.12735.

Ong, Chin-Tong, and Swarnaseetha Adusumalli. 2019. “Increased Intron Retention Is Linked to Alzheimer’s Disease.” *Neural Regeneration Research* 15 (2): 259–60. https://doi.org/10.4103/1673-5374.265549.

Orecchini, Elisa, Margherita Doria, Ambra Antonioni, Silvia Galardi, Silvia Anna Ciafrè, Loredana Frassinelli, Carmine Mancone, Claudia Montaldo, Marco Tripodi, and Alessandro Michienzi. 2017. “ADAR1 Restricts LINE-1 Retrotransposition.” *Nucleic Acids Research* 45 (1): 155–68. https://doi.org/10.1093/nar/gkw834.

Orecchini, Elisa, Loredana Frassinelli, and Alessandro Michienzi. 2017. “Restricting Retrotransposons: ADAR1 Is Another Guardian of the Human Genome.” *RNA Biology* 14 (11): 1485–91. https://doi.org/10.1080/15476286.2017.1341033.

Ostertag, E. M., and H. H. Kazazian. 2001. “Biology of Mammalian L1 Retrotransposons.” *Annual Review of Genetics* 35: 501–38. https://doi.org/10.1146/annurev.genet.35.102401.091032.

Papanikolaou, Sofia, George K. Bertsias, and Christoforos Nikolaou. 2021. “Extensive Changes in Transcription Dynamics Reflected on Alternative Splicing Events in Systemic Lupus Erythematosus Patients.” *Genes* 12 (8): 1260. https://doi.org/10.3390/genes12081260.

Park, Sang-Je, Jae-Won Huh, Young-Hyun Kim, Heui-Soo Kim, and Kyu-Tae Chang. 2012. “Intron Retention and TE Exonization Events in ZRANB2.” *Comparative and Functional Genomics* 2012: 170208. https://doi.org/10.1155/2012/170208.

Patchsung, Maturada, Sirapat Settayanon, Monnat Pongpanich, Dharm Mutirangura, Pornrutsami Jintarith, and Apiwat Mutirangura. 2018. “Alu SiRNA to Increase Alu Element Methylation and Prevent DNA Damage.” *Epigenomics* 10 (2): 175–85. https://doi.org/10.2217/epi-2017-0096.

Payer, Lindsay M., and Kathleen H. Burns. 2019. “Transposable Elements in Human Genetic Disease.” *Nature Reviews Genetics* 20 (12): 760–72. https://doi.org/10.1038/s41576-019-0165-8.

Petasny, Mayra, Mercedes Bentata, Andrea Pawellek, Mai Baker, Gillian Kay, and Maayan Salton. 2021. “Splicing to Keep Cycling: The Importance of Pre-MRNA Splicing during the Cell Cycle.” *Trends in Genetics* 37 (3): 266–78. https://doi.org/10.1016/j.tig.2020.08.013.

Petrova, Veronika, Renhua Song, Deep Consortium, Karl J. V. Nordström, Jörn Walter, Justin J.-L. Wong, Nicola J. Armstrong, John E. J. Rasko, and Ulf Schmitz. 2021. “Chromatin Accessibility Determines Intron Retention in a Cell Type-Specific Manner.” bioRxiv. https://doi.org/10.1101/2021.02.17.431609.

Peze-Heidsieck, Eugenie, Tom Bonnifet, Rania Znaidi, Camille Ravel-Godreuil, Olivia Massiani-Beaudoin, Rajiv L. Joshi, and Julia Fuchs. 2022.

"Retrotransposons as a Source of DNA Damage in Neurodegeneration." *Frontiers in Aging Neuroscience* 13. https://www.frontiersin.org/article/10.3389/fnagi.2021.786897.

Pimentel, Harold, Marilyn Parra, Sherry L. Gee, Narla Mohandas, Lior Pachter, and John G. Conboy. 2016. "A Dynamic Intron Retention Program Enriched in RNA Processing Genes Regulates Gene Expression during Terminal Erythropoiesis." *Nucleic Acids Research* 44 (2): 838–51. https://doi.org/10.1093/nar/gkv1168.

Pinson, Marie-Elisa, Franck Court, Aymeric Masson, Yoan Renaud, Allison Fantini, Ophélie Bacoeur-Ouzillou, Marie Barriere, et al. 2022. "L1 Chimeric Transcripts Are Expressed in Healthy Brain and Their Deregulation in Glioma Follows That of Their Host Locus." *Human Molecular Genetics* 31 (15): 2606–22. https://doi.org/10.1093/hmg/ddac056.

Piriyapongsa, Jittima, Nalini Polavarapu, Mark Borodovsky, and John McDonald. 2007. "Exonization of the LTR Transposable Elements in Human Genome." *BMC Genomics* 8 (1): 291. https://doi.org/10.1186/1471-2164-8-291.

Piskurek, Oliver, Masato Nikaido, Boeadi, Minoru Baba, and Norihiro Okada. 2003. "Unique Mammalian TRNA-Derived Repetitive Elements in Dermopterans: The t-SINE Family and Its Retrotransposition Through Multiple Sources." *Molecular Biology and Evolution* 20 (10): 1659–68. https://doi.org/10.1093/molbev/msg187.

Pizzi, Sara, Sarah Sertic, Simona Orcesi, Cristina Cereda, Marika Bianchi, Andrew P. Jackson, Federico Lazzaro, Paolo Plevani, and Marco Muzi-Falconi. 2015. "Reduction of HRNase H2 Activity in Aicardi–Goutières Syndrome Cells Leads to Replication Stress and Genome Instability." *Human Molecular Genetics* 24 (3): 649–58. https://doi.org/10.1093/hmg/ddu485.

Platt, Roy N., Michael W. Vandewege, and David A. Ray. 2018. "Mammalian Transposable Elements and Their Impacts on Genome Evolution." *Chromosome Research* 26 (1): 25–43. https://doi.org/10.1007/s10577-017-9570-z.

Pokatayev, Vladislav, Naushaba Hasin, Hyongi Chon, Susana M. Cerritelli, Kiran Sakhuja, Jerrold M. Ward, H. Douglas Morris, Nan Yan, and Robert J. Crouch. 2016. "RNase H2 Catalytic Core Aicardi-Goutières Syndrome-Related Mutant Invokes CGAS-STING Innate Immune-Sensing Pathway in Mice." *The Journal of Experimental Medicine* 213 (3): 329–36. https://doi.org/10.1084/jem.20151464.

Potenski, Catherine J., Anastasiya Epshtein, Christopher Bianco, and Hannah L. Klein. 2019. "Genome Instability Consequences of RNase H2 Aicardi-Goutières Syndrome Alleles." *DNA Repair*, RNases H: multiples roles in maintaining genome integrity, 84 (December): 102614. https://doi.org/10.1016/j.dnarep.2019.04.002.

Powell, Rebecca D., Paul J. Holland, Thomas Hollis, and Fred W. Perrino. 2011. "Aicardi-Goutières Syndrome Gene and HIV-1 Restriction Factor SAMHD1 Is a DGTP-Regulated Deoxynucleotide Triphosphohydrolase♦." *The Journal of Biological Chemistry* 286 (51): 43596–600. https://doi.org/10.1074/jbc.C111.317628.

Ravel-Godreuil, Camille, Olivia Massiani-Beaudoin, Philippe Mailly, Alain Prochiantz, Rajiv L. Joshi, and Julia Fuchs. 2021. "Perturbed DNA Methylation by Gadd45b Induces Chromatin Disorganization, DNA Strand

Breaks and Dopaminergic Neuron Death." *IScience* 24 (7): 102756. https://doi.org/10.1016/j.isci.2021.102756.
Ray, Debashish, Hilal Kazan, Kate B. Cook, Matthew T. Weirauch, Hamed S. Najafabadi, Xiao Li, Serge Gueroussov, et al. 2013. "A Compendium of RNA-Binding Motifs for Decoding Gene Regulation." *Nature* 499 (7457): 172–77. https://doi.org/10.1038/nature12311.
Reilly, Matthew T., Geoffrey J. Faulkner, Joshua Dubnau, Igor Ponomarev, and Fred H. Gage. 2013. "The Role of Transposable Elements in Health and Diseases of the Central Nervous System." *The Journal of Neuroscience* 33 (45): 17577–86. https://doi.org/10.1523/JNEUROSCI.3369-13.2013.
Ren, Pingping, Luying Lu, Shasha Cai, Jianghua Chen, Weiqiang Lin, and Fei Han. 2021. "Alternative Splicing: A New Cause and Potential Therapeutic Target in Autoimmune Disease." *Frontiers in Immunology* 12 (August): 713540. https://doi.org/10.3389/fimmu.2021.713540.
Rice, Gillian I., Yoandris Del Toro Duany, Emma M. Jenkinson, Gabriella Ma Forte, Beverley H. Anderson, Giada Ariaudo, Brigitte Bader-Meunier, et al. 2014. "Gain-of-Function Mutations in IFIH1 Cause a Spectrum of Human Disease Phenotypes Associated with Upregulated Type I Interferon Signaling." *Nature Genetics* 46 (5): 503–9. https://doi.org/10.1038/ng.2933.
Rice, Gillian I, Gabriella M A Forte, Marcin Szynkiewicz, Diana S Chase, Alec Aeby, Mohamed S Abdel-Hamid, Sam Ackroyd, et al. 2013. "Assessment of Interferon-Related Biomarkers in Aicardi-Goutières Syndrome Associated with Mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, and ADAR: A Case-Control Study." *The Lancet. Neurology* 12 (12): 1159–69. https://doi.org/10.1016/S1474-4422(13)70258-8.
Rice, Gillian I., Paul R. Kasher, Gabriella M. A. Forte, Niamh M. Mannion, Sam M. Greenwood, Marcin Szynkiewicz, Jonathan E. Dickerson, et al. 2012. "Mutations in ADAR1 Cause Aicardi-Goutières Syndrome Associated with a Type I Interferon Signature." *Nature Genetics* 44 (11): 1243–48. https://doi.org/10.1038/ng.2414.
Rice, Gillian I., Candice Meyzer, Naïm Bouazza, Marie Hully, Nathalie Boddaert, Michaela Semeraro, Leo A. H. Zeef, et al. 2018. "Reverse-Transcriptase Inhibitors in the Aicardi–Goutières Syndrome." *The New England Journal of Medicine* 379 (23): 2275–77. https://doi.org/10.1056/NEJMc1810983.
Richardson, Sandra R., Aurélien J. Doucet, Huira C. Kopera, John B. Moldovan, José Luis Garcia-Pérez, and John V. Moran. 2015. "The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes." *Microbiology Spectrum* 3 (2): 10.1128/microbiolspec.MDNA3-0061–2014. https://doi.org/10.1128/microbiolspec.MDNA3-0061-2014.
Rodero, Mathieu P., Alessandra Tesser, Eva Bartok, Gillian I. Rice, Erika Della Mina, Marine Depp, Benoit Beitz, et al. 2017. "Type I Interferon-Mediated Autoinflammation Due to DNase II Deficiency." *Nature Communications* 8 (December): 2176. https://doi.org/10.1038/s41467-017-01932-3.
Rosenblum, Michael D., Kelly A. Remedios, and Abul K. Abbas. 2015. "Mechanisms of Human Autoimmunity." *The Journal of Clinical Investigation* 125 (6): 2228–33. https://doi.org/10.1172/JCI78088.
Roy, Ethan, and Wei Cao. 2022. "Glial Interference: Impact of Type I Interferon in Neurodegenerative Diseases." *Molecular Neurodegeneration* 17 (1): 78. https://doi.org/10.1186/s13024-022-00583-3.

Roy, Ethan R., Gabriel Chiu, Sanming Li, Nicholas E. Propson, Rupa Kanchi, Baiping Wang, Cristian Coarfa, Hui Zheng, and Wei Cao. 2022. “Concerted Type I Interferon Signaling in Microglia and Neural Cells Promotes Memory Impairment Associated with Amyloid β Plaques.” *Immunity* 55 (5): 879-894.e6. https://doi.org/10.1016/j.immuni.2022.03.018.

Ruijtenberg, Suzan, and Sander van den Heuvel. 2016. “Coordinating Cell Proliferation and Differentiation: Antagonism between Cell Cycle Regulators and Cell Type-Specific Gene Expression.” *Cell Cycle (Georgetown, Tex.)* 15 (2): 196–212. https://doi.org/10.1080/15384101.2015.1120925.

Sakabe, Noboru Jo, and Sandro José de Souza. 2007. “Sequence Features Responsible for Intron Retention in Human.” *BMC Genomics* 8 (1): 59. https://doi.org/10.1186/1471-2164-8-59.

Saleh, Aurian, Angela Macia, and Alysson R. Muotri. 2019. “Transposable Elements, Inflammation, and Neurological Disease.” *Frontiers in Neurology* 10. https://doi.org/10.3389/fneur.2019.00894.

Samuelson, L. C., K. Wiebauer, C. M. Snow, and M. H. Meisler. 1990. “Retroviral and Pseudogene Insertion Sites Reveal the Lineage of Human Salivary and Pancreatic Amylase Genes from a Single Gene during Primate Evolution.” *Molecular and Cellular Biology* 10 (6): 2513–20. https://doi.org/10.1128/mcb.10.6.2513-2520.1990.

Sanz, Ignacio, Chungwen Wei, Scott A. Jenks, Kevin S. Cashman, Christopher Tipton, Matthew C. Woodruff, Jennifer Hom, and F. Eun-Hyung Lee. 2019. “Challenges and Opportunities for Consistent Classification of Human B Cell and Plasma Cell Populations.” *Frontiers in Immunology* 10. https://www.frontiersin.org/articles/10.3389/fimmu.2019.02458.

Sase, Sunetra, Asako Takanohashi, Adeline Vanderver, and Akshata Almad. 2018. “Astrocytes, an Active Player in Aicardi–Goutières Syndrome.” *Brain Pathology* 28 (3): 399–407. https://doi.org/10.1111/bpa.12600.

Sassaman, D. M., B. A. Dombroski, J. V. Moran, M. L. Kimberland, T. P. Naas, R. J. DeBerardinis, A. Gabriel, G. D. Swergold, and H. H. Kazazian. 1997. “Many Human L1 Elements Are Capable of Retrotransposition.” *Nature Genetics* 16 (1): 37–43. https://doi.org/10.1038/ng0597-37.

Scharer, Christopher D., Emily L. Blalock, Tian Mi, Benjamin G. Barwick, Scott A. Jenks, Tsuneo Deguchi, Kevin S. Cashman, et al. 2019. “Epigenetic Programming Underpins B Cell Dysfunction in Human SLE.” *Nature Immunology* 20 (8): 1071–82. https://doi.org/10.1038/s41590-019-0419-9.

Schmitz, Ulf, Natalia Pinello, Fangzhi Jia, Sultan Alasmari, William Ritchie, Maria-Cristina Keightley, Shaniko Shini, Graham J. Lieschke, Justin J-L Wong, and John E. J. Rasko. 2017. “Intron Retention Enhances Gene Regulatory Complexity in Vertebrates.” *Genome Biology* 18 (1): 216. https://doi.org/10.1186/s13059-017-1339-3.

Schneider, William M., Meike Dittmann Chevillotte, and Charles M. Rice. 2014. “Interferon-Stimulated Genes: A Complex Web of Host Defenses.” *Annual Review of Immunology* 32: 513–45. https://doi.org/10.1146/annurev-immunol-032713-120231.

Schweingruber, Christoph, Simone C. Rufener, David Zünd, Akio Yamashita, and Oliver Mühlemann. 2013. “Nonsense-Mediated MRNA Decay - Mechanisms of Substrate MRNA Recognition and Degradation in Mammalian Cells.” *Biochimica Et Biophysica Acta* 1829 (6–7): 612–23. https://doi.org/10.1016/j.bbagrm.2013.02.005.

Shaban, Nadine M., Scott Harvey, Fred W. Perrino, and Thomas Hollis. 2010. "The Structure of the Mammalian RNase H2 Complex Provides Insight into RNA·DNA Hybrid Processing to Prevent Immune Dysfunction." *The Journal of Biological Chemistry* 285 (6): 3617–24. https://doi.org/10.1074/jbc.M109.059048.

Sharangdhar, Tejaswini, Yoichiro Sugimoto, Jacqueline Heraud-Farlow, Sandra M Fernández-Moya, Janina Ehses, Igor Ruiz de los Mozos, Jernej Ule, and Michael A Kiebler. 2017. "A Retained Intron in the 3′-UTR of Calm3 MRNA Mediates Its Staufen2- and Activity-dependent Localization to Neuronal Dendrites." *EMBO Reports* 18 (10): 1762–74. https://doi.org/10.15252/embr.201744334.

Shen, Shihao, Juw Won Park, Zhi-xiang Lu, Lan Lin, Michael D. Henry, Ying Nian Wu, Qing Zhou, and Yi Xing. 2014. "RMATS: Robust and Flexible Detection of Differential Alternative Splicing from Replicate RNA-Seq Data." *Proceedings of the National Academy of Sciences of the United States of America* 111 (51): E5593-5601. https://doi.org/10.1073/pnas.1419161111.

Shivshankar, Pooja, Baharan Fekry, Kristin Eckel-Mahan, and Rick A. Wetsel. 2020. "Circadian Clock and Complement Immune System—Complementary Control of Physiology and Pathology?" *Frontiers in Cellular and Infection Microbiology* 10 (August): 418. https://doi.org/10.3389/fcimb.2020.00418.

Shlomchik, M. J., J. E. Craft, and M. J. Mamula. 2001. "From T to B and Back Again: Positive Feedback in Systemic Autoimmune Disease." *Nature Reviews. Immunology* 1 (2): 147–53. https://doi.org/10.1038/35100573.

Simpson, Sean R., Wayne O. Hemphill, Teesha Hudson, and Fred W. Perrino. 2020. "TREX1 – Apex Predator of Cytosolic DNA Metabolism." *DNA Repair* 94 (October): 102894. https://doi.org/10.1016/j.dnarep.2020.102894.

Song, Renhua, Shweta Tikoo, Rohit Jain, Natalia Pinello, Amy Y. M. Au, Rajini Nagarajah, Bo Porse, John E. J. Rasko, and Justin J.-L. Wong. 2022. "Dynamic Intron Retention Modulates Gene Expression in the Monocytic Differentiation Pathway." *Immunology* 165 (2): 274–86. https://doi.org/10.1111/imm.13435.

Speek, Mart. 2001. "Antisense Promoter of Human L1 Retrotransposon Drives Transcription of Adjacent Cellular Genes." *Molecular and Cellular Biology* 21 (6): 1973–85. https://doi.org/10.1128/MCB.21.6.1973-1985.2001.

Steiman-Shimony, Avital, Orr Shtrikman, and Hanah Margalit. 2018. "Assessing the Functional Association of Intronic MiRNAs with Their Host Genes." *RNA* 24 (8): 991–1004. https://doi.org/10.1261/rna.064386.117.

Stetson, Daniel B., Joan S. Ko, Thierry Heidmann, and Ruslan Medzhitov. 2008. "Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity." *Cell* 134 (4): 587–98. https://doi.org/10.1016/j.cell.2008.06.032.

Su, Zhongjing, and Dongyang Huang. 2021. "Alternative Splicing of Pre-MRNA in the Control of Immune Activity." *Genes* 12 (4): 574. https://doi.org/10.3390/genes12040574.

Sundaram, Vasavi, and Joanna Wysocka. 2020. "Transposable Elements as a Potent Source of Diverse Cis-Regulatory Sequences in Mammalian Genomes." *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1795): 20190347. https://doi.org/10.1098/rstb.2019.0347.

Szak, Suzanne T., Oxana K. Pickeral, Wojciech Makalowski, Mark S. Boguski, David Landsman, and Jef D. Boeke. 2002. "Molecular Archeology of L1 Insertions in the Human Genome." *Genome Biology* 3 (10): research0052.1. https://doi.org/10.1186/gb-2002-3-10-research0052.

Sznajder, Łukasz J., James D. Thomas, Ellie M. Carrell, Tammy Reid, Karen N. McFarland, John D. Cleary, Ruan Oliveira, et al. 2018. “Intron Retention Induced by Microsatellite Expansions as a Disease Biomarker.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (16): 4234–39. https://doi.org/10.1073/pnas.1716617115.

Tam, Oliver H., Lyle W. Ostrow, and Molly Gale Hammell. 2019. “Diseases of the NERVous System: Retrotransposon Activity in Neurodegenerative Disease.” *Mobile DNA* 10 (1): 32. https://doi.org/10.1186/s13100-019-0176-1.

Thomas, Charles A., Leon Tejwani, Cleber A. Trujillo, Priscilla D. Negraes, Roberto H. Herai, Pinar Mesci, Angela Macia, Yanick J. Crow, and Alysson R. Muotri. 2017. “Modeling of TREX1-Dependent Autoimmune Disease Using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation.” *Cell Stem Cell* 21 (3): 319-331.e8. https://doi.org/10.1016/j.stem.2017.07.009.

Tiňo, Peter. 2009. “Basic Properties and Information Theory of Audic-Claverie Statistic for Analyzing CDNA Arrays.” *BMC Bioinformatics* 10 (1): 310. https://doi.org/10.1186/1471-2105-10-310.

Tolmie, J. L., P. Shillito, R. Hughes-Benzie, and J. B. Stephenson. 1995. “The Aicardi-Goutières Syndrome (Familial, Early Onset Encephalopathy with Calcifications of the Basal Ganglia and Chronic Cerebrospinal Fluid Lymphocytosis).” *Journal of Medical Genetics* 32 (11): 881–84. https://doi.org/10.1136/jmg.32.11.881.

Tyzack, Giulia E, Jacob Neeves, Hamish Crerar, Pierre Klein, Oliver Ziff, Doaa M Taha, Raphaëlle Luisier, Nicholas M Luscombe, and Rickie Patani. 2021. “Aberrant Cytoplasmic Intron Retention Is a Blueprint for RNA Binding Protein Mislocalization in VCP-Related Amyotrophic Lateral Sclerosis.” *Brain* 144 (7): 1985–93. https://doi.org/10.1093/brain/awab078.

Ullrich, Sebastian, and Roderic Guigó. 2020. “Dynamic Changes in Intron Retention Are Tightly Associated with Regulation of Splicing Factors and Proliferative Activity during B-Cell Development.” *Nucleic Acids Research* 48 (3): 1327–40. https://doi.org/10.1093/nar/gkz1180.

Vanichkina, Darya P., Ulf Schmitz, Justin J. -L. Wong, and John E. J. Rasko. 2018. “Challenges in Defining the Role of Intron Retention in Normal Biology and Disease.” *Seminars in Cell & Developmental Biology*, Diversity of transcripts emanating from protein-coding genes, 75 (March): 40–49. https://doi.org/10.1016/j.semcdb.2017.07.030.

Volkman, Hannah E., and Daniel B. Stetson. 2014. “The Enemy within: Endogenous Retroelements and Autoimmune Disease.” *Nature Immunology* 15 (5): 415–22. https://doi.org/10.1038/ni.2872.

Vries, B. de, G. M. Steup-Beekman, J. Haan, E. L. Bollen, J. Luyendijk, R. R. Frants, G. M. Terwindt, et al. 2010. “TREX1 Gene Variant in Neuropsychiatric Systemic Lupus Erythematosus.” *Annals of the Rheumatic Diseases* 69 (10): 1886–87. https://doi.org/10.1136/ard.2009.114157.

Wahl, Markus C., Cindy L. Will, and Reinhard Lührmann. 2009. “The Spliceosome: Design Principles of a Dynamic RNP Machine.” *Cell* 136 (4): 701–18. https://doi.org/10.1016/j.cell.2009.02.009.

Wang, Qingqing, Erin G. Conlon, James L. Manley, and Donald C. Rio. 2020. “Widespread Intron Retention Impairs Protein Homeostasis in C9orf72 ALS Brains.” *Genome Research* 30 (12): 1705–15. https://doi.org/10.1101/gr.265298.120.

Wang, Xiao-Lan, and Lianjian Li. 2021. “Circadian Clock Regulates Inflammation and the Development of Neurodegeneration.” *Frontiers in Cellular and Infection Microbiology* 11. https://www.frontiersin.org/articles/10.3389/fcimb.2021.696554.

Wang, Yan, Jing Liu, B. O. Huang, Yan-Mei Xu, Jing Li, Lin-Feng Huang, Jin Lin, et al. 2015. “Mechanism of Alternative Splicing and Its Regulation.” *Biomedical Reports* 3 (2): 152–58. https://doi.org/10.3892/br.2014.407.

Wang, Yan, Fan Wang, and Xiaolei Zhang. 2021. “STING-Associated Vasculopathy with Onset in Infancy: A Familial Case Series Report and Literature Review.” *Annals of Translational Medicine* 9 (2): 176. https://doi.org/10.21037/atm-20-6198.

Werner, Markus, and Hassan Jumaa. 2015. “Chapter 6 - Proliferation and Differentiation Programs of Developing B Cells.” In *Molecular Biology of B Cells (Second Edition)*, edited by Frederick W. Alt, Tasuku Honjo, Andreas Radbruch, and Michael Reth, 75–97. London: Academic Press. https://doi.org/10.1016/B978-0-12-397933-9.00006-0.

Wessler, Susan R. 2006. “Transposable Elements and the Evolution of Eukaryotic Genomes.” *Proceedings of the National Academy of Sciences* 103 (47): 17600–601. https://doi.org/10.1073/pnas.0607612103.

Wicker, Thomas, François Sabot, Aurélie Hua-Van, Jeffrey L. Bennetzen, Pierre Capy, Boulos Chalhoub, Andrew Flavell, et al. 2007. “A Unified Classification System for Eukaryotic Transposable Elements.” *Nature Reviews Genetics* 8 (12): 973–82. https://doi.org/10.1038/nrg2165.

Wilkinson, Max E., Clément Charenton, and Kiyoshi Nagai. 2020. “RNA Splicing by the Spliceosome.” *Annual Review of Biochemistry* 89 (1): 359–88. https://doi.org/10.1146/annurev-biochem-091719-064225.

Wong, Justin J.-L., Dadi Gao, Trung V. Nguyen, Chau-To Kwok, Michelle van Geldermalsen, Rob Middleton, Natalia Pinello, et al. 2017. “Intron Retention Is Regulated by Altered MeCP2-Mediated Splicing Factor Recruitment.” *Nature Communications* 8 (1): 15134. https://doi.org/10.1038/ncomms15134.

Wong, Justin J.-L., William Ritchie, Olivia A. Ebner, Matthias Selbach, Jason W. H. Wong, Yizhou Huang, Dadi Gao, et al. 2013. “Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation.” *Cell* 154 (3): 583–95. https://doi.org/10.1016/j.cell.2013.06.052.

Xiang, Kun, Zhiwei Xu, Yu-Qian Hu, Yi-Sheng He, Guo-Cui Wu, Tian-Yu Li, Xue-Rong Wang, et al. 2021. “Circadian Clock Genes as Promising Therapeutic Targets for Autoimmune Diseases.” *Autoimmunity Reviews* 20 (8): 102866. https://doi.org/10.1016/j.autrev.2021.102866.

Xiao, Tsan. 2009. “Innate Immune Recognition of Nucleic Acids.” *Immunologic Research* 43 (1–3): 98–108. https://doi.org/10.1007/s12026-008-8053-x.

Xu, Jiayue, Dongsheng Wang, Hongliu Ma, Xueying Zhai, Yue Huo, Yue Ren, Weiqian Li, et al. 2022. “KHSRP Combines Transcriptional and Posttranscriptional Mechanisms to Regulate Monocytic Differentiation.” *Blood Science* 4 (3): 103–15. https://doi.org/10.1097/BS9.0000000000000122.

Y, Ge, and Porse Bt. 2014. “The Functional Consequences of Intron Retention: Alternative Splicing Coupled to NMD as a Regulator of Gene Expression.” *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* 36 (3). https://doi.org/10.1002/bies.201300156.

Yadegari, Hamideh, Arijit Biswas, Mohammad Suhail Akhter, Julia Driesen, Vytautas Ivaskevicius, Natascha Marquardt, and Johannes Oldenburg. 2016. “Intron

Retention Resulting from a Silent Mutation in the VWF Gene That Structurally Influences the 5′ Splice Site." *Blood* 128 (17): 2144–52. https://doi.org/10.1182/blood-2016-02-699686.
Yan, Nan. 2017. "Immune Diseases Associated with TREX1 and STING Dysfunction." *Journal of Interferon & Cytokine Research: The Official Journal of the International Society for Interferon and Cytokine Research* 37 (5): 198–206. https://doi.org/10.1089/jir.2016.0086.
Yang, Yun-Gui, Tomas Lindahl, and Deborah E. Barnes. 2007. "Trex1 Exonuclease Degrades SsDNA to Prevent Chronic Checkpoint Activation and Autoimmune Disease." *Cell* 131 (5): 873–86. https://doi.org/10.1016/j.cell.2007.10.017.
Yao, Jun, Dong Ding, Xueping Li, Ting Shen, Haihui Fu, Hua Zhong, Gang Wei, and Ting Ni. 2020. "Prevalent Intron Retention Fine-Tunes Gene Expression and Contributes to Cellular Senescence." *Aging Cell* 19 (12): e13276. https://doi.org/10.1111/acel.13276.
Yap, Karen, Zhao Qin Lim, Piyush Khandelia, Brad Friedman, and Eugene V. Makeyev. 2012. "Coordinated Regulation of Neuronal MRNA Steady-State Levels through Developmentally Controlled Intron Retention." *Genes & Development* 26 (11): 1209–23. https://doi.org/10.1101/gad.188037.112.
Yeo, Gene, and Christopher B. Burge. 2004. "Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals." *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 11 (2–3): 377–94. https://doi.org/10.1089/1066527041410418.
Yu, Le, and Pengda Liu. 2021. "Cytosolic DNA Sensing by CGAS: Regulation, Function, and Human Diseases." *Signal Transduction and Targeted Therapy* 6 (1): 1–15. https://doi.org/10.1038/s41392-021-00554-y.
Zhang, Fangming, and Reuven Rabinovici. 2007. "Adenosine Deaminase Acting on RNA 1 Accelerates Cell Cycle through Increased Translation and Activity of Cyclin-Dependent Kinase 2." *Shock (Augusta, Ga.)* 27 (2): 214–19. https://doi.org/10.1097/01.shk.0000238072.51653.f2.
Zhang, Qu, Hua Li, Hong Jin, Huibiao Tan, Jun Zhang, and Sitong Sheng. 2014. "The Global Landscape of Intron Retentions in Lung Adenocarcinoma." *BMC Medical Genomics* 7 (March): 15. https://doi.org/10.1186/1755-8794-7-15.
Zhang, Xiao, Rui Zhang, and Jinpu Yu. 2020. "New Understanding of the Relevant Role of LINE-1 Retrotransposition in Human Disease and Immune Modulation." *Frontiers in Cell and Developmental Biology* 8. https://www.frontiersin.org/articles/10.3389/fcell.2020.00657.
Zhang, Yanbin, Fenghua Yuan, Steven R. Presnell, Keli Tian, Yin Gao, Alan E. Tomkinson, Liya Gu, and Guo-Min Li. 2005. "Reconstitution of 5'-Directed Human Mismatch Repair in a Purified System." *Cell* 122 (5): 693–705. https://doi.org/10.1016/j.cell.2005.06.027.
Zhao, Ke, Juan Du, Xue Han, John L. Goodier, Peng Li, Xiaohong Zhou, Wei Wei, et al. 2013. "Modulation of LINE-1 and Alu/SVA Retrotransposition by Aicardi-Goutières Syndrome-Related SAMHD1." *Cell Reports* 4 (6): 1108–15. https://doi.org/10.1016/j.celrep.2013.08.019.
Zheng, Jian-Tao, Cui-Xiang Lin, Zhao-Yu Fang, and Hong-Dong Li. 2020. "Intron Retention as a Mode for RNA-Seq Data Analysis." *Frontiers in Genetics* 11. https://doi.org/10.3389/fgene.2020.00586.
Zhu, Li-Yuan, Yi-Ran Zhu, Dong-Jun Dai, Xian Wang, and Hong-Chuan Jin. 2018. "Epigenetic Regulation of Alternative Splicing." *American Journal of Cancer Research* 8 (12): 2346–58.

Zhu, Yong, Xiuye Wang, Elmira Forouzmand, Joshua Jeong, Feng Qiao, Gregory A. Sowd, Alan N. Engelman, Xiaohui Xie, Klemens J. Hertel, and Yongsheng Shi. 2018. “Molecular Mechanisms for CFIm-Mediated Regulation of MRNA Alternative Polyadenylation.” *Molecular Cell* 69 (1): 62-74.e4. https://doi.org/10.1016/j.molcel.2017.11.031.

Ziff, Oliver J, Doaa M Taha, Hamish Crerar, Benjamin E Clarke, Anob M Chakrabarti, Gavin Kelly, Jacob Neeves, Giulia E Tyzack, Nicholas M Luscombe, and Rickie Patani. 2021. “Reactive Astrocytes in ALS Display Diminished Intron Retention.” *Nucleic Acids Research* 49 (6): 3168–84. https://doi.org/10.1093/nar/gkab115.

Appendices

Table 1: Summary of genes used for interferon score calculation

Symbol	Ensembl ID	Description
IRF7	ENSG00000276561	interferon regulatory factor 7
IFI27	ENSG00000275214	interferon alpha inducible protein 27
EPSTI1	ENSG00000133106	epithelial stromal interaction 1
MX1	ENSG00000157601	MX dynamin like GTPase 1
SIGLEC1	ENSG00000088827	sialic acid binding Ig like lectin 1
USP18	ENSG00000184979	ubiquitin specific peptidase 18
LAMP3	ENSG00000078081	lysosomal associated membrane protein 3
RTP4	ENSG00000136514	receptor transporter protein 4
IFIT1	ENSG00000185745	interferon induced protein with tetratricopeptide repeats 1
IFIT3	ENSG00000119917	interferon induced protein with tetratricopeptide repeats 3
IFIT2	ENSG00000119922	interferon induced protein with tetratricopeptide repeats 2
SPATS2L	ENSG00000196141	spermatogenesis associated serine rich 2 like
HERC5	ENSG00000138646	HECT and RLD domain containing E3 ubiquitin protein ligase 5
IFI27	ENSG00000165949	interferon alpha inducible protein 27
PLSCR1	ENSG00000188313	phospholipid scramblase 1
SERPING1	ENSG00000149131	serpin family G member 1
CXCL10	ENSG00000169245	C-X-C motif chemokine ligand 10
SOCS1	ENSG00000185338	suppressor of cytokine signaling 1
IFITM1	ENSG00000185885	interferon induced transmembrane protein 1
HERC6	ENSG00000138642	HECT and RLD domain containing E3 ubiquitin protein ligase family member 6
IFIT5	ENSG00000152778	interferon induced protein with tetratricopeptide repeats 5
RSAD2	ENSG00000134321	radical S-adenosyl methionine domain containing 2
OAS1	ENSG00000089127	2'-5'-oligoadenylate synthetase 1
OAS3	ENSG00000111331	2'-5'-oligoadenylate synthetase 3
OAS2	ENSG00000111335	2'-5'-oligoadenylate synthetase 2
GBP1	ENSG00000117228	guanylate binding protein 1
IFI6	ENSG00000126709	interferon alpha inducible protein 6
IFI44L	ENSG00000137959	interferon induced protein 44 like
ISG15	ENSG00000187608	ISG15 ubiquitin like modifier
IRF7	ENSG00000185507	interferon regulatory factor 7
OASL	ENSG00000135114	2'-5'-oligoadenylate synthetase like
STAT1	ENSG00000115415	signal transducer and activator of transcription 1
C1QA	ENSG00000173372	complement C1q A chain

Table 2: Summary of genes used for proliferation score calculation

Symbol	Ensembl ID	Description
DLG5	ENSG00000274429	discs large MAGUK scaffold protein 5
TRIM27	ENSG00000234495	tripartite motif containing 27
LST1	ENSG00000235915	leukocyte specific transcript 1
BTG3	ENSG00000281484	BTG anti-proliferation factor 3
LST1	ENSG00000230791	leukocyte specific transcript 1
CDK7	ENSG00000277273	cyclin dependent kinase 7
MIF	ENSG00000276701	macrophage migration inhibitory factor
PTPRC	ENSG00000262418	protein tyrosine phosphatase receptor type C
SERPINF1	ENSG00000282307	serpin family F member 1
CDKN1C	ENSG00000273707	cyclin dependent kinase inhibitor 1C
CHRNA7	ENSG00000274542	cholinergic receptor nicotinic alpha 7 subunit
RBFOX2	ENSG00000277564	RNA binding fox-1 homolog 2
TCF19	ENSG00000224472	transcription factor 19
TRIM27	ENSG00000229006	tripartite motif containing 27
TCF19	ENSG00000206455	transcription factor 19
TCF19	ENSG00000224941	transcription factor 19
PRTN3	ENSG00000277804	proteinase 3
TRIM27	ENSG00000215641	tripartite motif containing 27
ELANE	ENSG00000277571	elastase, neutrophil expressed
AIF1	ENSG00000235588	allograft inflammatory factor 1
CCL23	ENSG00000276114	C-C motif chemokine ligand 23
LST1	ENSG00000206433	leukocyte specific transcript 1
TSPY1	ENSG00000258992	testis specific protein Y-linked 1
AIF1	ENSG00000206428	allograft inflammatory factor 1
AIF1	ENSG00000234836	allograft inflammatory factor 1
KIF15	ENSG00000280610	kinesin family member 15
TRIM27	ENSG00000233948	tripartite motif containing 27
TCF19	ENSG00000233890	transcription factor 19
CCL14	ENSG00000277236	C-C motif chemokine ligand 14
AIF1	ENSG00000235985	allograft inflammatory factor 1
LST1	ENSG00000231048	leukocyte specific transcript 1
CHRNA7	ENSG00000282088	cholinergic receptor nicotinic alpha 7 subunit
PTEN	ENSG00000284792	phosphatase and tensin homolog
LST1	ENSG00000234514	leukocyte specific transcript 1
TRIM27	ENSG00000237071	tripartite motif containing 27
FLT1	ENSG00000102755	fms related receptor tyrosine kinase 1
AIF1	ENSG00000237727	allograft inflammatory factor 1
TCF19	ENSG00000234674	transcription factor 19
ADRA1D	ENSG00000171873	adrenoceptor alpha 1D
MAPRE2	ENSG00000166974	microtubule associated protein RP/EB family member 2
SIRPG	ENSG00000089012	signal regulatory protein gamma
TRIM27	ENSG00000237462	tripartite motif containing 27
OSM	ENSG00000099985	oncostatin M
LST1	ENSG00000223465	leukocyte specific transcript 1
ATP5IF1	ENSG00000285390	ATP synthase inhibitory factor subunit 1
S100B	ENSG00000160307	S100 calcium binding protein B
LST1	ENSG00000226182	leukocyte specific transcript 1
LIF	ENSG00000128342	LIF interleukin 6 family cytokine

TCF19	ENSG00000224379	transcription factor 19
BTG3	ENSG00000154640	BTG anti-proliferation factor 3
MYO16	ENSG00000041515	myosin XVI
BCAR1	ENSG00000285460	BCAR1 scaffold protein, Cas family member
FLT3	ENSG00000122025	fms related receptor tyrosine kinase 3
COL18A1	ENSG00000182871	collagen type XVIII alpha 1 chain
ADAMTS1	ENSG00000154734	ADAM metallopeptidase with thrombospondin type 1 motif 1
MNAT1	ENSG00000020426	MNAT1 component of CDK activating kinase
NDN	ENSG00000288364	necdin, MAGE family member
MCTS1	ENSG00000288295	MCTS1 re-initiation and release factor
GAS6	ENSG00000183087	growth arrest specific 6
CDC123	ENSG00000151465	cell division cycle 123
TSPO	ENSG00000100300	translocator protein
NDP	ENSG00000124479	norrin cystine knot growth factor NDP
EEF1E1	ENSG00000124802	eukaryotic translation elongation factor 1 epsilon 1
SSTR3	ENSG00000278195	somatostatin receptor 3
ING1	ENSG00000153487	inhibitor of growth family member 1
SSTR1	ENSG00000139874	somatostatin receptor 1
TRIM27	ENSG00000204713	tripartite motif containing 27
EDN1	ENSG00000078401	endothelin 1
SSR1	ENSG00000124783	signal sequence receptor subunit 1
EIF5A	ENSG00000288145	eukaryotic translation initiation factor 5A
PDS5B	ENSG00000083642	PDS5 cohesin associated factor B
ADRA1A	ENSG00000120907	adrenoceptor alpha 1A
NRP1	ENSG00000099250	neuropilin 1
TNFSF13B	ENSG00000102524	TNF superfamily member 13b
ATP8A2	ENSG00000132932	ATPase phospholipid transporting 8A2
GNRH1	ENSG00000147437	gonadotropin releasing hormone 1
LYN	ENSG00000254087	LYN proto-oncogene, Src family tyrosine kinase
MAPRE1	ENSG00000101367	microtubule associated protein RP/EB family member 1
LDOC1	ENSG00000182195	LDOC1 regulator of NFKB signaling
MTCP1	ENSG00000214827	mature T cell proliferation 1
TOB2	ENSG00000183864	transducer of ERBB2, 2
ERG	ENSG00000157554	ETS transcription factor ERG
FGF10	ENSG00000070193	fibroblast growth factor 10
TPX2	ENSG00000088325	TPX2 microtubule nucleation factor
TFDP1	ENSG00000198176	transcription factor Dp-1
PRMT5	ENSG00000100462	protein arginine methyltransferase 5
POU3F2	ENSG00000184486	POU class 3 homeobox 2
E2F1	ENSG00000101412	E2F transcription factor 1
PCNA	ENSG00000132646	proliferating cell nuclear antigen
TIMP1	ENSG00000102265	TIMP metallopeptidase inhibitor 1
CD24	ENSG00000272398	CD24 molecule
PEMT	ENSG00000133027	phosphatidylethanolamine N-methyltransferase
OSMR	ENSG00000145623	oncostatin M receptor
MATK	ENSG00000007264	megakaryocyte-associated tyrosine kinase
BRCA2	ENSG00000139618	BRCA2 DNA repair associated
ICOSLG	ENSG00000160223	inducible T cell costimulator ligand
PGF	ENSG00000119630	placental growth factor
MAGED1	ENSG00000179222	MAGE family member D1
CDC16	ENSG00000130177	cell division cycle 16
UHRF2	ENSG00000147854	ubiquitin like with PHD and ring finger domains 2

CSE1L	ENSG00000124207	chromosome segregation 1 like
NDUFS4	ENSG00000164258	NADH:ubiquinone oxidoreductase subunit S4
KAT2B	ENSG00000114166	lysine acetyltransferase 2B
TCFL5	ENSG00000101190	transcription factor like 5
VEGFD	ENSG00000165197	vascular endothelial growth factor D
UBR5	ENSG00000104517	ubiquitin protein ligase E3 component n-recognin 5
APC	ENSG00000134982	APC regulator of WNT signaling pathway
CUL4A	ENSG00000139842	cullin 4A
FABP7	ENSG00000164434	fatty acid binding protein 7
CDC6	ENSG00000094804	cell division cycle 6
NAA35	ENSG00000135040	N-alpha-acetyltransferase 35, NatC auxiliary subunit
ANG	ENSG00000214274	angiogenin
FOXO4	ENSG00000184481	forkhead box O4
MKI67	ENSG00000148773	marker of proliferation Ki-67
NF2	ENSG00000186575	NF2, moesin-ezrin-radixin like (MERLIN) tumor suppressor
TXN	ENSG00000136810	thioredoxin
IL9	ENSG00000145839	interleukin 9
TGFBI	ENSG00000120708	transforming growth factor beta induced
PTK2B	ENSG00000120899	protein tyrosine kinase 2 beta
TUSC2	ENSG00000114383	tumor suppressor 2, mitochondrial calcium regulator
CDC25B	ENSG00000101224	cell division cycle 25B
MIF	ENSG00000240972	macrophage migration inhibitory factor
LAMA1	ENSG00000101680	laminin subunit alpha 1
IGFBP4	ENSG00000141753	insulin like growth factor binding protein 4
BUB3	ENSG00000154473	BUB3 mitotic checkpoint protein
TTK	ENSG00000112742	TTK protein kinase
MYC	ENSG00000136997	MYC proto-oncogene, bHLH transcription factor
OPRM1	ENSG00000112038	opioid receptor mu 1
NR6A1	ENSG00000148200	nuclear receptor subfamily 6 group A member 1
BCAT1	ENSG00000060982	branched chain amino acid transaminase 1
DNAJA2	ENSG00000069345	DnaJ heat shock protein family (Hsp40) member A2
SYK	ENSG00000165025	spleen associated tyrosine kinase
EPS8	ENSG00000151491	epidermal growth factor receptor pathway substrate 8
DELEC1	ENSG00000173077	deleted in esophageal cancer 1
KLF4	ENSG00000136826	Kruppel like factor 4
FRK	ENSG00000111816	fyn related Src family tyrosine kinase
VIP	ENSG00000146469	vasoactive intestinal peptide
GML	ENSG00000104499	glycosylphosphatidylinositol anchored molecule like
CBFA2T2	ENSG00000078699	CBFA2/RUNX1 partner transcriptional co-repressor 2
POLA1	ENSG00000101868	DNA polymerase alpha 1, catalytic subunit
SCIN	ENSG00000006747	scinderin
NPY	ENSG00000122585	neuropeptide Y
NANOG	ENSG00000111704	Nanog homeobox
TBX2	ENSG00000121068	T-box transcription factor 2
LGI1	ENSG00000108231	leucine rich glioma inactivated 1
TOB1	ENSG00000141232	transducer of ERBB2, 1
TRIB1	ENSG00000173334	tribbles pseudokinase 1
IL3	ENSG00000164399	interleukin 3
BCAR1	ENSG00000050820	BCAR1 scaffold protein, Cas family member
TGFBR2	ENSG00000163513	transforming growth factor beta receptor 2
CDK9	ENSG00000136807	cyclin dependent kinase 9
TSHR	ENSG00000165409	thyroid stimulating hormone receptor

MCTS1	ENSG00000232119	MCTS1 re-initiation and release factor
RASGRP4	ENSG00000171777	RAS guanyl releasing protein 4
FTH1	ENSG00000167996	ferritin heavy chain 1
VTI1B	ENSG00000100568	vesicle transport through interaction with t-SNAREs 1B
RACGAP1	ENSG00000161800	Rac GTPase activating protein 1
TNFRSF11A	ENSG00000141655	TNF receptor superfamily member 11a
PROK2	ENSG00000163421	prokineticin 2
APPL1	ENSG00000157500	adaptor protein, phosphotyrosine interacting with PH domain and leucine zipper 1
ATP5IF1	ENSG00000130770	ATP synthase inhibitory factor subunit 1
ADRA1B	ENSG00000170214	adrenoceptor alpha 1B
TBRG4	ENSG00000136270	transforming growth factor beta regulator 4
IGFBP6	ENSG00000167779	insulin like growth factor binding protein 6
KRT2	ENSG00000172867	keratin 2
PAWR	ENSG00000177425	pro-apoptotic WT1 regulator
BTC	ENSG00000174808	betacellulin
RBFOX2	ENSG00000100320	RNA binding fox-1 homolog 2
IFNL1	ENSG00000182393	interferon lambda 1
FABP3	ENSG00000121769	fatty acid binding protein 3
BTG1	ENSG00000133639	BTG anti-proliferation factor 1
EID2	ENSG00000176396	EP300 interacting inhibitor of differentiation 2
TNFSF9	ENSG00000125657	TNF superfamily member 9
CD70	ENSG00000125726	CD70 molecule
NCK1	ENSG00000158092	NCK adaptor protein 1
SPOCK1	ENSG00000152377	SPARC (osteonectin), cwcv and kazal like domains proteoglycan 1
IL4	ENSG00000113520	interleukin 4
SERPINF1	ENSG00000132386	serpin family F member 1
DNPH1	ENSG00000112667	2'-deoxynucleoside 5'-phosphate N-hydrolase 1
RAPGEF3	ENSG00000079337	Rap guanine nucleotide exchange factor 3
TCF19	ENSG00000137310	transcription factor 19
SERTAD1	ENSG00000197019	SERTA domain containing 1
IL27	ENSG00000197272	interleukin 27
NOX4	ENSG00000086991	NADPH oxidase 4
GPC4	ENSG00000076716	glypican 4
CSF1R	ENSG00000182578	colony stimulating factor 1 receptor
EGR4	ENSG00000135625	early growth response 4
REG1A	ENSG00000115386	regenerating family member 1 alpha
IGF1R	ENSG00000140443	insulin like growth factor 1 receptor
PDGFRA	ENSG00000134853	platelet derived growth factor receptor alpha
PLG	ENSG00000122194	plasminogen
CSF3	ENSG00000108342	colony stimulating factor 3
HTR1A	ENSG00000178394	5-hydroxytryptamine receptor 1A
CDKN2C	ENSG00000123080	cyclin dependent kinase inhibitor 2C
REG1B	ENSG00000172023	regenerating family member 1 beta
CUL5	ENSG00000166266	cullin 5
LST1	ENSG00000204482	leukocyte specific transcript 1
AIF1	ENSG00000204472	allograft inflammatory factor 1
ALOX15B	ENSG00000179593	arachidonate 15-lipoxygenase type B
FABP6	ENSG00000170231	fatty acid binding protein 6
LAMP3	ENSG00000078081	lysosomal associated membrane protein 3
IL11	ENSG00000095752	interleukin 11
BUB1B	ENSG00000156970	BUB1 mitotic checkpoint serine/threonine kinase B

MT3	ENSG00000087250	metallothionein 3
TM4SF4	ENSG00000169903	transmembrane 4 L six family member 4
SST	ENSG00000157005	somatostatin
APPL2	ENSG00000136044	adaptor protein, phosphotyrosine interacting with PH domain and leucine zipper 2
TGFA	ENSG00000163235	transforming growth factor alpha
GLP2R	ENSG00000065325	glucagon like peptide 2 receptor
ELF4	ENSG00000102034	E74 like ETS transcription factor 4
S1PR2	ENSG00000267534	sphingosine-1-phosphate receptor 2
IL1B	ENSG00000125538	interleukin 1 beta
TBRG1	ENSG00000154144	transforming growth factor beta regulator 1
CIAO1	ENSG00000144021	cytosolic iron-sulfur assembly component 1
MAS1	ENSG00000130368	MAS1 proto-oncogene, G protein-coupled receptor
KITLG	ENSG00000049130	KIT ligand
CHRNA7	ENSG00000175344	cholinergic receptor nicotinic alpha 7 subunit
TNFSF15	ENSG00000181634	TNF superfamily member 15
CRTAM	ENSG00000109943	cytotoxic and regulatory T cell molecule
SPHK1	ENSG00000176170	sphingosine kinase 1
IL12RB1	ENSG00000096996	interleukin 12 receptor subunit beta 1
NCK2	ENSG00000071051	NCK adaptor protein 2
IL15	ENSG00000164136	interleukin 15
IL9R	ENSG00000124334	interleukin 9 receptor
FGF18	ENSG00000156427	fibroblast growth factor 18
FZD3	ENSG00000104290	frizzled class receptor 3
EIF5A2	ENSG00000163577	eukaryotic translation initiation factor 5A2
EIF5A	ENSG00000132507	eukaryotic translation initiation factor 5A
JAG2	ENSG00000184916	jagged canonical Notch ligand 2
MS4A2	ENSG00000149534	membrane spanning 4-domains A2
USP8	ENSG00000138592	ubiquitin specific peptidase 8
TGFB1I1	ENSG00000140682	transforming growth factor beta 1 induced transcript 1
PIM2	ENSG00000102096	Pim-2 proto-oncogene, serine/threonine kinase
CRIP1	ENSG00000213145	cysteine rich protein 1
GHRL	ENSG00000157017	ghrelin and obestatin prepropeptide
SFTPD	ENSG00000133661	surfactant protein D
ABI1	ENSG00000136754	abl interactor 1
PRKD1	ENSG00000184304	protein kinase D1
KLF10	ENSG00000155090	Kruppel like factor 10
CDKN3	ENSG00000100526	cyclin dependent kinase inhibitor 3
TSC1	ENSG00000165699	TSC complex subunit 1
CHRM5	ENSG00000184984	cholinergic receptor muscarinic 5
FLT3LG	ENSG00000090554	fms related receptor tyrosine kinase 3 ligand
MNT	ENSG00000070444	MAX network transcriptional repressor
GFI1B	ENSG00000165702	growth factor independent 1B transcriptional repressor
PLK1	ENSG00000166851	polo like kinase 1
SSTR4	ENSG00000132671	somatostatin receptor 4
MXI1	ENSG00000119950	MAX interactor 1, dimerization protein
IL2	ENSG00000109471	interleukin 2
NME2	ENSG00000243678	NME/NM23 nucleoside diphosphate kinase 2
LRP1	ENSG00000123384	LDL receptor related protein 1
ERF	ENSG00000105722	ETS2 repressor factor
GTPBP4	ENSG00000107937	GTP binding protein 4
FGFBP1	ENSG00000137440	fibroblast growth factor binding protein 1

CDC14A	ENSG00000079335	cell division cycle 14A
UTP20	ENSG00000120800	UTP20 small subunit processome component
CD40LG	ENSG00000102245	CD40 ligand
CDK10	ENSG00000185324	cyclin dependent kinase 10
NME1	ENSG00000239672	NME/NM23 nucleoside diphosphate kinase 1
IGFBP7	ENSG00000163453	insulin like growth factor binding protein 7
CTF1	ENSG00000150281	cardiotrophin 1
IL18	ENSG00000150782	interleukin 18
TNFSF4	ENSG00000117586	TNF superfamily member 4
CSF1	ENSG00000184371	colony stimulating factor 1
PPARD	ENSG00000112033	peroxisome proliferator activated receptor delta
DKC1	ENSG00000130826	dyskerin pseudouridine synthase 1
LAMC1	ENSG00000135862	laminin subunit gamma 1
CD276	ENSG00000103855	CD276 molecule
NPM1	ENSG00000181163	nucleophosmin 1
SKP2	ENSG00000145604	S-phase kinase associated protein 2
TENM1	ENSG00000009694	teneurin transmembrane protein 1
ELF5	ENSG00000135374	E74 like ETS transcription factor 5
CNOT8	ENSG00000155508	CCR4-NOT transcription complex subunit 8
ADGRB1	ENSG00000181790	adhesion G protein-coupled receptor B1
GAS8	ENSG00000141013	growth arrest specific 8
DTYMK	ENSG00000168393	deoxythymidylate kinase
DLEC1	ENSG00000008226	DLEC1 cilia and flagella associated protein
ZEB1	ENSG00000148516	zinc finger E-box binding homeobox 1
TBX3	ENSG00000135111	T-box transcription factor 3
NDN	ENSG00000182636	necdin, MAGE family member
INSIG1	ENSG00000186480	insulin induced gene 1
HGS	ENSG00000185359	hepatocyte growth factor-regulated tyrosine kinase substrate
TDGF1	ENSG00000241186	teratocarcinoma-derived growth factor 1
PDF	ENSG00000258429	peptide deformylase, mitochondrial
FOSL1	ENSG00000175592	FOS like 1, AP-1 transcription factor subunit
CDKN2D	ENSG00000129355	cyclin dependent kinase inhibitor 2D
TBX5	ENSG00000089225	T-box transcription factor 5
GLI1	ENSG00000111087	GLI family zinc finger 1
SLC29A2	ENSG00000174669	solute carrier family 29 member 2
CDC25C	ENSG00000158402	cell division cycle 25C
FGF7	ENSG00000140285	fibroblast growth factor 7
PTCH1	ENSG00000185920	patched 1
RBBP7	ENSG00000102054	RB binding protein 7, chromatin remodeling factor
TSPAN31	ENSG00000135452	tetraspanin 31
CDK4	ENSG00000135446	cyclin dependent kinase 4
CDK7	ENSG00000134058	cyclin dependent kinase 7
FES	ENSG00000182511	FES proto-oncogene, tyrosine kinase
CKLF	ENSG00000217555	chemokine like factor
TGFB2	ENSG00000092969	transforming growth factor beta 2
ISG20	ENSG00000172183	interferon stimulated exonuclease gene 20
CDK6	ENSG00000105810	cyclin dependent kinase 6
GPNMB	ENSG00000136235	glycoprotein nmb
THPO	ENSG00000090534	thrombopoietin
ERBB2	ENSG00000141736	erb-b2 receptor tyrosine kinase 2
UBE2V2	ENSG00000169139	ubiquitin conjugating enzyme E2 V2
FOXP3	ENSG00000049768	forkhead box P3

IL21	ENSG00000138684	interleukin 21
RPS4X	ENSG00000198034	ribosomal protein S4 X-linked
CDK5	ENSG00000164885	cyclin dependent kinase 5
LY86	ENSG00000112799	lymphocyte antigen 86
TP53I11	ENSG00000175274	tumor protein p53 inducible protein 11
EHF	ENSG00000135373	ETS homologous factor
SRA1	ENSG00000213523	steroid receptor RNA activator 1
BST2	ENSG00000130303	bone marrow stromal cell antigen 2
BNC1	ENSG00000169594	basonuclin 1
PLAAT4	ENSG00000133321	phospholipase A and acyltransferase 4
PRL	ENSG00000172179	prolactin
UMOD	ENSG00000169344	uromodulin
ADRA2A	ENSG00000150594	adrenoceptor alpha 2A
BAP1	ENSG00000163930	BRCA1 associated protein 1
CDC7	ENSG00000097046	cell division cycle 7
CAPNS1	ENSG00000126247	calpain small subunit 1
ING5	ENSG00000168395	inhibitor of growth family member 5
FSCN1	ENSG00000075618	fascin actin-bundling protein 1
NAP1L1	ENSG00000187109	nucleosome assembly protein 1 like 1
CGRRF1	ENSG00000100532	cell growth regulator with ring finger domain 1
CUL2	ENSG00000108094	cullin 2
EMP2	ENSG00000213853	epithelial membrane protein 2
BMI1	ENSG00000168283	BMI1 proto-oncogene, polycomb ring finger
ING4	ENSG00000111653	inhibitor of growth family member 4
LAMB1	ENSG00000091136	laminin subunit beta 1
TIPIN	ENSG00000075131	TIMELESS interacting protein
CDKN2A	ENSG00000147889	cyclin dependent kinase inhibitor 2A
GLI2	ENSG00000074047	GLI family zinc finger 2
NAB2	ENSG00000166886	NGFI-A binding protein 2
CDKN2B	ENSG00000147883	cyclin dependent kinase inhibitor 2B
CD274	ENSG00000120217	CD274 molecule
S1PR3	ENSG00000213694	sphingosine-1-phosphate receptor 3
IL15RA	ENSG00000134470	interleukin 15 receptor subunit alpha
CKS2	ENSG00000123975	CDC28 protein kinase regulatory subunit 2
TGFBR1	ENSG00000106799	transforming growth factor beta receptor 1
AIMP1	ENSG00000164022	aminoacyl tRNA synthetase complex interacting multifunctional protein 1
EPGN	ENSG00000182585	epithelial mitogen
CHRM3	ENSG00000133019	cholinergic receptor muscarinic 3
CADM1	ENSG00000182985	cell adhesion molecule 1
CXCL10	ENSG00000169245	C-X-C motif chemokine ligand 10
VEGFA	ENSG00000112715	vascular endothelial growth factor A
CD74	ENSG00000019582	CD74 molecule
FGF5	ENSG00000138675	fibroblast growth factor 5
VEGFB	ENSG00000173511	vascular endothelial growth factor B
WARS1	ENSG00000140105	tryptophanyl-tRNA synthetase 1
PTHLH	ENSG00000087494	parathyroid hormone like hormone
EMP1	ENSG00000134531	epithelial membrane protein 1
PRG4	ENSG00000116690	proteoglycan 4
DAB2	ENSG00000153071	DAB adaptor protein 2
CHRNA10	ENSG00000129749	cholinergic receptor nicotinic alpha 10 subunit
RHOG	ENSG00000177105	ras homolog family member G

IL2RA	ENSG00000134460	interleukin 2 receptor subunit alpha
CD33	ENSG00000105383	CD33 molecule
IL12B	ENSG00000113302	interleukin 12B
INSL4	ENSG00000120211	insulin like 4
IFITM1	ENSG00000185885	interferon induced transmembrane protein 1
ALOX12	ENSG00000108839	arachidonate 12-lipoxygenase, 12S type
SESN1	ENSG00000080546	sestrin 1
B4GALT7	ENSG00000027847	beta-1,4-galactosyltransferase 7
DLG5	ENSG00000151208	discs large MAGUK scaffold protein 5
CDKN1A	ENSG00000124762	cyclin dependent kinase inhibitor 1A
PIM1	ENSG00000137193	Pim-1 proto-oncogene, serine/threonine kinase
HELLS	ENSG00000119969	helicase, lymphoid specific
IL5RA	ENSG00000091181	interleukin 5 receptor subunit alpha
IL7	ENSG00000104432	interleukin 7
GAB1	ENSG00000109458	GRB2 associated binding protein 1
CXCL1	ENSG00000163739	C-X-C motif chemokine ligand 1
CXCL5	ENSG00000163735	C-X-C motif chemokine ligand 5
ACVRL1	ENSG00000139567	activin A receptor like type 1
DUSP22	ENSG00000112679	dual specificity phosphatase 22
LRPAP1	ENSG00000163956	LDL receptor related protein associated protein 1
EPHB4	ENSG00000196411	EPH receptor B4
PTEN	ENSG00000171862	phosphatase and tensin homolog
GCG	ENSG00000115263	glucagon
VIPR1	ENSG00000114812	vasoactive intestinal peptide receptor 1
SSTR5	ENSG00000162009	somatostatin receptor 5
TNFRSF8	ENSG00000120949	TNF receptor superfamily member 8
RBBP4	ENSG00000162521	RB binding protein 4, chromatin remodeling factor
PRKRA	ENSG00000180228	protein activator of interferon induced protein kinase EIF2AK2
TNFSF13	ENSG00000161955	TNF superfamily member 13
CXCL8	ENSG00000169429	C-X-C motif chemokine ligand 8
DHPS	ENSG00000095059	deoxyhypusine synthase
AGGF1	ENSG00000164252	angiogenic factor with G-patch and FHA domains 1
IRF2	ENSG00000168310	interferon regulatory factor 2
CD164	ENSG00000135535	CD164 molecule
MFN2	ENSG00000116688	mitofusin 2
KLF11	ENSG00000172059	Kruppel like factor 11
NOP2	ENSG00000111641	NOP2 nucleolar protein
CHRM1	ENSG00000168539	cholinergic receptor muscarinic 1
SPEG	ENSG00000072195	striated muscle enriched protein kinase
ACHE	ENSG00000087085	acetylcholinesterase (Cartwright blood group)
MARK4	ENSG00000007047	microtubule affinity regulating kinase 4
INHA	ENSG00000123999	inhibin subunit alpha
EREG	ENSG00000124882	epiregulin
AREG	ENSG00000109321	amphiregulin
CLEC11A	ENSG00000105472	C-type lectin domain containing 11A
NAMPT	ENSG00000105835	nicotinamide phosphoribosyltransferase
TNFSF8	ENSG00000106952	TNF superfamily member 8
IL31RA	ENSG00000164509	interleukin 31 receptor A
KIF2C	ENSG00000142945	kinesin family member 2C
MXD4	ENSG00000123933	MAX dimerization protein 4
KRT16	ENSG00000186832	keratin 16
ENPP7	ENSG00000182156	ectonucleotide pyrophosphatase/phosphodiesterase 7

ENPEP	ENSG00000138792	glutamyl aminopeptidase
CCN1	ENSG00000142871	cellular communication network factor 1
CDH13	ENSG00000140945	cadherin 13
EBI3	ENSG00000105246	Epstein-Barr virus induced 3
CDC25A	ENSG00000164045	cell division cycle 25A
TBC1D8	ENSG00000204634	TBC1 domain family member 8
DCTN2	ENSG00000175203	dynactin subunit 2
RPS27	ENSG00000177954	ribosomal protein S27
PPM1D	ENSG00000170836	protein phosphatase, Mg2+/Mn2+ dependent 1D
CBFA2T3	ENSG00000129993	CBFA2/RUNX1 partner transcriptional co-repressor 3
CTBP2	ENSG00000175029	C-terminal binding protein 2
BUB1	ENSG00000169679	BUB1 mitotic checkpoint serine/threonine kinase
CEP43	ENSG00000213066	centrosomal protein 43
THAP12	ENSG00000137492	THAP domain containing 12
KLK8	ENSG00000129455	kallikrein related peptidase 8
SSTR2	ENSG00000180616	somatostatin receptor 2
PRTN3	ENSG00000196415	proteinase 3
BMP10	ENSG00000163217	bone morphogenetic protein 10
ELANE	ENSG00000197561	elastase, neutrophil expressed
REG3A	ENSG00000172016	regenerating family member 3 alpha
FLT4	ENSG00000037280	fms related receptor tyrosine kinase 4
EIF2AK2	ENSG00000055332	eukaryotic translation initiation factor 2 alpha kinase 2
PYY	ENSG00000131096	peptide YY
POU1F1	ENSG00000064835	POU class 1 homeobox 1
ARTN	ENSG00000117407	artemin
TNFRSF17	ENSG00000048462	TNF receptor superfamily member 17
IFRD2	ENSG00000214706	interferon related developmental regulator 2
CDK3	ENSG00000250506	cyclin dependent kinase 3
UHRF1	ENSG00000276043	ubiquitin like with PHD and ring finger domains 1
CDKN1C	ENSG00000129757	cyclin dependent kinase inhibitor 1C
CTBP1	ENSG00000159692	C-terminal binding protein 1
CD79A	ENSG00000105369	CD79a molecule
PA2G4	ENSG00000170515	proliferation-associated 2G4
ARHGEF1	ENSG00000076928	Rho guanine nucleotide exchange factor 1
NF1	ENSG00000196712	neurofibromin 1
CHEK1	ENSG00000149554	checkpoint kinase 1
CHRM4	ENSG00000180720	cholinergic receptor muscarinic 4
MIA	ENSG00000261857	MIA SH3 domain containing
CUL1	ENSG00000055130	cullin 1
EMP3	ENSG00000142227	epithelial membrane protein 3
RARRES1	ENSG00000118849	retinoic acid receptor responder 1
E4F1	ENSG00000167967	E4F transcription factor 1
CDC27	ENSG00000004897	cell division cycle 27
PRDX1	ENSG00000117450	peroxiredoxin 1
NASP	ENSG00000132780	nuclear autoantigenic sperm protein
NUP62	ENSG00000213024	nucleoporin 62
CD3E	ENSG00000198851	CD3 epsilon subunit of T-cell receptor complex
HCLS1	ENSG00000180353	hematopoietic cell-specific Lyn substrate 1
MAP3K11	ENSG00000173327	mitogen-activated protein kinase kinase kinase 11
SIPA1	ENSG00000213445	signal-induced proliferation-associated 1
ELN	ENSG00000049540	elastin
CDK2	ENSG00000123374	cyclin dependent kinase 2

TP53	ENSG00000141510	tumor protein p53
MXD1	ENSG00000059728	MAX dimerization protein 1
STAMBP	ENSG00000124356	STAM binding protein
FGF6	ENSG00000111241	fibroblast growth factor 6
CDK11B	ENSG00000248333	cyclin dependent kinase 11B
PRDM4	ENSG00000110851	PR/SET domain 4
COL4A3	ENSG00000169031	collagen type IV alpha 3 chain
TNFSF12	ENSG00000239697	TNF superfamily member 12
BRCA1	ENSG00000012048	BRCA1 DNA repair associated
KRT4	ENSG00000170477	keratin 4
PDAP1	ENSG00000106244	PDGFA associated protein 1
TRAIP	ENSG00000183763	TRAF interacting protein
MST1R	ENSG00000164078	macrophage stimulating 1 receptor
IL6	ENSG00000136244	interleukin 6
MPL	ENSG00000117400	MPL proto-oncogene, thrombopoietin receptor
CD86	ENSG00000114013	CD86 molecule
BTG4	ENSG00000137707	BTG anti-proliferation factor 4
HDGF	ENSG00000143321	heparin binding growth factor
PTN	ENSG00000105894	pleiotrophin
KHDRBS1	ENSG00000121774	KH RNA binding domain containing, signal transduction associated 1
MDM2	ENSG00000135679	MDM2 proto-oncogene
FGF4	ENSG00000075388	fibroblast growth factor 4
PDGFA	ENSG00000197461	platelet derived growth factor subunit A
BMPR2	ENSG00000204217	bone morphogenetic protein receptor type 2
EGFR	ENSG00000146648	epidermal growth factor receptor
DLG1	ENSG00000075711	discs large MAGUK scaffold protein 1
LMO1	ENSG00000166407	LIM domain only 1
CLK1	ENSG00000013441	CDC like kinase 1
CDK13	ENSG00000065883	cyclin dependent kinase 13
VHL	ENSG00000134086	von Hippel-Lindau tumor suppressor
PSPHP1	ENSG00000226278	phosphoserine phosphatase pseudogene 1
CCDC88A	ENSG00000115355	coiled-coil domain containing 88A
DERL2	ENSG00000072849	derlin 2
CCL14	ENSG00000276409	C-C motif chemokine ligand 14
CCL23	ENSG00000274736	C-C motif chemokine ligand 23
CCL3L3	ENSG00000276085	C-C motif chemokine ligand 3 like 3
CUL3	ENSG00000036257	cullin 3
CAPN1	ENSG00000014216	calpain 1
STIL	ENSG00000123473	STIL centriolar assembly protein
TIMELESS	ENSG00000111602	timeless circadian regulator
GLMN	ENSG00000174842	glomulin, FKBP associated protein
HOXC10	ENSG00000180818	homeobox C10
BNIPL	ENSG00000163141	BCL2 interacting protein like
RAF1	ENSG00000132155	Raf-1 proto-oncogene, serine/threonine kinase
LRP5	ENSG00000162337	LDL receptor related protein 5
CGREF1	ENSG00000138028	cell growth regulator with EF-hand domain 1
CDK5R1	ENSG00000176749	cyclin dependent kinase 5 regulatory subunit 1
ADAMTS8	ENSG00000134917	ADAM metallopeptidase with thrombospondin type 1 motif 8
TCIRG1	ENSG00000110719	T cell immune regulator 1, ATPase H+ transporting V0 subunit a3
CD81	ENSG00000110651	CD81 molecule
GNL3	ENSG00000163938	G protein nucleolar 3
IGF1	ENSG00000017427	insulin like growth factor 1

RERG	ENSG00000134533	RAS like estrogen regulated growth inhibitor
EVI5	ENSG00000067208	ecotropic viral integration site 5
DDX11	ENSG00000013573	DEAD/H-box helicase 11
GPX1	ENSG00000233276	glutathione peroxidase 1
CKS1B	ENSG00000173207	CDC28 protein kinase regulatory subunit 1B
CCKBR	ENSG00000110148	cholecystokinin B receptor
S100A11	ENSG00000163191	S100 calcium binding protein A11
CALCA	ENSG00000110680	calcitonin related polypeptide alpha
SFN	ENSG00000175793	stratifin
NUDC	ENSG00000090273	nuclear distribution C, dynein complex regulator
ETS1	ENSG00000134954	ETS proto-oncogene 1, transcription factor
KIF15	ENSG00000163808	kinesin family member 15
MDM4	ENSG00000198625	MDM4 regulator of p53
CD28	ENSG00000178562	CD28 molecule
SLAMF1	ENSG00000117090	signaling lymphocytic activation molecule family member 1
NRDC	ENSG00000078618	nardilysin convertase
NOTCH2	ENSG00000134250	notch receptor 2
CD47	ENSG00000196776	CD47 molecule
CDCA7	ENSG00000144354	cell division cycle associated 7
CXCR2	ENSG00000180871	C-X-C motif chemokine receptor 2
IL1A	ENSG00000115008	interleukin 1 alpha
SPDYA	ENSG00000163806	speedy/RINGO cell cycle regulator family member A
BLZF1	ENSG00000117475	basic leucine zipper nuclear factor 1
TAL1	ENSG00000162367	TAL bHLH transcription factor 1, erythroid differentiation factor
SCG2	ENSG00000171951	secretogranin II
ARHGEF2	ENSG00000116584	Rho/Rac guanine nucleotide exchange factor 2
IL12RB2	ENSG00000081985	interleukin 12 receptor subunit beta 2
BIN1	ENSG00000136717	bridging integrator 1
TACSTD2	ENSG00000184292	tumor associated calcium signal transducer 2
ERBB4	ENSG00000178568	erb-b2 receptor tyrosine kinase 4
PTPRC	ENSG00000081237	protein tyrosine phosphatase receptor type C
IL6R	ENSG00000160712	interleukin 6 receptor
CTNNBIP1	ENSG00000178585	catenin beta interacting protein 1
PDZK1	ENSG00000174827	PDZ domain containing 1
CD160	ENSG00000117281	CD160 molecule
TNFRSF9	ENSG00000049249	TNF receptor superfamily member 9
CREG1	ENSG00000143162	cellular repressor of E1A stimulated genes 1
QSOX1	ENSG00000116260	quiescin sulfhydryl oxidase 1
RUNX3	ENSG00000020633	RUNX family transcription factor 3
EPS15	ENSG00000085832	epidermal growth factor receptor pathway substrate 15
BTG2	ENSG00000159388	BTG anti-proliferation factor 2