Scuola Internazionale Superiore di Studi Avanzati

PhD Course in Functional and Structural Genomics

LINE-1 copy number variation in Alzheimer's disease

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

Candidate
Marta Maurutto

Supervisor
Prof. Stefano Gustincich

Co-Supervisor
Dr. Elena Agostoni

Academic Year 2013/14

This page intentionally left blank

Table of contents

This page intentionally left blank

Abstract

Transposable Elements (TEs) are a class of repetitive DNA sequences able to mobilize and change location in the genome. They make up almost 50% of the human genome and slightly less in the mouse (Lander et al., 2001; Waterston et al., 2002). The autonomous non-LTR retrotransposon Long Interspersed Nuclear Element-1 (L1) is for sure the most impactful and still active TE in the human and mouse genome (Richardson et al., 2014). Recently, it was shown that L1s are active in the mouse, rat and human neural progenitor cells (NPCs), and are more abundant in the genome of cells of the human hippocampus, than in other non-nervous tissues (Thomas and Muotri, 2012; Coufal et al., 2009; Reilly et al., 2013).

Besides human diseases caused by the direct effect of L1 insertion and few neurological diseases that were demonstrated to misregulate L1 retrotransposition, it is still unknown whether L1 retrotransposition can directly cause neurological disorders (Muotri et al., 2010; Coufal et al., 2011; Thomas et al., 2012; Erwin et al., 2014).

Alzheimer's disease is the main cause of dementia in the elderly, affecting almost two third of the world population over 65 (Bettens et al., 2013). It has been recently demonstrated that AD patients suffer from vitamin B12 deficiency and high homocystein content in blood, which contribute to the disregulation of S-adenosylmethionine synthesis and DNA methylation (Scarpa et al., 2006).

Since these DNA epigenetic alterations observed in AD patients could have an impact on L1 retrotransposition, we decided to investigate the Copy Number Variation (CNV) of L1 sequences in the genome of different tissues from AD patients and healthy controls (from three different cohorts), and in a mouse model of the disease.

By using the qPCR with Taqman probes, we observed in some of the tissues that we analyzed, a decreased amount of full length L1 sequences in AD patients compared to controls. Assuming that in AD patients there is a lower degree of DNA methylation, we can speculate that a higher retrotransposition of L1 elements occurred in certain neurons, may have caused the death of these cells, eventually leading to the detection of a lower amount of L1 sequences in AD patients.

We then developed new Taqman assays to study L1 CNV in the TgCRND8 mouse at P0 and at two stages of the adulthood (3 and 8 months). We observed in transgenic mice a higher amount of L1 sequences in the cortex at P0, in the hippocampus at 3 months,

while no difference were detectable at 8 months of age, supporting the hypothesis that in AD there is a higher L1 retrotransposition that causes cell death.

We also set up a technique called SPAM (SPlinkerette Analysis of Mobile elements), aimed at identifying the insertion sites of L1 sequences in the human genome. After a first test of the protocol, we used this technique to compare L1 insertion profile of AD patients and controls.

Interestingly, we found that AD susceptibility genes present several novel insertion sites that would warrant future investigation.

Finally we adapted the SPAM technique to a different repetitive element, the *roo* element, in a different organism: the *Drosophila melanogaster*, demonstrating that SPAM technique is a scalable approach, suitable for the integration sites discovery of different kinds of repetitive elements in different organisms.

Abbreviations

AD: Alzheimer's Disease

AIS: Annotated Integration Site

APP: Amyloid Precursor Protein

CNV: Copy Number Variation

CTF: C-Terminal Fragment

DNMT: DNA methyltransferase

dsDNA: double-strand DNA

EGFP: Enhanced Green Fluorescent Protein

ENi: Endonuclease-independent retrotransposition

EO-FAD: Early Onset Familial AD

fAβ: fibrillar Aβ peptide

FACS: Fluorescence-Activated Cell Sorting

gDNA: genomic DNA

GWAS: Genome-Wide Association Studies

hESC: human Embryonic Stem Cell

iPS: induced Pluripotent Stem

L1: LINE1, Long Interspersed Nuclear Element-1

LOAD: Late Onset AD

LTR: Long Terminal Repeats

NCLI: Non-Classical L1 Insertion

NFT: Neurofibrillary Tangle

NIS: Novel Integration Site

NPC: Neural Progenitor Cell

oAβ : soluble Oligomers of Aβ peptide

ON: Over Night

ORF: Open Reading Frame

PCR: Polymerase Chain Reaction

pfAβ: protofibrils of Aβ peptide

PNK: Polynucleotide Kinase

qPCR: quantitative real-time PCR

RT: Room Temperature

RNP: Ribonucleoprotein Particle

SINE: Short Interspersed Nuclear Element

SNP: Single Nucleotide Polymorfism

SPAM: SPlinkerette Analysis of Mobile elements

spPCR: splinkerette PCR

TE: Transposable Element

TPRT: Target Primed Reverse Transcription

TSD: Target Site Duplication

WGS: Whole-Genome Sequencing

Introduction

Transposable elements

Transposable elements (TEs), present in the genomes of all plants and animals, comprise a multitude of repetitive DNA sequences, all having the ability to mobilize and change locations in the genome (Kazazian, 2004).

First discovered in maize plants by the geneticist Barbara McClintock in the mid-1940s, they were initially considered a sort of genetic anomaly, and several decades later they acquired the label of *selfish DNA parasites*, since able to replicate independently and therefore being a potential threat towards genomic integrity (Fedoroff, 2012).

The ideas about TEs have evolved essentially over the past two decades, mostly thanks to the evidence that the extent of their presence in the genome of eukaryote is around 50%. Even if the modern view of transposable elements is still controversial, it is now clear that genomes have coevolved with them, on one side protecting themselves from their uncontrolled expansion, and on the other side taking advantage of their presence. TEs, and retrotransposons in particular, seem to have had an important role in driving genome evolution, influencing gene expression and contributing to tissue-specific transcriptional programs, in particular as enhancer-like elements and regulator of chromatin structure (Goodier and Kazazian, 2008; Bodega and Orlando, 2014).

According to their mechanism of mobilization, TEs can be universally divided in two main groups: class II TEs, or DNA transposons, that move throughout the genome using a “cut and paste” mechanism, and class I TEs, or retrotransposons, that multiply themselves in the genome with a “copy and paste” mechanism (Wicker et al., 2007).

DNA transposons in the mammalian genome

DNA transposons are mobile elements that move throughout the genome using the so called "cut-and-paste" mechanism: they are first cut out from their original position as double-stranded DNA and then reinserted elsewhere in the genome. DNA transposons encode a transposase enzyme that is able to catalyze both the excision and integration of the repetitive element by recognizing the flanking terminal inverted repeats (TIR), giving rise to a non-replicative mechanism during which the transposon is moved to a new genomic location without the generation of new copies of the element, although

there are few exceptions. During the insertion, the target site DNA is duplicated at both the ends of the transposable element, forming the so called target site duplications (TSD), which are unique for each different DNA transposon (Muñoz-López and García-Pérez, 2010).

The classification of DNA transposons is commonly made based on the sequence, the TIRs and /or TSDs. Among the subclass I there are: Tc1/*mariner*, PIF/*Harbinger*, hAT, Mutator, Merlin, Transib, P, *piggyBac* and CACTA. In the Subclass II there are *Helitron* and *Maverick* transposons, which are replicated and do not create double-strand breaks during their insertion. The most widespread TE family in nature is the Tc1/*mariner* that is present in diverse taxa as rotifers, fungi, plants, fish and mammal (Muñoz-López and García-Pérez, 2010) (Figure 1).

Figure 1: An example of a DNA transposon: the Tc1/*mariner* transposon. DNA transposons are flanked by inverted terminal repeats (ITRs) and have a single open reading frame (ORF) that encodes a transposase. They are also flanked by short direct repeats (DRs) created during the integration process (Ostertag and Kazazian, 2001).

Currently there are no active DNA transposons in mammals, and recent computational analyses indicated that their activity ceased in the primate lineage at least 37 million years ago (Pace and Feschotte, 2007). At present they comprise approximately the 3% of the human reference genome (Beck et al., 2011) and the 4% of the mouse genome (Keane et al., 2014).

Retrotransposons in the mammalian genome

Retrotransposons cover around the 40% of the sequenced mammalian genome (Lander et al., 2001; Waterston et al., 2002).

They can be subdivided in two main groups according to the presence of long terminal repeats (LTR): the LTR retrotransposons (like HERV in the human genome and IAP in mice) and the non-LTR retrotransposons, among which autonomous elements like LINE-1 (L1) and non-autonomous elements like Alu in humans and B1 in mice can be distinguished (Crichton et al., 2014). Autonomous elements encode for the enzymatic machinery that allows them to reverse transcribe their own RNA intermediate and to insert it in a new genomic location, while non-autonomous elements, not encoding the

necessary enzymes, exploit autonomous retrotransposons' machinery to transpose (Cordaux and Batzer, 2009).

LTR retrotransposons

LTR retrotransposons and retroviruses are very similar in structure: they both contain *gag* and *pol* genes that encode a viral particle coat (GAG) and a reverse transcriptase (RT), ribonuclease H (RH), and integrase (IN), able to create the cDNA from RNA and insert it in a new genomic location. The fundamental difference stands in the fact that retroviruses encode for a functional envelope protein that mediates their movement from one cell to another, while LTR retrotransposons do not encode for this gene or contain only a partial and non-functional gene, therefore being able only to reinsert in the genome of the same cell (Ostertag and Kazazian, 2001; Kazazian, 2004). These retrotransposons include elements such as the mouse intracisternal A-particles (IAPs) and the human endogenous retroviruses (HERVs) (Kuff and Lueders, 1988; Bénit et al., 1999; Ostertag and Kazazian, 2001) (Figure 2).

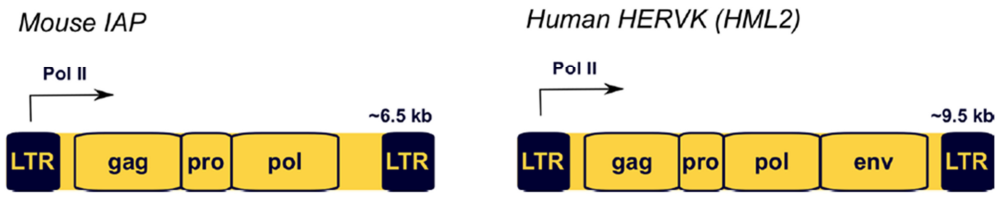


Figure 2: Mouse and human examples of LTR retrotransposons. Transcription regulatory regions are indicated with filled rectangles, and the main protein coding regions with open rectangles. Transcriptional start sites are shown with an arrow. Some LTR retrotransposons, like IAP, have lost the env gene present in their infectious progenitors (Crichton et al., 2014).

The present-day human genome contains ~400000 copies of HERVs that seem to be replication-incompetent, because of the numerous mutations and deletions that they accumulated. Although not apparently active, HERVs are classified as transposons for their proven transpositional activity during the human evolution and for the existence of animal ERVs, such as the mouse IAPs, that are currently able to mobilize (Katoh and Kurata, 2013).

IAP is the most active ERV class in mouse genome, with 1000–2500 copies estimated to be present in the haploid genome. Although the number of IAP elements is smaller than that of L1 elements, these retrotransposons might be more active in the mouse (Sharif et al., 2013).

Non-autonomous non-LTR retrotransposons

The major class of non-autonomous retrotransposons is represented by short interspersed nuclear elements (SINEs) which alone make up around 10% of the sequenced mammalian genome (Crichton et al., 2014). These elements derive from a series of small RNA polymerase III transcripts such as 7SL RNA, 5S rRNA, and tRNAs, and depend on L1-encoded proteins, in particular L1 ORF2, to mediate their retrotransposition, since they do not encode for any protein (Dewannieux et al, 2003; Dewannieux and Heidmann, 2005; Crichton et al., 2014).

The most important family of SINEs in the human genome is represented by Alu elements (~300 bp), while the corresponding in the mouse genome is represented by the B1 elements (~140 bp), both deriving from the signal recognition particle 7SL cellular RNA. Although having the same origin, mutations and rearrangements have made these elements quite different in structure (Crichton et al., 2014) (Figure 3).

In the human genome there are approximately 1100000 copies of Alu elements, whereas roughly 550000 B1 elements are present in the mouse genome (Ponicsan et al., 2010).

Alu elements are characterized by a dimeric structure (coming from the fusion of two similar but not identical monomers) in which the left monomer is separated from the right one by an A-rich linker region. At the 5'end there is an internal RNA polymerase III promoter (A and B boxes), while at the 3'end there is a polyA tail of variable length. Moreover, since the Alu sequence does not contain a polymerase III terminator, usually the transcripts include a portion of the downstream genomic sequence until the next terminator (Cordaux and Batzer, 2009; Hancks and Kazazian, 2012).

Figure 3: Mouse and human examples of SINE retrotransposons. Transcription regulatory regions are indicated with filled rectangles, and the main protein coding regions with open rectangles. Transcriptional start sites are shown with an arrow (Crichton et al., 2014).

Unlike the human Alu elements, the mouse B1 is a monomer (Vassetzky et al., 2003).

Like Alu elements, B1's internal promoter is composed of two boxes (A and B) spaced by 30-45 bp. During the early stages of the evolution of this element, a 29-bp tandem

duplication caused the formation of a further B box (B') located at 79-82bp from box A. Despite the large distance between boxes A and B', they form an active promoter (Koval et al., 2011).

Less abundant but still important are the other four SINE families present in the murine genome: B2, B4/RSINE, ID, and MIR (Ohshima and Okada, 1994).

Another important non-autonomous retrotransposon, present only in the human genome, is represented by SVA elements (Figure 4). SVA (SINE-R/VNTR/Alu) elements account for about 0.2% of the human genome (Hancks and Kazazian, 2012).

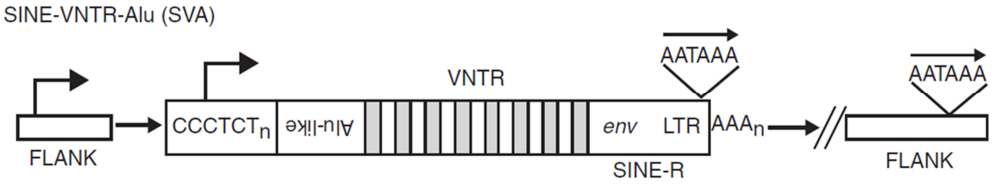


Figure 4: Structure of the human full-length canonical SINE-VNTR-Alu (SVA). From the 5'end there are a CCCTCT repeat of varying length, a sequence sharing homology to two antisense Alu fragments (Alu-like), a variable number of GC-rich tandem repeats (VNTR), a partial envelope (env) and right LTR sequence derived from an extinct HERV-K10 (SINE-R). SVAs are RNA PolII transcripts, however whether SVAs encode their own promoter is unknown. Transcription of SVA RNAs may occur upstream (black bent arrow) of a genomic SVA or may be initiated throughout the SVA (black bent arrow). SVA RNAs terminate at a polyA signal (AATAAA) located at the 3'end of the SINE-R, but may also bypass this signal for a downstream polyA signal. Likewise, SVA genomic insertions also terminate in a polyA tail (AAAn) and are flanked by a TSD (Hancks and Kazazian, 2012).

They are ~2 kb long and composed of an hexameric repeat region, followed by an inverted Alu-like sequence, a variable number of tandem repeats region, a HERV-K10-like region (SINE-R) and a polyA tail of variable length. Although SVA elements do not contain an internal promoter, they are transcribed by RNA polymerase II and the resultant RNA transcript seems to be mobilized by the L1-encoded proteins (Beck et al., 2011; Cordaux and Batzer, 2009).

The mammalian genome includes another class of non-autonomous retrotransposons: the processed pseudogenes. Processed pseudogenes are cDNA copies of mRNA molecules that have been inserted into the genome by the L1 enzymatic machinery. They tipically lack intronic RNA, usually have polyA tails, and are flanked by TSDs (Ostertag and Kazazian, 2001).

Processed pseudogenes do not encode a functional protein, but although they accumulated frameshift mutations and premature stop codons during evolution, few of them are transcriptionally active (Ding et al., 2006).

Autonomous non-LTR retrotransposons

Long interspersed nuclear element-1 (LINE1 or L1) is the most prevalent family of autonomous retrotransposons in mammals (Huang et al., 2012). The great majority of these elements are inactive due to 5'truncations, inversions or point mutations, but a part of them are still active, and with their de novo integration have been demonstrated to be able to cause even diseases in both mice and humans (Crichton et al., 2014; Huang et al., 2012; Maksakova et al., 2006; Ostertag and Kazazian, 2001).

The human Long interspersed nuclear element-1 (L1)

About half of the human genome is composed of transposable elements (de Koning et al., 2011). The most abundant of these elements is represented by the L1 element, which alone makes up about 17% of our DNA (~500000 copies, of which ~5000 are full length). L1 elements are autonomous retrotransposons that are still active, with a mobilization that started several millions of years ago (Konkel and Batzer, 2010; Viollet et al., 2014).

L1 elements can be divided into several subfamilies (pre-Ta, Ta-0, Ta-1, Ta1-d, Ta1-nd) based upon sequence polymorphisms contained within their 5′ and 3′UTRs (Beck et al., 2010). A subset of human L1 elements called Ta family that is still active in the human genome arose ~4 Myr and subsequently differentiated into two major subfamilies, Ta-0 and Ta-1. Ta-1 is younger than Ta-0, but thanks to its current high replicative activity, it presently forms at least 50% of the entire human Ta family (Boissinot et al., 2000).

After the peak of retrotranspositional activity occurred around 40 Myr ago, L1's activity declined, and the human genome currently harbors only ~80 to 100 potentially active L1 elements per diploid genome. Of those, only 6-8 have been defined as "hot", and seem to be involved in the majority of new and recent insertions (Brouha et al., 2003; Ohshima et al., 2003).

The full-length human L1 retrotransposon is ~6 kb long. It is composed by a 910-bp 5'UTR region, two ORFs (ORF1 and ORF2) both necessary for L1 retrotransposition, separated by a 63-bp inter-ORF region, and a 205-bp 3'UTR, containing a weak but functional polyadenylation signal. At the end of the 3'UTR sequence there is a polyA tail of variable length, and at both the ends of the L1 element there are the so called target site duplications (TSDs) of 2–20 bp, deriving from the insertional event (Szak et al., 2002; Babushok and Kazazian, 2007) (Figure 5).

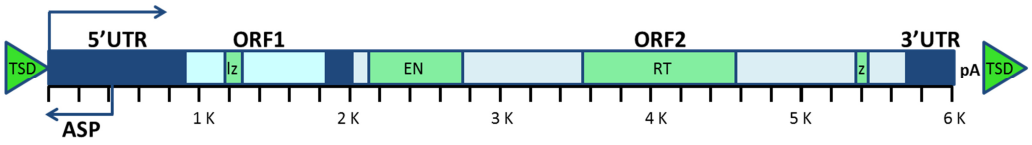


Figure 5: Structure of the human L1. The human L1 is 6kb long and is composed of a 5'untraslated region (UTR) having both sense and antisense promoter activity (ASP), two open reading frames, ORF1 and ORF2, a 3'UTR and a polyadenilation signal (pA). ORF1 encodes for a 40kDa RNA-binding protein with a leucine zipper (lz) domain. ORF2 encodes for a 150kDa protein with endonuclease (EN) and reverse-transcriptase (RT) activities and present a zinc knuckle (z) domain at its 3'end. the structure is included between target site duplication (TSD) of the insertion site (Babushok and Kazazian, 2007).

The protein encoded by ORF1 is a ~40-kDa protein (p40 or ORF1p) mostly composed by basic residues and structurally characterized by the presence at the N-terminus of a leucine zipper domain. ORF1p has been demonstrated to possess nucleic acid binding and chaperone activity, but it has still to be clarified its precise function, that seems to be crucial for L1 retrotransposition (Martin et al., 2005; Goodier et al., 2007).

The protein encoded by ORF2 is a ~150-kDa protein (ORF2p) with a double domain: endonuclease (EN) and reverse transcriptase (RT), and a cysteine-rich domain at the C-terminus. The origins of the EN domain, present in several non-LTR retrotransposons, go to the early eukaryotes, in which it derived from the host cells' apurinic/apyrimidinic endonuclease (APE) domain (Malik et al., 1999).

When the L1 element inserts in a new genomic location through the target primed reverse transcription (TPRT), it recognizes the EN cleavage consensus site 5′-TTTT/A-3′, which facilitates the EN interaction with the DNA minor groove (Berry et al., 2006).

The RT domain, which is present in all the non-LTR retrotransposons, has been demonstrated to possess both RNA- and DNA-dependent DNA polymerase activities, and is the peculiar mediator of L1 retrotransposition (Malik et al., 1999; Piskareva and Schmatchenko, 2006).

The cysteine-rich domain at the C-terminus, conserved in all the L1 elements, is characterized by the presence of a zinc knuckle-like region, which seems to be involved in ORF2-DNA interactions during L1 integration or in facilitating polymerization by RT (Ostertag and Kazazian, 2001; Babushok and Kazazian, 2007). Mutations at the level of this domain have been demonstrated to reduce L1 retrotransposition activity in a cultured cell assay, without an involvement of RT activity (Moran et al., 1996).

<u>The murine Long interspersed nuclear element-1 (L1)</u>

About 10% of the entire mouse genome is composed of L1 elements (more than 100000 copies), but most of them are inactive, as consequence of 5'-end truncations, inversions

or mutations, so the total of potentially active elements is ~3000 (Goodier et al., 2001). Murine L1's structure is very similar to the human one (Figure 6).

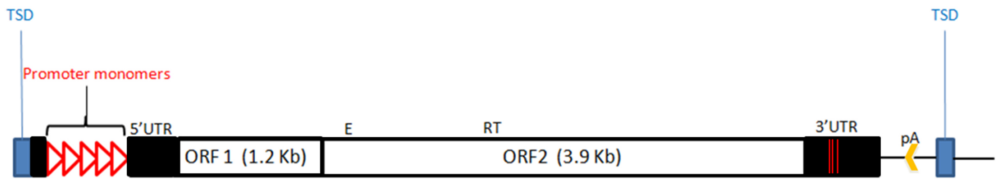


Figure 6: Structure of the mouse L1. The 5'UTR region is composed of a variable number of monomers (~200 bp, red triangles) having promoter activity. Polymorphisms are present at the 3'UTR, here indicated with red vertical axis (Mears and Hutchison, 2001).

The major differences can be observed in the promoter region: indeed, the 5'UTR region is characterized by the presence of repeated conserved monomers of about 200bp followed by a short non monomeric region. Some *in vitro* experiments, aiming at testing the activity of different regions in the 5'UTR of a mouse L1 family, demonstrated that the promoter activity depends on the monomeric region, and it seems to increase according to the number of these sequences (Ostertag and Kazazian, 2001).

Indeed, monomers can vary in terms of number and sequence, and based on these differences, it is possible to separate L1 elements in different subclasses (V, A, Tf and Gf), all deriving from the same common ancestor. The V family, without identifiable monomers, seems to be the oldest one and it is inactive. L1 elements from the A family, with approximately 6500 full length elements, are characterized by a monomeric structure at the 5'UTR. Some of these full length L1s seem to be active, due to the presence of intact ORF1 and ORF2 sequences, and are transcribed.

The Tf and Gf families, composed by several L1 elements each, derive from a common ancestor (the L1element of the F family) and are characterized by a peculiar monomeric structure. The Tf family is represented by 1800 active elements among 3000 full length members, whereas the most recently discovered Gf type includes 400 active elements among 1500 full length members (Goodier et al., 2001).

L1 retrotransposition

During its evolution, the mammalian genome has been largely expanded and shaped by the direct retrotransposition of L1 elements, but also by the mobilization *in trans* of other non-autonomous mobile elements and the generation of processed pseudogenes (Richardson et al., 2014). The study of L1 structure has elucidated many aspects of the mechanism of L1 retrotransposition, a multi-step process which eventually results in a de novo insertion.

Molecular mechanisms of L1 retrotransposition

L1 retrotransposition begins with the transcription by RNA polymerase II of L1 sequence starting from its own internal promoter, leading to the generation of a bicistronic mRNA (Figure 7a). This mRNA molecule has a polyA tail that can be encoded by its own weak but functional polyadenylation signal, or by a signal present in the downstream genomic sequence, in this case leading to the so called L1-mediated 3'-transduction. Concerning further typical RNA modifications, it is still not known whether L1 transcripts are added with a 7-methylguanosine cap, while it is clear that, since L1 elements do not contain introns, they do not require splicing (Ostertag and Kazazian, 2001).

The next step involves the transport of the L1 RNA molecule to the cytoplasm through a still unclear mechanism (Figure 7b), where the ORF1 and ORF2 sequences are translated into proteins (Figure 7c). After that, multiple copies of ORF1p and only few copies of ORF2p together with the L1 RNA molecule bind each other, creating a stable ribonucleoprotein complex (RNP) (Figure 7d) (Kulpa and Moran, 2006; Beck et al., 2011).

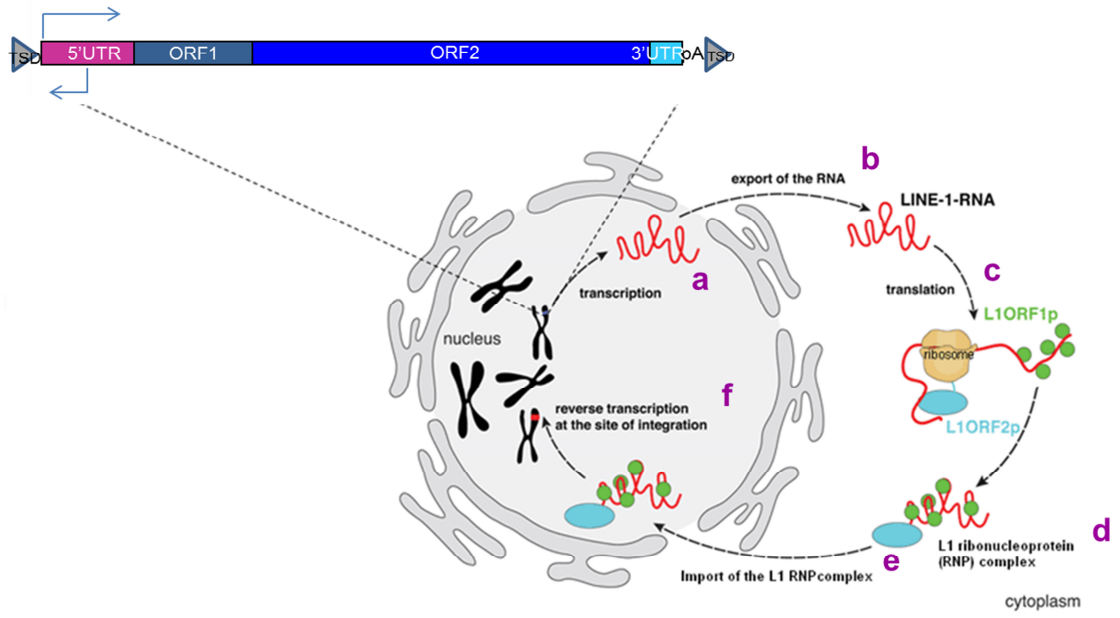


Figure 7: L1 retrotransposition mechanism. After L1 sequence transcription by RNA pol II (a), the mRNA molecule is exported from the nucleus to the cytoplasm (b), where it is translated in its two encoded proteins: ORF1p and ORF2p (c). These proteins bind to an L1 mRNA molecule to form a ribonucleoprotein (RNP) complex (d), which is then transported back to the nucleus (e), where reverse transcription and integration of the new L1 copy occur (f) (Adapted from www.pei.de).

This complex, through a mechanism that has not yet been clarified, but probably involving active transport through a nuclear pore or nuclear membrane breakdown at mitosis or meiosos, is carried back into the nucleus (Figure 7e).

Once in the nucleus, two different L1 integration mechanisms can be mediated (Figure 7f): the canonical target primed reverse transcription (TPRT) or the endonuclease-independent (ENi) retrotransposition, also known as non-classical L1 insertion (NCLI) (Viollet et al., 2014) (Figure 8).

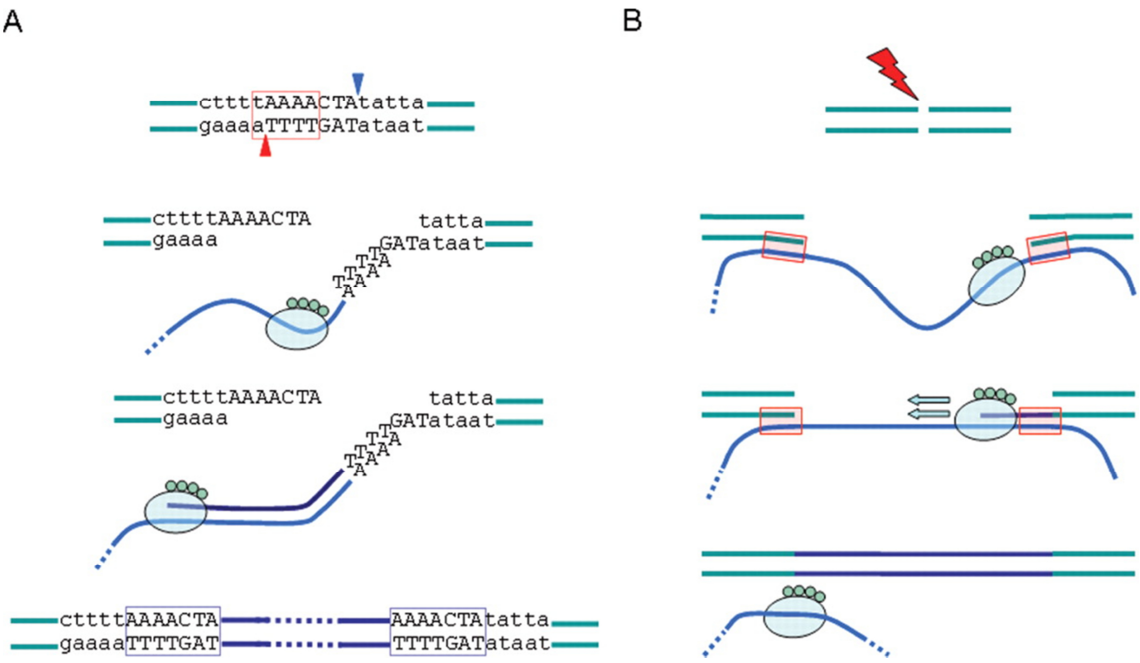


Figure 8: Comparison between TPRT and ENi L1 insertions. (A) Classical TPRT-mediated L1 insertion. First-strand cleavage by the L1 endonuclease (red arrowhead) at the 5'-TTTT/A-3' consensus (red dotted box) allows L1 mRNA (blue line) to anneal to genomic DNA using its poly(A) tail. Reverse transcriptase activity of L1 ORF2 (green oval) synthesizes L1 cDNA (purple line) using L1 mRNA as template and 3'OH from nicked genomic DNA as primer. Second-strand cleavage (blue arrowhead) occurs 7–20 bp downstream from first-strand cleavage site, creating staggered nicks which are later filled in to form TSDs (blue dotted boxes). Attachment of the L1 cDNA and synthesis of the second strand complete the insertion process. (B) Schematic representation of an ENi event. Following creation of a genomic double-strand break (red thunderbolt), free-floating L1 mRNA (blue line) attaches to newly separated ends using small stretches of complementary bases. Once gap is bridged, it may be filled in by DNA synthesis by either the L1 RT, cellular repair polymerases or both (Sen et al., 2007).

The target primed reverse transcription (TPRT) consists of a coupled reverse transcription/integration process (Morrish et al. 2002). During TPRT, ORF2p endonuclease activity produces a single-strand nick in the genomic DNA preferentially at the consensus sequence 5′-TTTT/A-3′. The ORF2p reverse transcriptase activity, priming the reaction within the polyA tail, extends the free 3′-OH group using the L1 RNA as a template (Viollet et al., 2014). After that, the second strand at the integration site is cleaved and used to prime the synthesis of the cDNA second strand. The typical

hallmarks of this TPRT-derived integration mechanism include the target site duplications (TSDs), 7-20 bp sequences present at each end of the new L1 element, and a dA-rich tail of variable length (Sen et al., 2007; Cordaux and Batzer, 2009).

Another feature of this integration process is the fact that the majority of the newly inserted L1 elements are 5' truncated, and therefore unable to retrotranspose any longer. This truncation may be caused by an inefficiency of the reverse transcriptase in the polymerization process of the new cDNA copy, or by the activity of a cellular RNAse H, reflecting a possible attempt by the host defense machinery to protect the genomic integrity (Ostertag and Kazazian, 2001; Beck et al., 2011).

In the endonuclease-independent (ENi) retrotransposition process, at the level of a pre-existing double-strand break, without the need for a further endonuclease cleavage, L1 mRNA molecules can attach to the protruding ends using small stretches of complementary bases. At this point the L1 reverse transcriptase, the host repair polymerase or both synthetize the missing DNA bases, leading to an L1 integration that lacks the structural features of the TPRT-mediated insertion.

The typical hallmarks of this alternative process are unusual structures caused by L1 integration at atypical target sequences, L1 truncations predominantly at the 3′ends and lack of TSDs (Morrish et al., 2002). L1 integrations mediated by the ENi mechanism have been observed at the level of telomeres (Viollet et al., 2014), although it seems that it is a significantly less efficient process, rarely found in vivo (Morrish et al., 2002; Babushok et al., 2006).

Molecular mechanisms influencing L1 retrotransposition

Given the potentially dangerous effect of new L1 insertions in the genome, cells have evolved some defense mechanisms able to block retrotranspositional activity and therefore to ensure genome stability across generations (Crichton et al., 2014) (Figure 9).

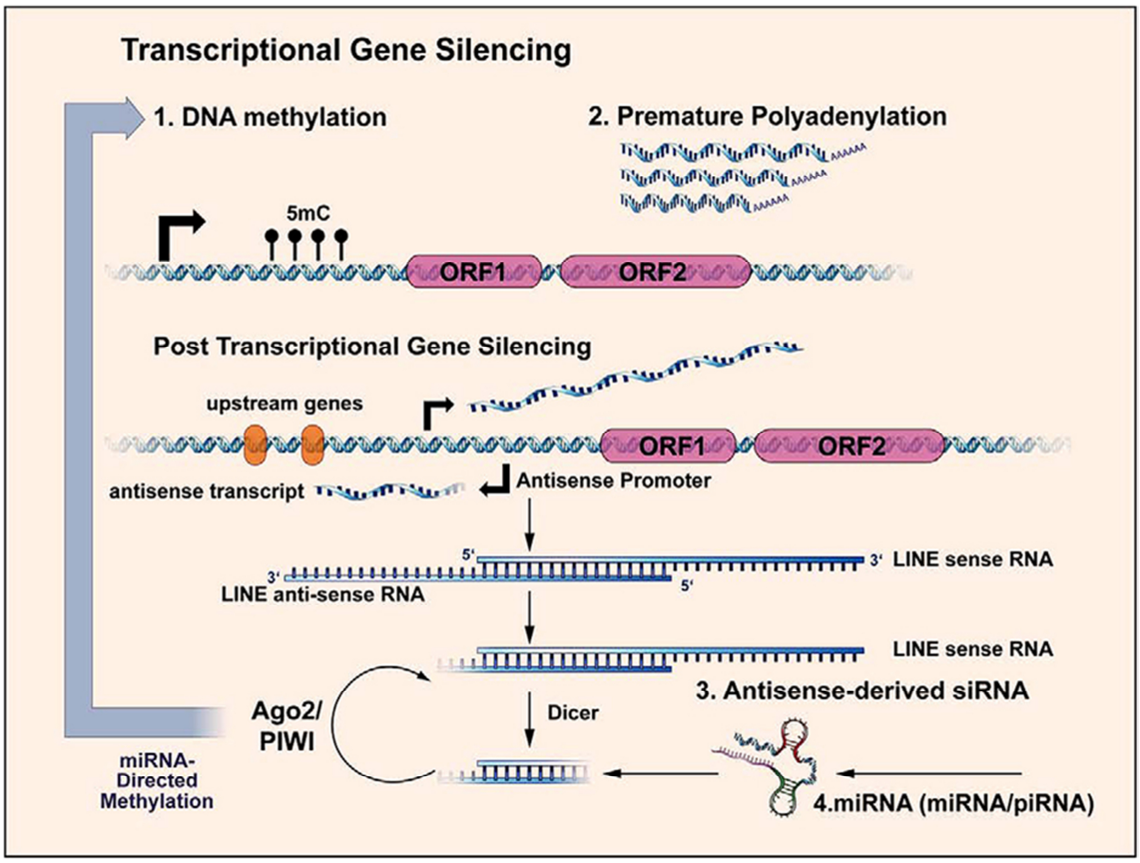


Figure 9: Cellular mechanisms limiting L1 retrotransposition. L1 transcription is suppressed by DNA methylation and can be interrupted by premature polyadenylation. RNA interference can also regulate L1 post-transcriptionally via small RNAs facilitated by Argonaute and PIWI proteins. Small RNAs may in-turn direct DNA methylation (St. Laurent et al., 2010).

These cellular mechanisms can act at different levels: transcriptional, post-transcriptional and likely post-translational (Bogerd et al. 2006). The three principal mechanisms are: extensive DNA methylation at L1 promoter sequence, premature polyadenilation and RNA interference using miRNA, piRNA and L1 antisense-derived RNAs.

L1 promoter methylation

DNA methylation in mammals commonly occurs on cytosine residues at the level of CpG dinucleotides, usually causing transcriptional repression. The promoter region at the 5'UTR of the L1 sequence contains several CpG dinucleotides that can be methylated, with a consequential repression of L1 mRNA expression and therefore L1 retrotransposition (Yoder et al., 1997). As a consequence, cellular enzymes involved in DNA methylation, such as methyltransferases, are important in controlling L1 retrotransposition. In the case of the de novo methylatransferase 3 (DNMT3), it has been demonstrated that its absence in mouse germ cells is linked to demethylation of L1 promoter and higher L1 RNA levels (Bourc'his and Bestor, 2004).

Another anzyme that has been reported to be essential in regulating L1 expression is the methyl-CpG-binding protein 2 (MeCP2), involved in the global DNA methylation. Indeed neural precursor cells knock out for MeCP2 show an increased L1 retrotransposition compared to wild type cells (Muotri et al., 2010).

L1-mRNA truncation by premature polyadenylation

The 3'UTR sequence of the L1 element has been demonstrated to contain 19 canonical and non-canonical potential polyadenylation (PolyA) signals that can be alternatively used leading to the generation of L1 transcripts of variable length. The premature transcription termination can have, as a consequence, the generation of truncated L1 forms unable to mobilize, helping the host cells in maintaining the genomic integrity (Perepelitsa-Belancio and Deininger, 2003).

It has been also demonstrated that some of these PolyA signals are weak, so that a part of the downstream genomic sequence can be included in the L1 transcript and eventually inserted in a new genomic location together with the L1 element (Belancio et al., 2007; Goodier et al., 2000; Moran et al., 1999; Pickeral et al., 2000).

L1 silencing through RNA interference (RNAi)

Another mechanism involved in L1 retrotransposition repression in host genome is represented by RNA interference, which silences L1-mRNAs post-transcriptionally. Small silencing RNAs in metazoa can be divided into three classes: small interfering RNAs (siRNAs), microRNAs (miRNAs) and PIWI-interacting RNAs (piRNAs), characterized by different biogenesis processes (Malone and Hannon, 2009).

Both miRNAs and siRNAs are processed through cleavage of their precursors by the Dicer endonuclease, followed by the association to the AGO protein, while piRNAs undergo a Dicer-independent mechanisms. Moreover, piRNAs are associated with an AGO protein family that is specifically expressed in the cells of the germline (the PIWI proteins), while siRNAs and miRNAs are associated with AGO proteins that are ubiquitously expressed (Siomi et al., 2011). Even if an increase in L1 transcripts content was found in embryonic stem cells of dicer knockout mice (Kanellopoulou et al., 2005), it is still unclear whether L1s are targeted by siRNA (Goodier and Kazazian, 2008).

On the other hand, a well established mechanism for post transcriptional inhibitions of L1 mRNA is mediated by piRNAs, (24-32 nt in length), that are processed from single-stranded RNA precursors and transcribed mostly from intergenic repetitive sequences called piRNA clusters. The main function of this class of small non-coding RNA is to

protect the genomic integrity of germ cells from the potentially detrimental action of "genomic parasites" such as the transposable elements.

After being processed from the precursor RNA molecules, the primary piRNAs are associated to the PIWI proteins and guided to the target. Once the piRNA molecule recognizes the L1-derived mRNA, Slicer can mediate the disruption of the mRNA or the production of a secondary piRNA. This secondary piRNA can trigger an amplification process, also known as "ping-pong cycle", during which the piRNA binds to the antisense strand of the target mRNA causing its disruption, with the production of another antisense piRNA molecule which again can target the sense L1 mRNA (Siomi et al. 2011). piRNAs actually seem to be involved also in methylation-mediated L1 silencing, by directing the DNA methylation machinery to L1 elements, as shown in PIWI-null mutant mice (Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008).

In mouse spermatogonia, the lack of MIWI2 and MILI (two PIWI proteins) causes an increase in L1 elements activity (Carmell et al. 2007; Reuter et al. 2011).

L1's retrotransposition effects

The most direct effect of L1 retrotransposition is the increase of the genome size, indeed L1 elements and Alu sequences together added about 750 million bases (Mb) to the human genome, with important consequences not only from the structural, but also from the functional point of view (Cordaux and Batzer, 2009).

Genomic structural modifications

L1 elements can alter the genome structure in many ways:

Insertional mutagenesis. This is the first genomic modification induced by L1 insertions that has been described. In this case L1 elements insert in the exon of a gene, inducing an interruption of the coding sequence (Kazazian et al., 1988).

Deletions at the insertion site. When a new L1 element inserts in the genome it can cause the loss of the genomic sequence close to the integration site, probably due to an involvement of the host DNA repair apparatus, that identifies the double strand break and repairs it causing the loss of genomic DNA (Gilbert et al., 2002).

In cultured human cells it has been observed that the 10% of the integrations mediated by en engineered L1 element where characterized by genomic loss, as also observed in the human and chimpanzee genomes (Beck et al., 2011). In the human genome it was successfully characterized a deletion of 46 kb in the PDHX gene, causing pyruvate dehydrogenase complex deficiency, that revealed the presence of a full length L1

element in the coding sequence of the gene, in a region corresponding to the deleted part (Miné et al. 2007).

Non allelic homologous recombination. The presence of several copies of LINE and SINE elements randomly distributed in the genome can lead to crossing over between these sequences, and therefore non allelic homologous recombination (NAHR) with resulting structural variations that seem to account for approximately the 0.3% of human genetic diseases (Belancio et al., 2008). Through the alignment between the human and the chimpanzee genomes it was possible to identify the presence of 55 human specific L1-NAHR events (Han et al., 2005).

3' and 5'transduction. As previously mentioned, the polyadenylation signal present at the 3'UTR of the L1 element is functional but weak, and therefore often substituted by downstream stronger signals. This mechanism, the so called 3' transduction, causes the transcription and possibly the retrotransposition of a segment of genomic sequence present at the 3' of the L1 element. 3' transduction events were reported in mouse and human, where 15 out of 66 uncharacterized L1 sequences were demonstrated to carry 3'genomic sequences with an average length of 207 bp (Goodier et al., 2000).

Usually shorter and less common is the 5' transduction that can be identified only at the level of full length L1 elements: in this case the transduction occurs when the L1 elements is transcribed starting from an upstream promoter, causing the retrotranpsosition of a segment of genome upstream to the L1 5'UTR (Beck et al., 2011).

Heterochromatinization. In 1998, with the description of the lyonization process that leads to the chromosome X inactivation, it was hypothesized for the first time a possible involvement of L1 elements in the heterochromatinization of the X chromosome (Lyon, 1998). Only recently it has been demonstrated that actually L1 elements participate during the process in two steps: silent L1s create a heterochromatic compartment in wich genes are recruited, while a subset of active and expressed L1s help in X-chromosome inactivation propagation to those genes that are prone to escape (Chow et al., 2010).

Transposition-mediated toxicity. After insertional mutagenesis, cell cycle arrest and apoptosis are other dangerous toxic effects that L1 retrotransposition can have in cells (Gasior et al., 2006; Belgnaoui et al., 2006). Indeed it seems that the production of several DNA double strand breaks mediated by the endonuclease expressed by the L1

element can lead to apoptosis and senescence (Gire et al., 2004; Wallace et al., 2008; Gao et al., 1998).

Potential effects on gene expression

According to the specific site of integration, the structural variations induced by L1 insertions can affect gene transcription and expression by the generation of splice sites, adenylation signals and new promoters that can finally generate new reorganized transcription units (Faulkner et al., 2009).

When a new L1 element falls in an intergenic region or an intron it can have no effects, while if the integration occurs in an exon or a regulatory sequence it can heavily modify gene expression or function mainly by destroying coding or regulating sequences (Viollet et al., 2014).

It has been observed that when an L1 element inserts in an intron actually it can cause the alternative splicing of a coding gene by exon skipping or exonization. If the new L1 integration disrupts a splice site that is bypassed during splicing, than it causes exon skipping, and therefore the generation of a defective transcript, while if the new L1 integration contains a donor or acceptor splice site, the L1 sequence can be included in the transcript as an exon of variable length according to the position of the splice site (Zemojtel et al., 2007). Also the activity of the antisense promoter (ASP) can alter gene expression: in vitro experiments on human embryonic stem cells actually demonstrated that the ASP can be used as alternative promoter in driving the transcription of the genomic sequence upstream to the 5'UTR of the L1 element also in a tissue-specific way, inducing tissue-specific gene expression of peculiar genes (Mätlik et al., 2006; Macia et al. 2011).

L1 retrotransposition in germline and during development

It is not clear when L1 retrotransposition exactly occurs during development in mammals, but the total number of L1 and Alu elements that populate mammalian genomes suggests that they mobilize in the germline (Levin and Moran, 2011).

As opposed to somatic retrotransposition, when L1 mobilization occurs in germline or pluripotent cells, it can be inherited by the following generations (Erwin et al., 2014). Several studies that employ endogenous and engineered L1 elements support this thesis: for instance, it has been shown that the mouse full-length L1 RNA and the L1 ORF1p are co-expressed in leptotene and zygotene spermatocytes during meiotic prophase (Branciforte and Martin 1994). The mouse ORF1p has been also demonstrated to be

expressed in the cytoplasm of oocytes during specific stages of their development (Trelogan and Martin, 1995). Thanks to experiments performed on transgenic mice, it has been observed that an engineered human L1 can mobilize in male germ cells, and finally, in human, an engineered human L1 seems to be able to mobilize in oocytes (Ostertag et al., 2002; Georgiou et al., 2009; Levin and Moran, 2011).

L1 mobilization during these developmental phases can be linked to the typical pattern of DNA hypomethylation that can be observed in the cells of the germline, fundamental for the epigenetic reprogramming that occurs during germ cells specification (Smallwood and Kelsey, 2012). Indeed, it has been demonstrated that germ cell populations of mice lacking de novo methyltransferase 3-like (Dnmt3L) present higher concentrations of L1 transcripts (Bourc'his and Bestor 2004).

Recently, Kano and colleagues showed the presence of high concentrations of L1-mRNA also in mouse embryos and they demonstrated that probably L1 mobilization and integration occurs more offen during embryogenesis instead of germline. In particular, by using an L1 transgenic mouse model, they detected high levels of L1-mRNA expression not only in germ cells but also in embryos, particularly at preimplantation stages and later on at E10.5 (Kano et al. 2009).

The group of Spadafora also demonstrated that the RT encoded by L1 has a fundamental role during the early embryonic development: by incubating mouse zygotes with the non-nucleoside RT inhibitor nevirapine or microinjecting murine zygotes with morpholino-modified antisense oligonucleotides against the L1 5'end region, they observed an arrest of development at the two- and four-cell stage (Pittoggi et al., 2003; Beraldi et al., 2006).

Since these somatic retrotransposition events are not incorporated into germ cells, they are not heritable and don't accumulate in the genome of all cells. Clearly, these events provide sources of genomic diversity within distinct somatic cells of an individual, generating somatic mosaicism in a particular organ (Vitullo et al., 2012) (Figure 10).

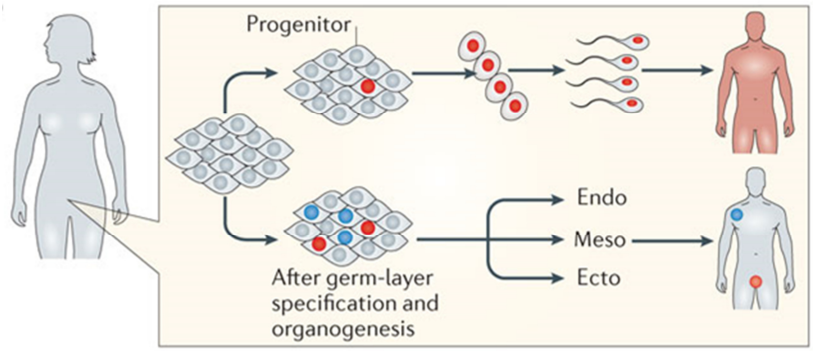


Figure 10: Consequences of somatic retrotransposition events. Somatic retrotransposition can happen at any time during embryogenesis. Retrotransposition events that occur in early pluripotent progenitor cells will result in somatic mosaicism: these unique cells will contribute to all tissues of the body of the individual, including the germ line. Somatic retrotransposition that happens after germ-layer specification and organogenesis, however, results in tissue-specific insertions that are not hereditary (Erwin et al., 2014).

Moreover, Malki et al. suggested that L1 elements, activated during the epigenetic reprogramming of the embryonic germline, could be involved in the process of fetal oocyte attrition (FOA) in mice, a process of elimination of more than two-thirds of meiotic prophase I (MPI) oocytes before birth. They showed that wild-type fetal oocytes can present different nuclear levels of L1 ORF1p and that experimental elevation of L1 expression is linked to increased MPI defects, FOA, oocyte aneuploidy, and embryonic lethality. They hypothesize that FOA is involved in the selection of those oocytes that have the least L1 activity, since they represent a lower risk for the following generations (Malki et al., 2014).

Somatic L1 retrotransposition in the brain

The nervous system is a complex network made by different subtypes of cells. At the same time, cells that belong to the same subtype can display different features, from both the structural and the functional point of view. The factors that are involved in the determination of these diversities comprise the epigenetic regulation, the alternative splicing and the post-translational modifications. The discovery of neurons with different genotypes, the so called somatic mosaicism, makes the nervous system more complex than ever thought (Erwin et al., 2014) (Figure 11).

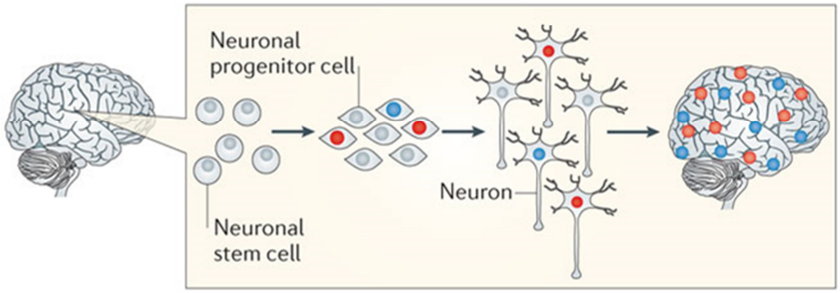


Figure 11: Consequences of somatic retrotransposition events in the brain. Somatic retrotransposition increases as neural stem cells differentiate into neurons and results in neurons with unique genomes (Erwin et al., 2014).

The first evidence of neuronal somatic mosaicism mediated by L1 occurring in the mammalian brain came from the research performed by Fred H. Gage and colleagues, who demonstrated that an engineered human L1 can retrotranspose in adult rat NPCs in vitro and in the mouse brain in vivo. They also demonstrated that the transcription factor SOX2 (SRY (sex determining region Y)-box 2) seems to repress L1 transcription in rat adult hippocampal neural stem cells, indeed during neuronal differentiation the low expression of SOX2 corresponds to a higher L1 transcription and retrotransposition (Muotri et al., 2005).

In 2009 they demonstrated that the engineered human L1 was able to transpose also in NPCs isolated from the human fetal brain and in NPCs derived from human embryonic stem cells hESCs. Moreover, as for the transcription factor SOX2, they proved that MeCP2's expression (the methyl-CpG-binding protein 2), previously demonstrated to associate with the L1 promoter and to be able to repress L1 transcription (Yu et al., 2001), was lower during neuronal differentiation than in mature neurons, proposing a model in which a lower methylation of L1 promoter during brain development may cause a higher L1 expression and probably retrotransposition (Coufal et al., 2009; Thomas and Muotri, 2012).

The same group in 2010 confirmed the correlation berween the absence of MeCP2 and the increased L1 retrottansposition in rodents, according to a model of L1 expression control in which the activation of L1 retrotransposition corresponds to the progression from neural stem cells to neural progenitors (Muotri et al., 2010; Erwin et al., 2014; Singer et al., 2010) (Figure 12).

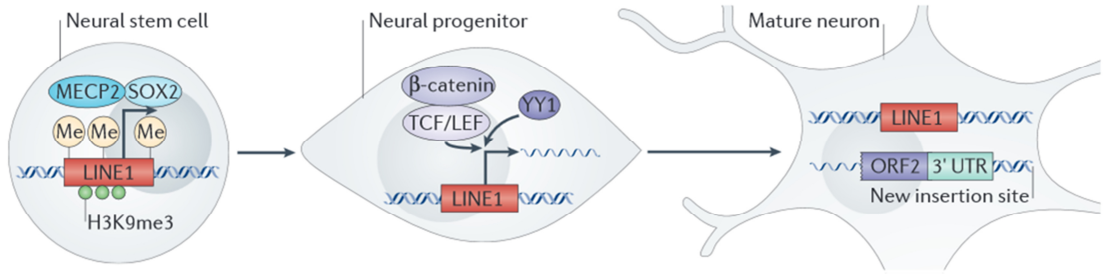


Figure 12: Regulation of retrotransposition in neural progenitors. In neural stem cells, L1 promoter is repressed by DNA methylation, H3K9me3 modifications, MECP2 (which binds to the methylated DNA) and SOX2. As neural stem cells transition to progenitors, SOX2 is no longer present. The LINE1 promoter assumes an open chromatin state and becomes de-methylated. MECP2 can no longer bind. The WNT transcription factors, β-catenin and members of the TCF/LEF family activate transcription, perhaps with the cooperation of another transcription factor, YY1. This results in an increase in LINE1 transcription in the progenitor and active retrotransposition (Erwin et al., 2014).

In order to assess the copy number variation (CNV) of L1 elements in human tissues, Coufal and colleagues also set up a protocol of Taqman qPCR and applied it to different human tissues and brain regions, allowing to estimate the presence of approximately 80 more L1 copies in the hippocampus compared to other organs such as heart and liver. They observed that there is a substantial variability between individuals, and also that the hippocampus, probably because of its neurogenic niche, seems to harbor a higher copy number of L1 elements compared also to other brain regions (Coufal et al., 2009; Richardson et al., 2014) (Figure 13).

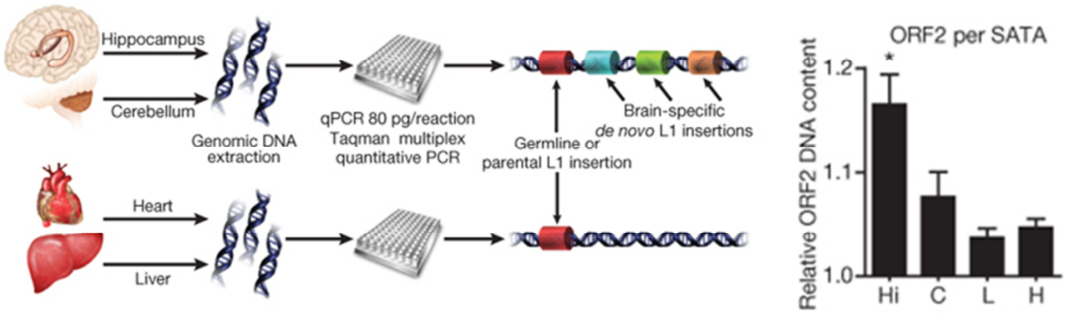


Figure 13: Experimental scheme of the multiplex quantitative PCR analyses of L1 copy number in human tissues and relative quantity of L1, standardized such that the lowest liver value was normalized to 1.0. Hi = hippocampus, C = cerebellum, H = heart, and L = liver (Coufal et al., 2009).

The CNV analysis is highly informational, but it does not give any clue about the exact insertion site in the genome of the L1 elements. More recent next-generation sequencing approaches developed in the last few years allow the identification of endogenous L1's insertion sites.

In 2011 the group of our collaborator Geoffrey Faulkner published a work in which a new high-throughput method, called RC-seq (retrotransposon capture sequencing) was applied on samples from the hippocampus and caudate nucleus of three individuals in

order to identify L1, Alu and SVA insertions. Thanks to this technique it was possible to identify known and novel retrotransposon insertions with uniquely mapped read pairs: besides several germline mutations, they found 7743 putative somatic L1 insertions, 13692 somatic Alu insertions and 1350 SVA insertions (Baillie et al., 2011).

Another method aimed at mapping L1 insertions, published by Evrony and colleagues in 2012, allowed the amplification and sequencing of genomes extracted from single human neurons. They performed the L1 insertion profiling of 300 neurons deriving from cerebral cortex and caudate nucleus of three individuals: in this way they were able to distinguish >80% of germline insertions from single neurons end to estimate that the rate of somatic insertions specific for each neuron was <0.6 in cortex and caudate, with most neurons lacking somatic insertions (Evrony et al., 2012).

This last observation was in contrast with what previously reported by Coufal and colleagues that estimated, by qPCR copy number assay, a rate of new insertions comprised between 80 and 800 for each hippocampal neuron.

These contrasting observations lead to several hypotheses: probably different individuals naturally present a different rate of L1 new insertions because influenced by environmental factors. Moreover it was demonstrated that some individuals can present a higher number of L1 elements in the hippocampus compared to the frontal cortex (Coufal et al., 2009; Baillie et al., 2011). The CNV assay performed by qPCR might be less accurate, but at the same time single cell sequencing might not detect insertions if the sequencing is performed on the wrong population (Erwin et al., 2014).

A further interesting observation comes from a paper published in 2009 by Muotri et al., in which they demonstrated the presence of higher L1 activity in neurons of mice exposed to voluntary exercise compared to sedentary mice, meaning that probably neuronal progenitor cells support de novo retrotransposition upon exposure to environmental factors, contributing to the physiological neuronal plasticity (Muotri et al., 2009, Thomas et al., 2012).

Human diseases linked to L1 retrotransposition

Besides skin, breast and colon cancers, the three still active non-LTR retrotransposons (L1, Alu and SVA) have been demonstrated to cause about the 0.27% of human genetic diseases, with at least 50 different disorders, exhaustively listed in the paper by Kaer and Speek (Callinan and Batzer , 2006; Kaer and Speek, 2013).

The authors describe 21 disease-causing mutations deriving from L1 retrotransposition (Table 1): fourteen are insertions into exons, one into the 3'UTR and six into introns. The majority of the exonic insertions are in the sense orientation, but actually any orientation would be effective since in this case the insertion causes a disruption of the coding sequence (Brouha et al., 2002; Holmes et al., 1994; Kazazian et al., 1988; Kimberland et al., 1999; Kondo-Iida et al., 1999, Li et al., 2001; Miki et al., 1992; Mukherjee et al., 2004; Schwahn et al., 1998; Yoshida et al., 1998). For most of these insertions the exact mechanism of interference with transcription is not clear, but six of them were demonstrated to cause exon skipping and partial exonization (Awano et al., 2010; Bernard et al., 2009; Narita et al., 1993; van den Hurk et al., 2003; Wimmer et al., 2011).

Concerning the intronic insertions, also in this case not all the mutational mechanisms are clear. Three of them were demonstrated to interfere with the recognition of the polypyrimidine or 3′ss signal causing exon skipping (Kondo-Iida et al., 1999; Martinez-Garay et al., 2003; Meischl et al., 2000), while two of them, involving the insertion of a full-length L1 element in the sense orientation, seemed to alter transcription or mRNA instability in one case, and to cause L1 exonization in the other case (Schwahn et al., 1998; Samuelov et al., 2011; Kaer and Speek, 2013).

Besides human diseases caused by the direct effect of L1 retrotransposition, some neurological diseases, such as Rett syndrome and ataxia telangiectasia have been recently demonstrated to misregulate L1 retrotransposition, probably contributing to some aspects of the diseases (Thomas et al., 2012).

Rett syndrome (RTT) is a neurodevelopmental disorder caused by mutations in the MeCP2 gene that typically affects girls, with symptoms such as autism, loss of speech, hand-wringing, anxiety, and motor deterioration (Amir et al., 1999). It is still not clear the function of MeCP2 in neurons and its precise contribution in the pathogenesis of Rett syndrome, but it has been recently demonstrated that MeCP2 is involved in L1 regulation (Muotri et al., 2010; Skene et al., 2010, Yu et al., 2001). As previously mentioned, MeCP2 is involved in global DNA methylation and particularly, in neural stem cells, it can combine to the methylated L1 promoter, where it forms a repressive complex with HDAC1 that is able to block L1 expression (Muotri et al. 2010; Coufal et al., 2009). Moreover, in RTT patients it has been observed a hypomethylation of L1 promoter, a higher L1 retrotransposition and a higher L1 copy number compared to controls (Muotri et al. 2010). Whether the higher L1 retrotransposition is really

involved in RTT pathogenesis is still unclear, but the increased genomic diversity induced by L1 insertions may contribute to the heterogeneity of the disease (Thomas et al., 2012).

Table 1: L1 insertions causing human diseases. Adapted from Kaer and Speek, 2013.

Disease	Affected gene	Chr	Insertion position	Strand	Mechanism
Hemophilia A	Coagulation factor VIII (F8)	X	Exon 14	Antisense	ND
Hemophilia A	Coagulation factor VIII (F8)	X	Exon 14	Sense	ND
Hemophilia B	Coagulation factor IX gene (F9)	X	Exon 5	Sense	ND
Hemophilia B	Coagulation factor IX gene (F9)	X	Exon 7	Sense	ND
Familial polyposis coli (colon cancer)	Adenomatous polyposis coli (APC)	5	Exon 16	Sense	ND
Duchenne muscular dystrophy (DMD)	Dystrophin (DMD)	X	Exon 44	Antisense	Exon 44 skipping
Duchenne muscular dystrophy (DMD)	Dystrophin (DMD)	X	Exon 48	Sense	Exon 48 skipping?
Duchenne muscular dystrophy (DMD)	Dystrophin (DMD)	X	Exon 67	Antisense	Exon 67 skipping
X-linked dilated cardiomyopathy (XLDCM)	Dystrophin (DMD)	X	Exon 1	Antisense	ND
X-linked retinitis pigmentosa (XLRP)	Retinitis pigmentosa 2 (RP2)	X	Intron 1	Sense	ND
Beta-thalassemia	Hemoglobin, beta (HBB)	11	Intron 2	Antisense	ND
Fukuyama-type congenital muscular dystrophy (FCMD)	Fukutin (FKTN)	9	Intron 7	Sense	Exon 7-8 skipping
Fukuyama-type congenital muscular dystrophy (FCMD)	Fukutin (FKTN)	9	3'UTR?	ND	ND
Chronic granulomatous disease (CGD)	Cytochrome b-245, beta polypeptide (CYBB)	X	Intron 5	Sense	Combinations of exon skipping and L1 exonization
Chronic granulomatous disease (CGD)	Cytochrome b-245, beta polypeptide (CYBB)	X	Exon 4	Sense	ND
Choroideremia (CHM)	Choroideremia (Rab escort protein 1) (CHM)	X	Exon 6	Antisense	Exon 6 skipping
Coffin-Lowry syndrome (CLS)	Ribosomal protein S6 kinase (RPS6KA3)	X	Intron 3	Antisense	Exon 4 skipping
Ataxia with oculomotor apraxia 2 (AOA2)	Senataxin (STX)	9	Exon 12	(Anti)sense	Partial exon skipping or exonization
Chanarin-Dorfman syndrome (CDS)	Abhydrolase domain containing 5 (ABHD5)	3	Intron 3	Sense	Exonization
Neurofibromatosis type 1 (NF1)	Neurofibromin 1 (NF1)	17	Exon 23	Sense	Exon 23 skipping
Neurofibromatosis type 1 (NF1)	Neurofibromin 1 (NF1)	17	Exon 39	Sense	Partial exonization

Ataxia telangiectasia (A-T) is a rare, autosomal recessive neurodegenerative disease caused by mutations in the Ataxia Telangiectasia Mutated (ATM) gene. Common aspects of the disease are loss of motor function, dilatation of blood vessels, immunodeficiency and a series of severe complications that lead to premature death. ATM is a serine/threonine protein kinase involved in the detection and response to DNA double strand breaks, through the blocking of the cell cycle until the damage is repaired. In case of ATM deficiency, double-strand breaks are not repaired and cells accumulate genomic mutations (Bar-Shira et al., 2002).

Coufal and colleagues in 2011 demonstrated that the brain of both ATM ko mice and A-T affected patients present higher levels of L1 retrotransposition and longer L1 insertions compared to the controls. They speculated that normally ATM might detect

the activity of L1 ORF2p and inhibit L1 insertion, while a dysfunctional ATM could prevent an effective DNA damage response during L1 integration, eventually leading to neurodegeneration (Coufal et al., 2011; Thomas et al., 2012).

Finally, it has been recently reported an altered L1 retrotransposition in schizophrenia. In particular, Bundo and colleagues demonstrated the presence of an increased L1 copy number in neurons from the prefrontal cortex of affected patients and in induced pluripotent stem cells-derived neurons containing the 22q11 deletion (one of the highest risk factors for schizophrenia). Taking advantage of whole-genome sequencing, they detected in patients brain-specific L1 insertions, localized preferentially in synapse- and schizophrenia-related genes. Further experiments on animal models aimed at identifying the causes of this L1 copy number alteration, suggested that the high neural L1 retrotransposition may be triggered by environmental and/or genetic risk factors, possibly contributing to the susceptibility and some aspects of the disease (Bundo et al., 2014).

Even if not directly caused by L1 insertion, several genetic diseases have been demonstrated to be linked to L1 because originating from the insertion of Alu sequences (56 cases reported) or SVA (7 cases reported), both non autonomous retrotransposons that take advantage of L1's enzymatic machinery to mobilize (Kaer and Speek, 2013).

Strategies for mapping L1 insertions

The majority of L1 elements present in the human and mouse genomes are immobile because of the mutations and truncations accumulated during evolution, but a small niche of them is still able to multiply throughout the genome, and that's why several strategies aimed at identifying their precise insertion sites have been developed in the last years (O'Donnell and Burns, 2010).

The first assays to be developed were based mainly on PCR, followed by gel separation of the amplicon, as the amplification typing of L1 active subfamilies (ATLAS) (Badge et al., 2003), the L1 display (Sheen et al., 2000), and the L1 insertion dimorphisms identification by PCR (LIDSIP) (Pornthanakasem and Mutirangura, 2004).

These techniques allowed a first investigation of the massive L1 polymorphism present in the human genome, but they were not suitable for L1 mapping in large numbers of samples (O'Donnell and Burns, 2010).

Beck and colleagues applied a fosmid based sequencing strategy to study active full length L1 elements in the human genome with the advantages to be independent from

PCR fidelity, compared to PCR-mediated techniques, and to detect large insertions and deletions in repetitive regions (Korbel et al., 2007; Kidd et al., 2008; Beck et al., 2010).
Recent technological advances allowed the development of several different methods that, exploiting high resolution microarrays or deep sequencing, allow the identification of mobile elements' insertion sites on a genome-wide scale (Gabriel et al., 2006; Iskow et al., 2010; Ewing et al., 2010; Witherspoon et al., 2010)
For instance, in 2010 Huang and colleagues mapped L1 insertions in the human genome using a ligation mediated PCR method called vectorette PCR: in this technique, synthetic adapters are ligated to fragmented DNA, followed by the PCR amplification of the genomic DNA flanking the mobile element using primers specific for the adapter and the L1 sequence. The insertions are then identified by labelling and hybridizing of the amplicons to genomic tiling microarrays or by deep sequencing (Huang et al., 2010).
One year later, Baillie and colleagues published the description of a new high-throughput protocol that they called retrotransposon capture sequencing (RC-seq). In this method, fragmented genomic DNA is hybridized to arrays targeting the 5′ and 3′ termini of mobile elements, sequenced and mapped using a computational pipeline aimed at identifying known and novel insertions (Baillie et al., 2011).

Drosophila melanogaster's *roo* element

About 10% of the *Drosophila melanogaster*'s genome is composed of transposable elements, with >1500 copies per gamete in the euchromatic part of the genome and ~2000 in its heterochromatic part (Maside et al., 2001; Kaminker et al., 2002).
With the publication of the Release 3 genomic sequence of *D. melanogaster* in 2002 it was reported for the first time an analysis from several points of view of the transposable elements present in the euchromatic part of the genome: indeed it was possible to identify 96 families of transposable elements: 49 LTR retrotransposon families, 27 LINE-like families, 19 DNA transposon families and the *Foldback* (FB) family (Celniker et al., 2002; Kaminker et al., 2002).
The largest family of the LTR retrotransposons' class is represented by the *roo* element, with a total of 541 copies (full length and LTR fragments) (de la Chaux and Wagner, 2009). The most important difference between *roo* and the other LTR elements stands in the fact that it encodes for an envelope (*env*) gene (probably non-functional) in addition to the *gag* and *pol* genes encoded by all the other LTR retrotransposons (Frame

et al., 2001). The *roo* element is 9092 bp long, it contains one ORF with 7083 nucleotides that comprise the *gag*, *pol* and *env* genes, it's strongly expressed during embryogenesis and it can induce gene mutations (Brunner et al., 1999) (Figure 14).

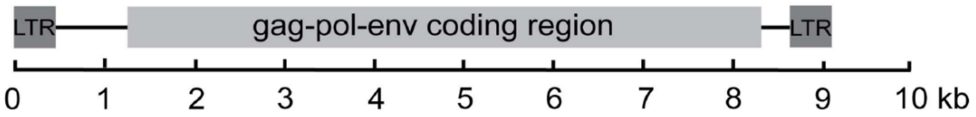


Figure 14: Drosophila roo element structure. The canonical roo element is 9092 bp long, has one 7083 bp long ORF (light gray box), and is flanked by 428/429 bp long LTR sequences (dark gray boxes). The ORF encodes the gag, pol and env genes (de la Chaux and Wagner, 2009).

From a comparison of *roo* distribution between twelve different *Drosophila* species, arose that in *Drosophila melanogaster*, the one with the highest number of *roo* element, the majority of these elements are young, with more than half harboring identical LTR sequences, indicating very recent insertion (de la Chaux and Wagner, 2009).

Alzheimer's disease

Alzheimer's disease is the leading cause of dementia in the elderly and leads to death within 3 to 10 years after appearance of symptoms (Bettens et al., 2013; Isik, 2010). Firstly described by the German neuropathologist Alois Alzheimer in 1907, it is a progressive neurodegenerative disease, clinically characterized by memory loss, cognitive deterioration, development of psychiatric and behavioral disorder, and impairment of activities of daily life (Hong-Qi et al., 2012).

This condition of severe dementia can have many consequences, such as immobility, swallowing disorders and malnutrition that can increase the risk of developing pneumonia, the most frequent cause of death among elderly individuals affected by AD (Brunnström and Englund, 2009; Kalia, 2003). Neuropathological hallmarks of the disease are amyloid plaques and neurofibrillary tangles, together with synaptic and neuronal loss as well as astrocytic gliosis (Chin, 2011).

The last twenty years of research and more than 1 billion of US dollars spent for clinical trials have actually failed to yield an effective drug treatment, able to treat or prevent AD (Kosik, 2013).

Epidemiology

Worldwide, approximately 35.5 million people are estimated living with dementia to date, and according to the World Health Organization this number is going to double by 2030 and triple by 2050. About 70% of these cases can be attributed to Alzheimer's disease. Among 60 years old populations, it seems that in North America and Western Europe there is the highest prevalence and incidence of dementia, followed by Latin America and Asia, with the incidence rate increasing exponentially with age. Similar patterns of prevalence and incidence can be observed also for AD (Ferri et al., 2005; Matthews et al., 2013; Reitz and Mayeux 2014) (Figure 15).

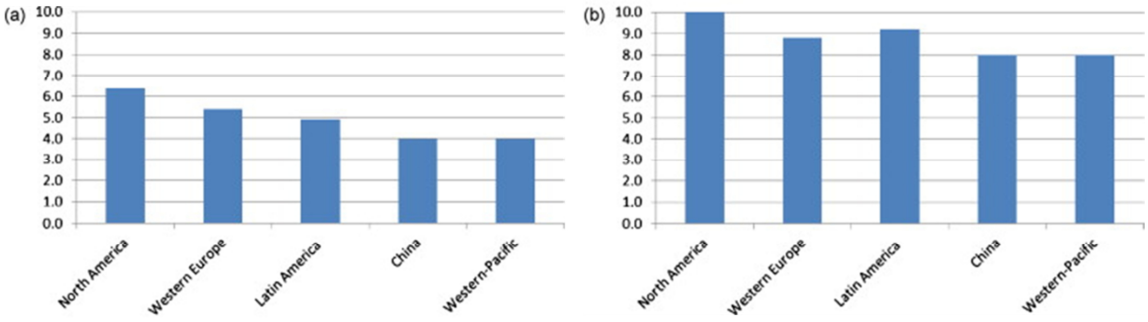


Figure 15: (a) Global prevalence of dementia (%). (b) Incidence rates (per 1000 individuals in the population). Adapted from Reitz and Mayeux 2014.

Indeed, besides the global ageing, the main negative consequence deriving from the improved healthcare over the last century is for sure the higher incidence of AD and other related dementias (from the World Alzheimer Report 2009).

Only in the USA, payments for health care, long-term care, and hospice for people affected by AD are projected to increase from $203 billion in 2013 to $1.2 trillion in 2050 (in 2013 dollars) (Thies et al., 2013).

Clinical aspects of AD

AD is commonly classified according to its age of onset. A small subset of AD patients show an early onset (in their late 40s or early 50s), which is called early-onset familial AD (EO-FAD), characterized by an aggressive progression and short relative survival time,but more than 95% of all the AD patients actually develop the disease when they are aged >65 years, which is called late-onset AD (LOAD). The two forms of the disease present a different pattern of genetic epidemiology (Jiang et al., 2013; Panegyres and Chen, 2013).

AD can affect people in different ways. In the majority of cases the onset of the disease corresponds to a gradual loss of memory, in particular of recent informations, just because the first neurons to degenerate are those involved in the creation of new memories. As soon as other brain regions involved in different functions die, the affected indivuals start suffering from other symptoms.

The following are common symptoms of AD, listed in the US 2014 Alzheimer's Disease Facts and Figures:

- Memory loss that disrupts daily life
- Challenges in planning or solving problems
- Difficulty completing familiar tasks at home, at work, or at leisure

- Confusion with time or place
- Trouble understanding visual images and spatial relationships
- New problems with words in speaking or writing
- Misplacing things and losing the ability to retrace steps
- Decreased or poor judgment
- Withdrawal from work or social activities
- Changes in mood and personality

The progression's rate of the disease can be very different depending on the individual. As the disease advances, patients experience a progressive decline in their cognitive and practical capacities. When they reach the final stage of AD, patients are no longer able to communicate, recognize people and they are completely dependent on people that take care of them (Thies et al., 2013).

In 1984, the Alzheimer's Association and the National Institute of Neurological Disorders and Stroke established for the first time a common and homogeneous list of criteria in order to make the diagnosis of this complex disease as much consistent as possible (McKhann et al., 1984).

In 2011, the NIA and the Alzheimer's Association proposed new criteria and guidelines that updated those published in 1984, and classified AD in three stages (preclinical phase, symptomatic predementia phase and dementia phase) characterized by a continuum between and within each stage. The new guidelines introduced the use of brain imaging and cerebrospinal fluid biomarkers as means for clinicians to diagnose AD as soon as possible in its course to allow prompt treatments in order to prevent further neuronal damages (Jiang et al., 2013; Thies et al., 2013). In 2012, the NIA and the Alzheimer's Association introduced also new guidelines to help pathologists in describing and categorizing brain changes associated with AD and other dementias (Hyman et al., 2012).

None of the pharmacological treatments that are available today for AD can slow or stop the death and malfunctioning of neurons in the brain of AD patients. The U.S. Food and Drug Administration approved five drugs that are able to give relief from the symptoms of AD temporarily, by increasing the amount of acetylcholine or blocking the activity of the neurotransmitter glutamate in the brain. The effectiveness of these drugs is variable across the population (Thies et al., 2013).

Besides drugs, also nonpharmacological therapies, such as cognitive training and behavioral interventions can be used to improve quality of life or reduce behavioral symptoms such as depression, apathy, sleep disturbances, agitation, and aggression, although only few have been demonstrated to be truly effective (Olazarán et al., 2010; Thies et al., 2013).

Neuropathological aspects of AD

In most cases of Alzheimer's disease brain samples present a small degree of cerebral cortical atrophy, which involves mainly the frontotemporal association cortex, almost without an involvement of the primary motor, sensory, and visual areas. However, among elderly subjects, it is not possible to observe any difference in brain weight as well as cerebral cortical thickness between age-matched normal and AD affected individuals. On the contrary, when examining cases of early-onset familial AD (EO-FAD) compared with age-matched controls, it can be detected a clear difference in brain weight or the degree of cerebral cortical atrophy (Perl, 2010).

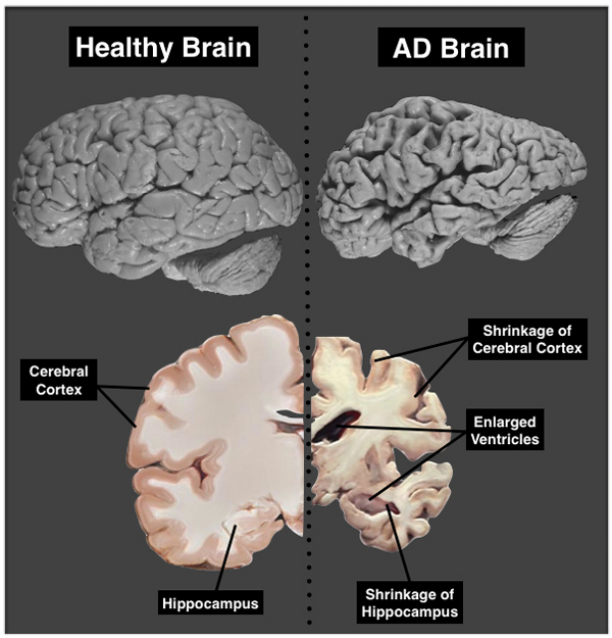


Figure 16: Comparison between a healthy and AD affected brain (from www.knowingneurons.com)

Moreover in AD brains is possible to observe a loss of tissue that generally leads to an expansion of the lateral ventricles and a significant atrophy of the hippocampus, with an enlargement of the adjacent temporal horn of the lateral ventricle (Perl, 2010) (Figure 16). It is also relatively common to find some cortical micro infarcts, lacunar infarcts in

the basal ganglia, and demyelination of the periventricular white matter (Serrano-Pozo et al., 2011).

Actually, an exhaustive diagnosis of Alzheimer's disease can be made only upon the histological examination of the post mortem brain specimen, even if, virtually, any of the morphological abnormalities peculiar for AD can also be seen, to some extent, in the brains of elderly healthy individuals. That's why a precise diagnosis often requires the observation of the lesions' spreading in certain areas of the brain (Perl, 2010).

Neurofibrillary Tangles

Neurofibrillary tangles (NFTs) are considered a cardinal microscopic lesion of AD and are fundamental for making the pathological diagnosis. They appear as parallel fibrils surrounding the nucleus and spreading toward the apical dendrite (Figure 17). When a NFT occurs with a more rounded shape (for instance in neurons within the substantia nigra and locus ceruleus), it is called globoid neurofibrillary tangle (Perl, 2010).

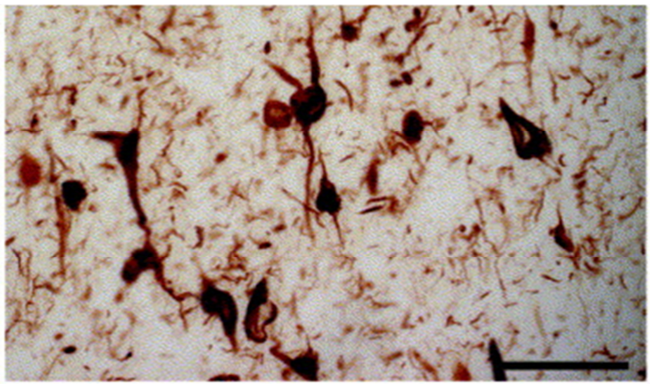


Figure 17: A representative microphotograph of neurofibrillary tangles. Tangles were visualized by immunostaining with an anti-PHF1 specific antibody. Scale bar: 62.5 µm (LaFerla and Oddo, 2005).

The main component of NFTs is the microtubule-associated protein tau. Tau protein is normally present in the axon where it stabilizes the microtubules, and its binding to tubulin is regulated by its phosphorylation state that is determined by the action of kinases and phosphatases (Mandell and Banker, 1996).

In pathological conditions, abnormal modifications such as hyper-phosphorylation and acetylation of tau protein decrease its binding to tubulin and this, in turn, causes the self-aggregation of tau into insoluble filaments, which form the NFTs (Iqbal and Grundke-Iqbal, 2008; Iqbal et al., 2010). There are several other protein constituents associated with neurofibrillary tangles, such as ubiquitin, cholinesterases, and beta-amyloid 4 (βA4), but tau seems to be the crucial component of most of these structures

(Mohamed et al., 2013; Perl, 2010). To date, no mutations in the tau gene were found in patients affected by AD. However, it has been recently suggested that tau polymorphisms might be considered new risk factors for AD (Gerrish et al., 2012).

Studies have shown that the extent and distribution of neurofibrillary tangles follow a reproducible pattern and they correlate with both the degree of dementia and the duration of illness, suggesting that these abnormalities must have a direct impact on the normal functionality of the brain (Bierer et al., 1995; Braak and Braak, 1991).

Moreover, according to a recent hypothesis, the primary deposition of tau may have a role in the formation, release and accumulation of Aβ in the form of extracellular plaque (in contrast with the *amyloid cascade hypothesis* mentioned below) (Braak and Del Tredici, 2013).

According to the spreading of NFTs in the brain, Braak established in the 90s a staging classification of sporadic AD. As clearly reported by Ferrer in 2012, "stages I and II are defined by the presence of NFTs in the entorhinal cortex (stage I) and progression to the transentorhinal cortex and mild involvement of the CA1 region (stage II). Limbic stages III and IV implicate the presence of NFTs in the upper and inner layers of the entorhinal cortex, transentorhinal cortex, CA1 region of the hippocampus, subiculum, anterodorsal thalamic nucleus, amygdala, magnocellular nuclei of the basal forebrain (including Meynert nucleus), tubero-mammillary nucleus (stage III), plus associated areas of the temporal cortex, striatal neurons, raphe nucleus and locus ceruleus (stage IV). Neocortical stages V and VI require, in addition, NFTs in cortical association areas, claustrum, reticular nucleus of the thalamus and substantia nigra (stage V), plus primary sensory areas (stage VI)" (Braak and Braak, 1991; Ferrer, 2012).

NFTs have been observed at the level of entorhinal and transentorhinal cortices in 70–80% of individuals 65 years old, belonging to a non-biased general population, and dying for reasons not related to neurological diseases. This percentage has been demonstrated to reach the 90% in 80 years old individuals, and to remain steady in centenarians (Braak and Braak, 1997; Braak et al., 2011; Ferrer, 2012). Individuals at the I-II stages don't present any cognitive impairment, while neurological symptoms as early memory deficits and mild cognitive impairment occur at stages III-IV, whereas dementia does not appear until stage V-VI, when the brain is filled with NFTs and senile plaques (Braak et al., 2006; Ferrer, 2012).

Although NFTs are considered a cardinal histopathological feature of Alzheimer's disease, these lesions can be found in association with many other diseases, such as

postencephalitic parkinsonism, posttraumatic dementia and amyotrophic lateral sclerosis/parkinsonism-dementia complex of Guam (Perl, 2010).

Amyloid plaques

Amyloid plaques, deriving from the deposition and aggregation of extracellular Aβ peptide in the brain, are considered as the major neuropathological hallmarks of AD (Kim et al., 2014).

Aβ peptides with 40 or 42 aminoacids (Aβ40 or Aβ42) derive from the cleavage of the amyloid precursor protein (APP), a transmembrane protein containing a large extracellular domain, a hydrophobic transmembrane domain and a short intracellular domain (Serrano-Pozo et al., 2011).

APP polypeptides can be different in terms of post-translational modifications and length. Indeed alternative splicing produces three major isoforms of APP (695, 751 and 770 residues) which are differentially expressed: the 751 and 770 isoforms are expressed in both non-neuronal and neural cells, whereas the 695 isoform is highly expressed in neurons and less expressed in non-neuronal cells (Cavallucci et al., 2012).

The APP protein goes through several proteolytic cleavages performed by enzyme complexes with α-, β- and γ-secretase activity, that cause the formation of large soluble secreted fragments and membrane-associated C-terminal fragments (CTF) (Figure 18).

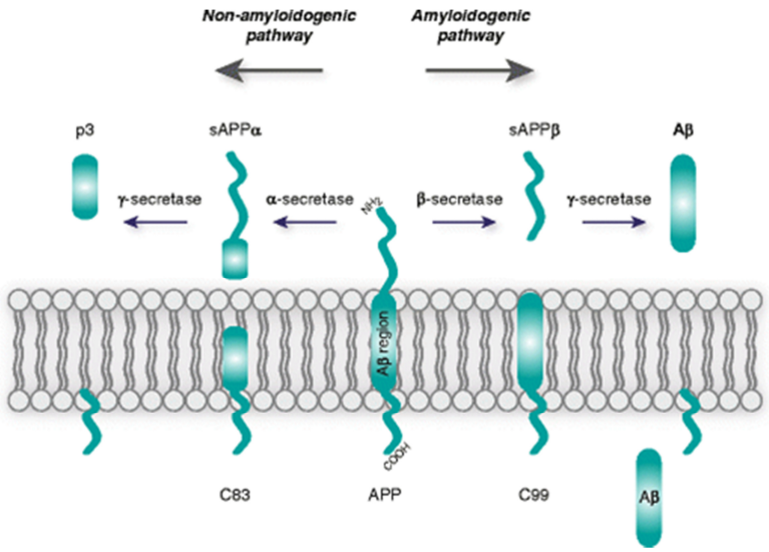


Figure 18: Schematic representation of APP processing. The APP transmembrane protein is cleaved sequentially by means two distinct pathways: the nonamyloidogenic pathway (left) and the amyloidogenic pathway (right). The initial cleavage of APP by α-secretase in the region containing Aβ sequence prevents the formation of Aβ peptide. By contrast, the alternative cleavage carried out by β-secretase leads to the release of Aβ peptide from the following γ-secretase cleavage (Cavallucci et al., 2012).

First of all APP protein can be processed through the prevalent non-amyloidogenic pathway, in which the α-secretase cleaves the APP protein leading to the formation of a membrane-retained CTF composed of 83 residues and a large N-terminal soluble fragment which is released in the extracellular space. After this first step, the CTF fragment is cleaved by the γ-secretase, with the production of a short fragment called p3. At the end of this process no Aβ peptides are produced.

On the other hand, in the amyloidogenic pathway the first cleavage of the APP protein is performed by the β-secretase, with the production of a membrane-retained CTF of 99 amino acids (C99), and a soluble fragment released in the extracellular space. The cleavage of C99 by the γ-secretase leads to the release of the Aβ peptide.

The majority of Aβ peptides have 40 residues (Aβ40), but also a longer form of 42 residues (Aβ42) can be produced. The Aβ42 peptide is more hydrophobic and prone to aggregation than the Aβ40 variant, and it has been demonstrated to be the most abundant form present in the amyloid plaques (Cavallucci et al., 2012).

Both Aβ40 and Aβ42 peptides can be observed at low levels in the healthy brain, where they seem to play physiological functions such as maintaining the synaptic function and plasticity (Kamenetz et al., 2003).

Aβ peptides tend to accumulate because an imbalance between Aβ production, aggregation and clearance occurs.

Early-onset familial forms of AD usually present an increased Aβ production and aggregation, while sporadic AD forms are particularly characterized by a reduced clearance of Aβ (Crews and Masliah, 2010).

There are different physical conformations of Aβ aggregates, such as soluble oligomers (oAβ), protofibrils (pfAβ), nonfibrillar insoluble oligomers, and fibrillar (fAβ) forms. Furthermore, amyloid plaques can be of two main types: diffuse and compact. Diffuse plaques contain expecially nonfibrillar insoluble oligomers of Aβ, while compact plaques are composed of fAβ and have been demonstrated to contain a larger numbers of microglial cells and astrocytes than adjacent brain regions. They also contain abnormal neurites (neuritic plaques) with abnormally phosphorylated tau. Moreover it has been demonstrated that there is a correlation between the density of neuritic plaques and the severity of dementia (Shah et al., 2010) (Figure 19).

Aβ peptides and Aβ amyloid plaques typically are located outside the cell, but it seems that also intracellular Aβ can play an important role in AD (Friedrich et al., 2010).

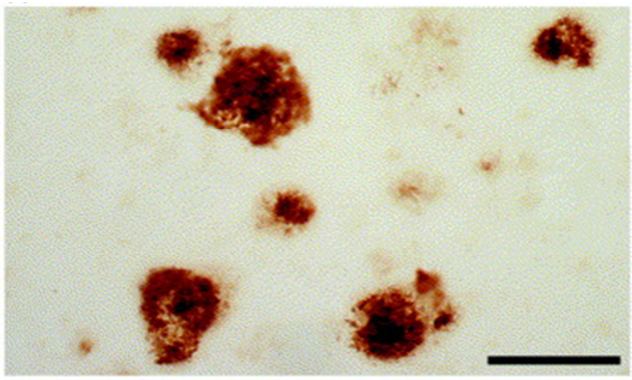


Figure 19: A representative microphotograph of amyloid plaques in the AD brain. Amyloid plaques were visualized by immunostaining with an anti-Aβ42 specific antibody. Scale bar: 125 µm (LaFerla and Oddo, 2005).

Although the *amyloid cascade hypothesis* (that places Aβ deposition at the top of a sequence of events that eventually leads to AD) is still controversial and it is not yet established which form of Aβ is the pathogenic one, in vitro and in vivo evidences indicate that fAβ exerts powerful toxic effects on neurons, contributing directly to oxidative stress, mitochondrial dysfunction, impaired synaptic transmission, the disruption of membrane integrity, and impaired axonal transport (Crouch et al., 2008; LaFerla and Oddo 2005; Shah et al., 2010). Furthermore the fact that the first mutations demonstrated to cause inherited forms of familial AD were identified in the APP gene, provides further evidence that APP plays a central role in AD pathogenesis (Galimberti and Scarpini, 2012).

Genetics of Alzheimer's disease

As previously mentioned, Alzheimer's disease can be divided into early-onset familial AD (EO-FAD), affecting patients younger than 65 years, and late-onset (LOAD), affecting patients older than 65 years. Both the two forms of AD have a genetic component (Bettens et al., 2013). EO-FAD is most often caused by rare, fully penetrant mutations, and is usually characterized by Mendelian inheritance (Tanzi, 2012). LOAD, on the other hand, is caused by an interplay between genetic and environmental factors (defined as a disease with a complex genetic background) (Bettens et al., 2013).

Early-onset familial Alzheimer's disease (EO-FAD)

Mutations able to cause EO-FAD have been observed at the level of three different genes: amyloid beta precursor protein (APP), presenilin 1 (PSEN1), and presenilin 2 (PSEN2) (Table 2). Both clinical and pathological aspects of the disease can vary

according to the locus of the mutation and the precise position in the gene (Ridge et al., 2013).

Table 2: Early-onset familial Alzheimer's disease genes and their pathogenic effects (Adapted from Tanzi, 2012).

Gene	Protein	Chromosome	Mutations	Molecular phenotype
APP	Amyloid β (A4) protein precursor	21q21	24 (duplication)	Increased $A\beta_{42}/A\beta_{40}$ ratio Increased Aβ production Increased Aβ aggregation
PSEN1	Presenilin 1	14q24	185	Increased $A\beta_{42}/A\beta_{40}$ ratio
PSEN2	Presenilin 2	1q31	14	Increased $A\beta_{42}/A\beta_{40}$ ratio

In the case of APP gene, duplications (as in Down's syndrome patients) are sufficient in many cases to cause EO-FAD, due to increased Aβ42 production and deposition. Mutations in this gene account for 13–16% of all EO-FAD cases (Raux et al., 2005; Sleegers et al., 2006) (Figure 20).

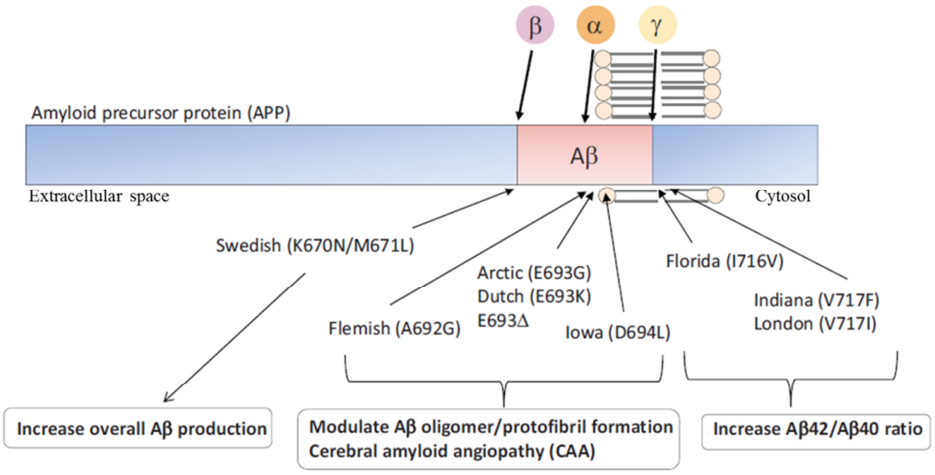


Figure 20: Schematic diagram of APP mutations and their amyloidogenic effect. APP is proteolytically cleaved by α-, β- and γ-secretase activities, and α-secretase cleavage precludes toxic Aβ to be produced. FAD-linked genetic mutations on APP exhibit different effects on APP processing and amyloidogenic properties of Aβ peptides (Kitazawa et al., 2012).

The Swedish, Arctic, and London variants are three different kinds of APP mutations that occur in different domains of the APP gene and lead to EO-FAD by different mechanisms. The Arctic mutation (inside the Aβ domain) is dominantly inherited and fully penetrant with an average age of onset of 57 years and leads to Aβ protofibril formation. The Swedish mutation is a double mutation that occurs upstream to the Aβ domain, causing an increase in total Aβ production and changes in its extracellular localization, while the London mutation occurs downstream to the Aβ domain, resulting in higher Aβ42 levels (Ridge et al., 2013).

PSEN1 gene is located in chromosome 14 (14q24.3) and has at least two isoforms. 185 mutations in this gene were identified as responsible for genetic forms of AD (Cruts and Van Broeckhoven, 1998). According to the precise position of the mutation, different forms of EO-FAD can occur, with differences in age at onset, rate of progression and severity (Heckmann et al., 2004). PSEN1 is part of the γ-secretase complex, meaning that mutations in this locus lead basically to an increase of Aβ42 production. If the mutation falls upstream to the protein position 200, it causes a pathology similar to sporadic AD, whereas mutations falling at subsequent positions cause a more severe amyloid angiopathy (Ryan and Rossor, 2010; Ridge et al., 2013).

PSEN2 gene is located in chromosome 1 (1q31-q42) and has two known isoforms. AD-causing mutations in this locus are less common compared to PSEN1 and the resulting pathology is less severe. There are 14 pathogenic mutations known to occur in this locus, and, as for PSEN2 mutations, they seem to increase Aβ42 production (Ridge et al., 2013).

Besides APP, PSEN1 and PSEN2, mutations in other three genes have been identified as possible causes of EO-FAD. An AD case occurred in a Belgian family has been linked to a missense mutation of the tau gene (Rademakers et al., 2003). Polymorphisms in the chromosome 7q36 have been reported to cause EO-FAD in a Dutch family, and also a missense mutation in the gene PEN2 have been found in an AD family (Rademakers et al., 2005; Sala Frigerio et al., 2005).

Late-onset Alzheimer's disease (LOAD)

Together with the old age, inherited genetic risk factors, exhaustively reviewd by Karch and Goate, seem to play a fundamental role in the pathogenesis of at least 80% of AD cases (Tanzi, 2012).

For many years, only one genetic risk factor, the APOE ε4 allele, has been implicated in late-onset and early-onset AD, but new technologies, such as large-scale genome-wide association studies (GWAS), that allow the analysis of millions of polymorphisms in thousands of subjects at a time, revealed new genes associated to LOAD risk (Bettens et al., 2013). These genes have been well described in the review by Karch and Goate published in 2014, and summarized below.

Apolipoprotein E (APOE) is the most important risk factor for LOAD. APOE gene is located on chromosome 19q13.2 and encodes for three alleles (ε2, ε3, ε4). The ε4 allele is correlated to an increased AD risk, in particular: one APOEε4 allele increases AD risk 3-fold, and two APOEε4 alleles increase AD risk by 12-fold, with a decrease in age

of onset. On the contrary, the ε2 allele is associated with a decreased risk and later age of onset (Karch and Goate, 2014).

APOE is involved in lipoprotein metabolism and plays several important roles in the central nervous system (Kim et al., 2009a). In particular, it is able to bind Aβ, altering the clearance of the soluble Aβ and its aggregation, and it can also regulate Aβ metabolism by interacting with the LRP1 receptor (Verghese et al., 2013). Indeed, APOEε4 allele carriers have been demonstrated to present a faster and higher Aβ deposition compared to APOEε4-negative individuals (Karch and Goate, 2014; Morris et al., 2010).

Since 2009, European and international genome-wide association collaborations allowed the discovery of at least nine new risk loci for AD, involved in lipid metabolism, inflammatory response and endocytosis (Karch and Goate, 2014) (Figure 21).

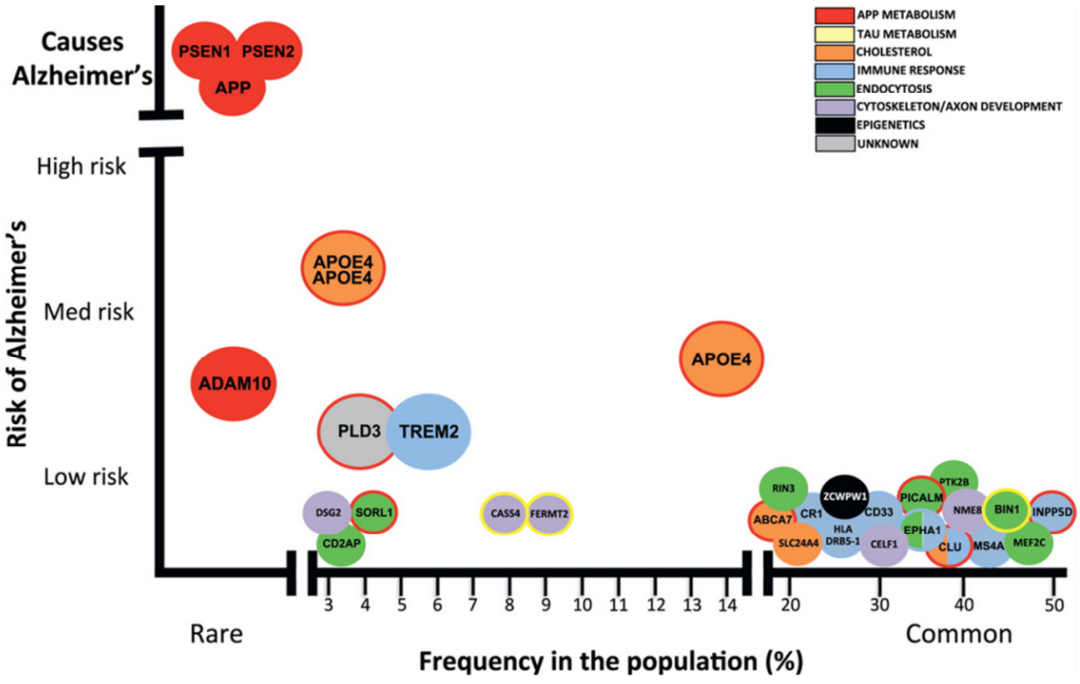


Figure 21: Rare and common variants contribute to Alzheimer's disease risk. GWAS, genome-wide association studies (Adapted from Karch and Goate, 2014).

Besides APOE, variants of other two genes involved in cholesterol metabolism were found to be associated to AD: CLU and ABCA7.

Clusterin (CLU) is an apolipoprotein, whose gene is located on chromosome 8p21.1, and it encodes three alternative transcripts. Single nucleotide polymorphisms (SNPs) at the level of this gene have been demonstrated to be protective against LOAD, or could

be linked to plasma clusterin levels (high levels of clusterin in the plasma have been associated with brain atrophy, disease severity, and disease progression). Clusterin expression has been observed to be higher in AD brains, where it can be detected at the level of amyloid plaques. Since purified clusterin is able to interact with Aβ influencing fibril formation in vitro, it could be involved in Aβ clearance, amyloid deposition, and neuritic toxicity (Calero et al., 2000; Rizzi et al., 2009; DeMattos et al., 2004; Karch and Goate, 2014).

The ATP-binding cassette transporter A7 (ABCA7) is a transporter across the cellular membrane of substrates like cholesterol and can inhibit Aβ secretion. ABCA7 gene is present in the chromosome 19, and two different alternative transcripts can derive, both expressed in the brain (Kim et al., 2008). GWAS have identified several SNPs, inside or near the ABCA7 gene, which can be significantly associated to LOAD as risk factors (Vasquez et al., 2013; Kim et al., 2013; Karch and Goate, 2014).

GWAS studies allowed the identification of gene variants associated to LOAD, involved also in neuroinflammation and dysregulation of the immune response, which are other two fundamental aspects of AD (Holtzman et al., 2011).

In particular, it has been demonstrated that several SNPs occurring in the gene encoding for the complement receptor 1 (*CR1*) are associated to AD (Liu and Niu, 2009). Moreover, CR1 can be present as high-expression or low-expression alleles, determining a different amount of CR1 copies per cell: the higher is CR1 protein expression, the higher is the clearance of immune complexes, with a resulting lower risk of developing AD (Krych-Goldberg et al., 2002; Rogers et al., 2006; Karch and Goate, 2014).

SNPs identified near to *CD33*, a transmembrane receptor expressed in myeloid cells and microglia, seem to reduce LOAD risk (Malik et al., 2013). When CD33 is activated by the binding of sialic acid it can inhibit monocytes. Since CD33 expression is increased in microglia, and Aβ fagocytosis seems to be inhibited in immortalized microglial cells expressing CD33, genetic variations of this gene may influence Aβ clearance and other pathways mediated by microglia in the brain (Griciuc et al., 2013; Karch and Goate, 2014).

The MS4A gene cluster is composed of three genes that have been demonstrated to be associated with the inflammatory response: MS4A4A, MS4A4E, and MS4A6E. SNPs localized near to these genes have been associated to both an increase and a decrease in LOAD risk (Karch et al., 2012; Karch and Goate, 2014).

Finally, TREM2 that encodes for a transmembrane receptor expressed in microglia, is involved in the regulation of phagocytosis and inflammation. Missense mutations occurring in this gene seem to increase LOAD risk (Bertram et al., 2013; Karch and Goate, 2014).

Also mutations in genes involved in endocytosis and synaptic functions have been demonstrated to be associated to LOAD risk: SNPs at the level of the bridging integrator 1 (BIN1) for example, that is involved in endocytosis, immune response, calcium homeostasis, tau processing and apoptosis, have been linked to an increased risk for LOAD (Ren et al., 2006; Chapuis et al., 2013; Karch and Goate, 2014).

SNPs present at the 5' of another gene, the phosphatidylinositol binding clathrin assembly protein (PICALM), mainly expressed in neurons and involved in APP trafficking, have been associate to a reduced LOAD risk, while few SNPs found in the gene encoding for the CD2-associated protein (CD2AP), that is involved in cytoskeletal reorganization and intracellular trafficking, are associated to an increased LOAD risk (Karch and Goate, 2014).

Also SNPs located near to the gene EPH Receptor A1 (*EPHA1*) and sortilin-related receptor L1 (SORL1), involved in intercellular signaling and vesicle trafficking respectively, have been associated to reduced LOAD risk (Martínez et al., 2005; Rogaeva et al., 2007). Even if through an unknown mechanism, also rare coding variants of the phospholipase D3 gene (PLD3) seem to confer risk for LOAD (Cruchaga et al., 2014; Karch and Goate, 2014).

Recently two rare mutations in the ADAM10 gene were reported, able to cause AD at an average age of 70 years in seven out of 1000 LOAD families tested (Kim et al., 2009b). ADAM10 encodes for the main neuronal α-secretase, which cleaves APP protein avoiding the production of β-amyloid: the two mutations have been demonstrated to impair the α-secretase activity, with a resulting β-amyloid accumulation (Kim et al., 2009b; Tanzi, 2012).

All these polymorphisms identified by the GWAS studies usually occur in >5% of the population, but their effect is quite small, with an increased or decreased risk of ~0.10- to 0.15-fold, considering that the risk associated to the presence of the APOEε4 allele is increased a four- to 15-fold. This means that probably a major part of the genetic features determining LOAD are still unknown and can not be explained with the currently known susceptibility genes (Tanzi, 2012).

Epigenetics of Alzheimer's disease

Recent studies performed on human patients and animal models have showed that epigenetic mechanisms, which determine how and when genes are expressed, without altering the genetic code, contribute to AD (Tsankova et al., 2007).

Indeed many chronic neurological and psychiatric diseases may have at least a partial epigenetic aetiology (Gräff and Mansuy, 2008). In the specific case of AD, if we consider that genetic aspects, as previously mentioned, do not fully explain its pathogenesis, epigenetic modifications or environmental factors could likely contribute to some aspects of the disease. Epigenetic changes may indeed help to explain, for instance, why some family members develop the disease while others do not (Fraga et al., 2005). Evidences suggest that the interplay between genes and environment have an important role in the pathophysiology of different types of dementias through epigenetic mechanisms (Chouliaras et al., 2010). The best understood epigenetic mechanisms through which genome and environment influence the phenotype are DNA methylation and histone modifications (Alagiakrishnan et al., 2012).

DNA methylation

DNA methylation, which consists of an addition of a methyl group to the carbon 5 of the pyrimidine ring in the base cytosine, typically occurs in the CpG dinucleotides. In the human genome, there are regions, the CpG islands, where clusters of CpG dinucleotides are found concentrated, usually in or close to gene promoter regions. Methylation of cytosines in these regions avoids the binding of transcription factors and therefore gene transcription (Alagiakrishnan et al., 2012).

The DNA methyltransferases (DNMT) enzymes catalyze the reactions of *de novo* DNA methylation and maintenance of methylation marks (Fitzsimos et al., 2014).

Initial epigenetic studies on AD focused their attention on DNA methylation at the level of the APP gene, but studies made by different groups led to conflicting results: West and colleagues, for instance, observed hypomethylation of the APP gene promoter in an AD patient, whereas Barrachina and colleagues did not find any AD-related abnormalities in methylation at the level of this gene (West et al., 1995; Barrachina and Ferrer, 2009). While Tohgi and colleagues found an age-related decrease in promoter methylation of the APP gene in the human cerebral cortex (Tohgi et a., 1999; Fitzsimos et al., 2014).

It is still unclear whether APP gene methylation is specifically altered in AD or not, but strong evidence suggests the presence of a general altered DNA methylation in AD (Fitzsimos et al., 2014). Indeed studies have shown the presence, in the AD brain, of a severe reduction in S-adenosylmethionine (SAM) concentration, a methyl donor crucial for DNMTs activity, and in global DNA methylation (Morrison et al., 1996; Mastroeni et al., 2008). Additional studies have found reduced 5-mC levels in APP/PS1 transgenic mice and in the hippocampus, entorhinal cortex and cerebellum of AD patients (Chouliaras et al., 2013). Surprisingly, Aβ itself has been shown to affect DNA methylation (Chen et al., 2009; Fitzsimos et al., 2014).

Although the consequences of an AD-related altered DNA methylation are still unknown, some affected genes have been identified: Scarpa and colleagues, for instance, showed that PS1 was hypomethylated. Since the protein encoded by PS1 is involved in Aβ production, increased PS1 expression may enhance Aβ formation (Scarpa et al., 2003; Fitzsimos et al., 2014).

Histone modification

Histones can undergo many different posttranslational modifications: acetylation, methylation, phosphorylation, ADP-ribosylation, ubiquitylation or SUMOylation of specific amino acid residues of N-terminal histone tails. Since histones are involved in DNA packaging, these modifications can influence DNA structure, by favouring either an open, euchromatin state resulting in transcriptional activation or a closed, heterochromatin state resulting in gene silencing. Depending on which histone residue is modified, the same post-translational modification can have different effects, but in general, however, histone acetylation activates gene transcription (Alagiakrishnan et al., 2012). Besides DNA methylation, it has been suggested that also alterations in histone acetylation may be involved in AD pathogenesis. Indeed it has been demonstrated that aberrant histone acetylation levels are present in animal models of AD (Gräff et al., 2011). Peleg and colleagues noticed an association between alterations in gene expression and abnormal H4 acetylation and impaired memory function in fear conditioning in aged mice. Interestingly, these deficits were compensated by the administration of HDAC (histone deacetylase) inhibitors in the hippocampus (Peleg et al., 2010). More recent studies have indicated that HDAC2, involved in the regulation of memory and synaptic plasticity, might be directly implicated. Using CK-p25 mice as model for AD-like neurodegeneration, Gräff and colleagues found a significant increase of HDAC2 in the hippocampus and prefrontal cortex of these mice with a parallel

hypoacetylation of H2bK5, H3K14, H4K5 and H4K12. Importantly, hypoacetylation negatively correlated with mRNA expression of genes related to learning, memory and synaptic plasticity. Finally, Gräff and colleagues validated their findings in postmortem human brain samples deriving from patients affected by sporadic AD at different Braak stages. With these experiments they demonstrated that HDAC2 is significantly increased in the hippocampus and entorhinal cortex of AD patients in any Braak stage, including I and II, suggesting that an altered HDAC2 activity might be one of the earlier events during the pathogenesis of AD (Gräff et al., 2012).

The TgCRND8 mouse model of AD

In the last years researchers have created several transgenic mouse models overexpressing APP and/or presenilin with one or more mutations linked to familial AD. By using these models it has been possible to recapitulate and study many of the features of AD, such as myloid neuropathology, cerebral amyloid angiopathy, synaptic loss, dystrophic neurites, reactive gliosis and impairments in synaptic plasticity, learning and memory. Each mouse model exhibits these characteristics to variable extents, and they can be useful means to investigate the roles of APP, Aβ, and amyloid pathology in the pathogenesis of AD (Chin, 2011). APP transgenic mice present many of the key features of AD except for neurofibrillary tangles (NFTs). Even if no mutations in tau have been linked to AD, in order to recapitulate both the main neuropathological hallmarks of AD, a lot of different transgenic mice expressing combinations of mutant APP, presenilin, and tau have been created so far (Ballatore et al., 2007).

At present, none of the current mouse models of AD fully recapitulate the complex neuropathology of the human AD brain, but they are still usefull in the study of proteins, pathologies, and lesions involved in the pathogenesis of AD (Chin, 2011; Radde et al., 2008).

The TgCRND8 mouse is an early-onset transgenic mouse model of Alzheimer's disease. It encodes a double mutant form of amyloid precursor protein 695 (Swedish + Indiana) under the control of the PrP gene promoter. The Swedish double mutation increases the affinity of APP for the β-secretase, favouring the amyloidogenic pathway and increasing Aβ production; the Indiana mutation increases the processing of Aβ42 in relation to Aβ40 by the γ-secretase (Figure 20).

Cerebral Aβ amyloid deposits can be detected at 3 months of age in the hippocampus and neocortex as in other mouse models, and at 5 months is possible to observe the presence of dense-cored plaques and neuritic pathology. By 8 months of age, amyloid deposition spreads also to the cerebellum and brainstem. At 3 months, mice already present an impairment in acquisition and learning reversal in the reference memory version of the Morris water maze that can be related to the high production of Aβ42 at the cerebral level (Chishti et al., 2001). Gliosis and neuritic dystrophy are evident at 5 months, while vascular deposition can be observed around 6 months. The mortality of these mice is extremely precocious, indeed only approximately 50% of them reach 9-10 months of age (Chin, 2011).

Materials and methods

Tissue samples

The human genomic DNA samples used in the copy number variation (CNV) experiments were extracted from tissues of three different cohorts of patients. The Italian cohort comprised samples of temporal cortex taken from 11 patients affected by AD, and 13 healthy controls. These samples were provided by Prof. Tagliavini (Istituto Neurologico Carlo Besta, Milan).

The Spanish cohort, received from Prof. Isidro Ferrer (Bellvitge Neuropathology Institute, Barcelona), comprised samples of frontal cortex from 10 AD patients at the final Braak stages V-VI (severe AD), 10 patients at Braak stages I-II (mild AD), and 7 healthy controls.

The Brazilian cohort, provided by Prof. Lea Grinberg (Brain Bank of Sao Paulo), comprised samples of frontal cortex, temporal cortex, hippocampus, cerebellum and an extra-nervous tissue: the kidney, taken from 10 AD patients at Braak stages IV-VI and 10 controls at Braak stages 0-II.

Clinical investigations were conducted according to the principles of the Declaration of Helsinki. All the material was anonymous and the specimens were coded by numbers.

The TgCRND8 mice samples were provided by Prof. Scarpa and Dr. Fuso (Sapienza University, Rome). In particular we received hippocampus, cortex and kidney of 32 P0 mice (16 WT + 16TgCRND8), 24 adult mice at 3 months of age (12 WT + 12 TgCRND8), and 10 adult mice at 8 months of age (6 WT + 4 TgCRND8). Wild type and transgenic mice had a 129sv genetic background.

Genomic DNA extraction

Dissected tissues (almost 30 mg) were homogenized at room temperature in 2mL of lysis buffer (Tris pH 8.0 100mM; EDTA pH 8.0 5mM; SDS 0.5%; NaCl 150mM) using a glass-Teflon potter. RNA was digested by incubation at 37°C for 1 hour with Rnase A (40 µg/mL) (Sigma). After having added the Proteinase K (Roche) with a final concentration of 10 µg/mL samples were incubated O/N at 37°C. The day after, the genomic DNA was extracted using the phenol/chloroform/isoamyl alcohol method: 1 volume of phenol (water-saturated, pH 8.0) (Sigma) was added to the samples, followed by a centrifugation at 10000 rpm for 20 minutes. The aqueous upper phase was

collected in a new tube, and 1 volume of phenol : (chloroform-isoamyl alcohol (24:1)) was added, followed by a centrifugation at 10000 rpm for 10 minutes. The upper aqueous phase was again collected in a new tube, 1 volume of chloroform : isoamyl alcohol (24:1) was added and centrifuged at 10000 rpm for 10 minutes. The resulting upper aqueous phase was finally collected in a new tube and DNA was precipitated by adding two volumes of 100% ethanol. DNA white flakes were transferred into fresh tubes containing 200 µL of 70% ethanol and centrifuged at 12000 rpm for 15 minutes at 4°C in order to wash and gradually hydrate them. After ethanol removal, the DNA pellets were air dried and dissolved in 300 µL of Tris 10mM pH 8.0 O/N. The DNA quality was finally assessed by gel electrophoresis using a 0.9% ethidium bromide agarose gel.

Genomic DNA quantification

Since L1 sequences are extremely numerous in the mammalian genome, it was necessary to use a precisely quantified small amount of genomic DNA, to be able to discriminate small variations. According to the literature (Coufal et al., 2009), the optimal amount of genomic DNA for an L1 copy number variation analysis is 80 pg, corresponding approximately to 12 human genomes.

In order to obtain such an accurate genomic dilution we took advantage of the Quant-iT™ PicoGreen® dsDNA kit (Invitrogen), an ultrasensitive dsDNA quantification method that allows the quantification of dsDNA solutions in concentrations ranging from 25 pg/mL to 1000 ng/mL.

First of all the purified genomic DNA concentration for each sample was spectrophotometrically measured using NanoDROP (Thermo Scientific). According to the starting concentration, DNA samples were diluted in TE buffer (10mM Tris-HCl, 1 mM EDTA, pH 7.5) to a final concentration of 80 pg/µL. The supplied λ bacteriophage dsDNA was used to prepare a five-points standard curve at the following concentrations: 0.1 ng/mL, 0.5 ng/mL, 1ng/mL, 5 ng/mL, 10 ng/mL. Each point of the standard curve, the DNA samples and the blank were diluted in TE buffer to a final volume of 100 µL, loaded in duplicate in a microtiter plate, and 100 µL of working solution was added to each well (Quant-iT™ PicoGreenR dsDNA reagent 200X diluted in TE buffer).

After a 5 minutes incubation at RT, standard curve, DNA samples and blank were measured in duplicate using SpectraMax M5 Multimode Microplate Reader (Molecular

Devices). By plotting the measured fluorescence versus DNA concentration, a standard curve plot was generated and its equation used to assess the DNA concentration of the DNA samples. Final DNA concentrations ranging from 60 to 100 pg/µL were accepted for the PCR analysis.

Quantitative real-time PCR (qPCR) with Taqman probes

The qPCR technique

Quantitative Real-Time PCR is a technique that, through amplification and detection of nucleic acids, allows to determine the starting quantity of a template with accuracy, specificity and high sensitivity. Unlike conventional PCR where the amplification product is detected by an end-point method, with qPCR is possible to detect the amplicon at each step of amplification, as the reaction progresses. This detection is possible thanks to the use of fluorescent molecules, such as DNA-binding dyes or fluorescently labeled sequence-specific probes, whose fluorescent signal intensity is proportional to the amount of DNA amplified in each cycle.

Unlike the DNA-binding dyes often used in gene expression analysis, the introduction of fluorescently labeled sequence-specific probes (such as Taqman probes) in qPCR experiments has increased the specificity of the signal and the reproducibility of the assay.

In this study we performed multiplex qPCRs using Taqman probes differentially labeled (with FAM or VIC fluorophore) and specifically designed to hybridize with the target DNA sequences. The relative quantification of the target sequence was possible thanks to the co-amplification in the same tube of the target DNA sequence together with an internal control, a DNA sequence present in constant number between different individuals.

The experiments were carried out using a 7900HT Fast Real Time PCR System (Applied Biosystem) with the following conditions:

- 0.6 µL of (10 µM) target forward primer
- 0.6 µL of (10 µM) target reverse primer
- 0.2 µL of (10 µM) target probe
- 0.6 µL of (10 µM) control forward primer
- 0.6 µL of (10 µM) control reverse primer
- 0.2 µL of (10 µM) control probe

- 10 µL of 2X iQ Multiplex Powermix (Biorad)
- 1 µL of gDNA (=80 pg of DNA)
- Up to 20 µL with water

qPCR cycling program was set as represented in Figure 22.

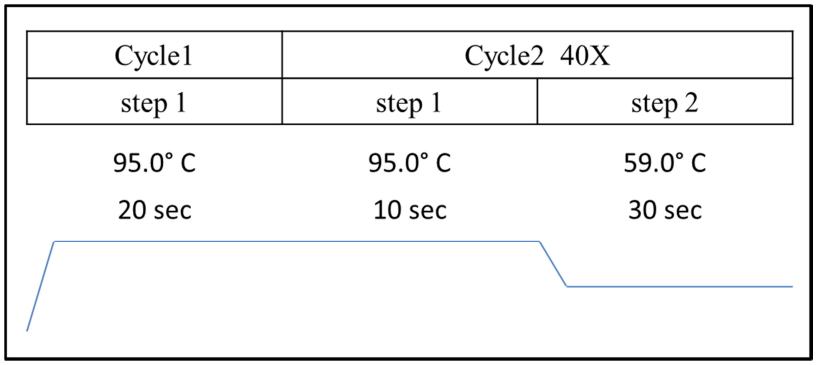


Figure 22: qPCR cycling conditions.

Standard curves of genomic DNA ranging from 20 pg to 640 pg were used to verify that the 80 pg dilution tested was within the linear range of reaction. Primer efficiency and multiplexing efficacy was verified by linear regression to the standard curve with a slope near -3.32 representing acceptable amplification efficiency.

Data obtained from the co-amplifications of the target DNA sequence and the internal invariable control were analyzed using the $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen, 2001) method. For each sample, the average Ct values of the duplicate for both target and control probes were first calculated and then the average value of the reference probe was subtracted to that of the test probe, obtaining the value of ΔCt:

ΔCt = average Ct test probe – average Ct reference probe

The highest value of ΔCt was chosen as calibrator and all samples were normalized relative to this calibrator value. In this way ΔΔCt was calculated:

ΔΔCt = ΔCt sample – ΔCt calibrator

Finally, the ratio was calculated as follows:

$2^{-\Delta\Delta Ct}$ = normalized amount of target

At least 3 technical replicas of each assay were performed for each sample.

Human L1 Taqman copy number variation assay

The Taqman qPCR assays used in this study for the analysis of L1 copy number variation in the human genome were in part adapted from those published by Coufal and colleagues (Coufal et al., 2009), and in part designed in our laboratory (Figure 23).

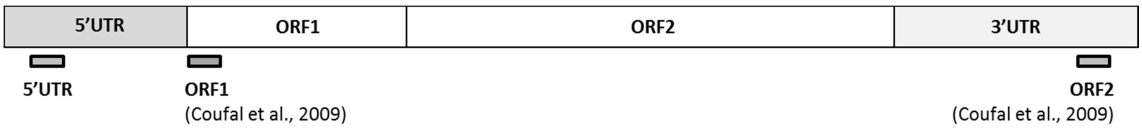


Figure 23: The human L1 Taqman CNV assays. The assay that detects both the 5'-truncated and the full length forms of the L1 element was called ORF2 by Coufal and colleagues, even if designed in the 3'UTR portion. The ORF1 assay (designed by Coufal et al.) and the 5'UTR assay that we designed detect only the full length forms.

This Taqman qPCR strategy was planned taking into account the fact that L1 5'UTR probes detect the L1 full length forms, corresponding to original L1 elements, whereas probes complementary to the ORF2 region detect the entire L1 repertoire, including both full length and truncated forms, which correspond to both original and retrotransposed L1 elements.

In order to estimate the total amount of human L1 sequences present in the human genome, we employed a Taqman assay which detects the L1 ORF2 content (a portion of the L1 element that is present in both the full length and the 5'-truncated forms), and a probe that is specific for the invariant high copy number internal control SATA (α-satellite sequences, about 1 million copies) (Table 3).

In order to quantify the L1 full length form, we took advantage of the ORF1 Taqman assay published by Coufal and colleagues, using as invariant control an assay that we designed in our lab. Indeed, after an analysis of possible invariant genes at low copy number (compatible with the amount of full length L1 present in the human genome), the Taqman assay that we designed on the glycerhaldeyde 3-phosphate dehydrogenase or GAPDH (65 copies in the human genome) resulted as the most stable.

Furthermore, since Coufal's assay for the full length L1 amplifies the ORF1 portion, to be sure to measure the relative content of the complete L1 sequence, we designed a new further Taqman assay at the very beginning of L1 5'UTR, to be used with the GAPDH assay. The new 5'UTR Taqman assay was designed by alignment in the L1Base, and considering the analysis performed by Lavie and colleagues on L1 promoter region, to define the very beginning of the 5'UTR region (Lavie et al., 2004; Penzkofer et al., 2005).

Table 3: The Taqman qPCR assays for the analysis of L1 copy number variation in the human genome.

Assay	Forward primer (5'→3')	Reverse primer (5'→3')	Probe (5'→3')
L1 ORF2 (Coufal et al., 2009)	TGCGGAGAAATAGGAACACTTTT	TGAGGAATCGCCACACTGACT	CTGTAAACTAGTTCAACCATT **(FAM)**
L1 ORF1 (Coufal et al., 2009)	GAATGATTTTGACGAGCTGAGAGAA	GTCCTCCCGTAGCTCAGAGTAATT	AAGGCTTCAGACGATC **(FAM)**
L1 5'UTR	GAGGTACCGGGTTCATCTCA	TCACCCCTTTCTTTGACTCG	TAGGGAGTGCCAGACAGTGG **(FAM)**
SATA (Coufal et al., 2009)	GGTCAATGGCAGAAAAGGAAAT	CGCAGTTTGTGGGAATGATTC	TCTTCGTTTCAAAACTAG **(VIC)**
GAPDH	CCCTTCATTGACCTCAACTACATG	TGGGATTTCCATTGATGACAAGC	CGTTCTCAGCCTTGACGGTGCCAT **(VIC)**

Data analysis was performed using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001), as previously explained. Standardization was performed considering the highest control ΔCt value as calibrator.

Mouse L1 Taqman copy number variation assay

In order to measure L1 copy number variation in the mouse genome, we developed new Taqman assays based on those one used for the human genome.

Since the murine L1 5'UTR region contains different repetitive monomers which characterize different L1 subfamilies, we designed specific probes for the 5'UTR of each active mouse L1 type (A, Tf and Gf), allowing us to discriminate between the different mouse L1 subfamilies. The same approach was used to design assays specific for the 3'UTR region of each L1 subfamily, taking advantage of a polymorphic region present in the 3'UTR sequence. On the other hand, we also designed probes and primers complementary to the ORF2 region of the L1 sequence able to detect the total amount of L1 sequences, without discriminating between the different L1 families present in the mouse genome (Figure 24 and Table 4).

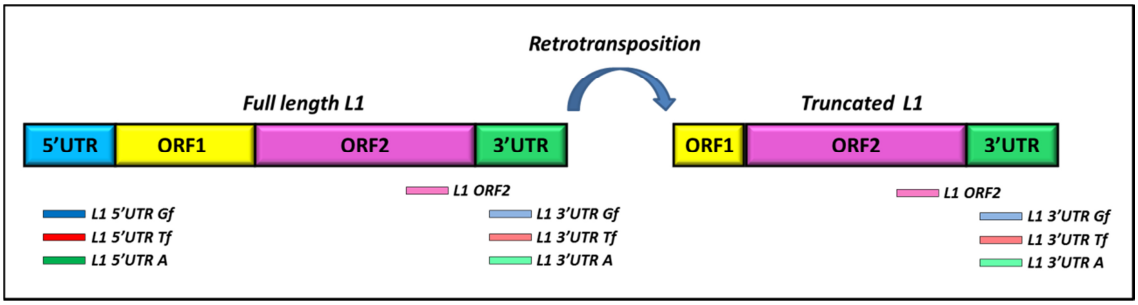


Figure 24: Schematic representation of L1 retrotransposition and multiplex quantitative PCR assays for L1 copy number variation analysis in mouse. Different Taqman assays, represented with colored boxes, were designed on the ORF2 domain (L1 ORF2 probe), 3'UTR region (L1 3'UTR Gf, Tf, A) or 5'UTR region (L1 5'UTR Gf, Tf, A). Since the vast majority of L1 insertions are truncated at the 5'end, L1 ORF2 probe detects the total number of L1 copies, including both full-length and truncated forms, as well as the 3'UTR probes designed to detect different L1 subfamilies(L1 Gf, L1 Tf, L1 A), whereas the 5'UTR probes for each L1 subfamily measure the number of LINE1 full-length forms.

In order to perform a relative quantification of the truncated and the full length L1 copies, we used, as internal controls, Taqman probes designed on non-mobile repetitive

genomic sequences with high or low repeats number. In particular we designed an assay for a high copy number invariant sequence: the centromeric microsatellite sequence or MICSAT (about 500000 copies), that we used together with the assay that detects the total L1 sequences in the mouse genome (ORF2 probe) or the assay that detects the 3'UTR sequence of the A or Tf family. Since the A and Tf families are the most ancient and still retrotransposing families in the mouse genome, they are highly numerous, and require an invariant control at high copy number. The 3'UTR sequence of the Gf family, which is much younger and less numerous, was quantified using another assay at low copy number, that we designed on the genomic sequence of glyceraldehyde 3-phosphate dehydrogenase or GAPDH (362 copies).

By doing this, we were able to assess and compare the amount of original full length L1 and the number of total L1 elements (both full length and truncated forms). In addition, since in the mouse genome the original full length L1 elements account for barely 1% of the total amount of L1 elements, the relative quantification of ORF2 content represented an estimate of L1 rate of retrotransposition.

In summary mouse samples were tested using the following combinations of assays:

1) L1 ORF2/MICSAT: estimates the total amount of L1 elements (both intact and truncated forms) providing the rate of L1 retrotransposition;

2) 3'UTR/GAPDH or MICSAT: provides the total number of L1s belonging to a specific family (both intact and truncated forms) and estimates of the rate of specific L1 family retrotransposition;

3) 3' UTR/5' UTR: represents the rate of L1 retrotransposition of specific L1 families.

Table 4: The Taqman qPCR assays (primers and probe) for the analysis of L1 copy number variation in the mouse genome.

Assay	Forward primer (5'→3')	Reverse primer (5'→3')	Probe (5'→3')	
ORF2 (common)	CCCTCAACAGAGGAATGGAT	CCATCCATTTGGCTAGGAAT	AAATGTGGTACATCTACACAATGGA	**(FAM)**
5'UTR **Tf** family	TGAGCACTGAAACTCAGAGGAG	GATTGTTCTTCTGGTGATTCTGTTA	GAATCTGTCTCCCAGGTCTG	**(FAM)**
5'UTR **A** family	TGCCCACTGAAACTAAGGAGA	GCTTGTTCTTCAGGTGACTCTGT	TGCTACCCTCCAGGTCTGCT	**(FAM)**
5'UTR **Gf** family	CCAAACACCAGATAACTGTACACC	CGTGGGAGACAAGCTCTCTT	TGAAAGAGGAGAGCTTGCCT	**(FAM)**
3'UTR **Tf** family	CATGGAAGGAGTTACAGAAACAA	ATCCCTGGATATGGCAGTC	TGAGATGAAAGGATGGACCA	**(FAM)**
3'UTR **A** family	CATGGAAGGAGTTACAGAGACGG	ATCCCTGGCTATGGCAGTC	TGAGATGAAAGGATGGACCA	**(FAM)**
3'UTR **Gf** family	CTTGGAAGGAGTTACAGAGACAA	ATCCCCGGATATGCTAGTG	TGAGATAAAAGGATGGACCA	**(FAM)**
MICSAT	GAACATATTAGATGAGTGAGTTAC	GTTCTACAAATCCCGTTTCCAA	ACTGAAAAACACATTCG	**(VIC)**
GAPDH	CGACCCCTTCATTGACCTC	CTCCACGACATACTCAGCACC	CTCCACTCACGGCAAATTC	**(VIC)**

Data analysis was performed using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001), as previously explained. Standardization was performed considering the highest control ΔCt value as calibrator.

Statistical analysis

Each qPCR assay was performed at least three times on all the samples. Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

The three replicas were analyzed individually and considering the mean values for each sample with the Wilcoxon rank sum test for equal medians, and with the paired, two-tailed, t-student test using MATLAB® Statistics Toolbox.

SPlinkerette Analysis of Mobile elements (SPAM)

Splinkerette PCR (spPCR) is a widely used PCR approach to clone flanking DNA regions of transposable elements and viruses in the genome. In this technique, genomic DNA is fragmented and ligated to a synthetic double stranded oligonucleotide (the splinkerette) that contains a compatible sticky end and a stable hairpin loop. Two rounds of nested PCR are then performed to amplify the genomic sequence between the integrant and the annealed splinkerette, followed by a sequencing reaction. The hairpin structure present in the adaptor was shown to increase the specificity of the ligation, in respect to other ligation-mediated PCRs reaction, and to prevent the amplification of unspecific products (Uren et al.; 2009).

In this study we set up a protocol (SPAM) inspired by the spPCR, aimed at identifying L1 insertion sites in the genome (Figure 25). The vast majority of the spPCR protocols use enzymatic digestion to create DNA fragments, but this procedure can introduce amplification biases, linked to an uneven genomic distribution of restriction enzyme recognition sites (Koudijs et al., 2011). In order to avoid possible amplification biases we decided to substitute the enzymatic digestion with sonication, as previously done by Koudijs et al., and create 300 bp DNA fragments.

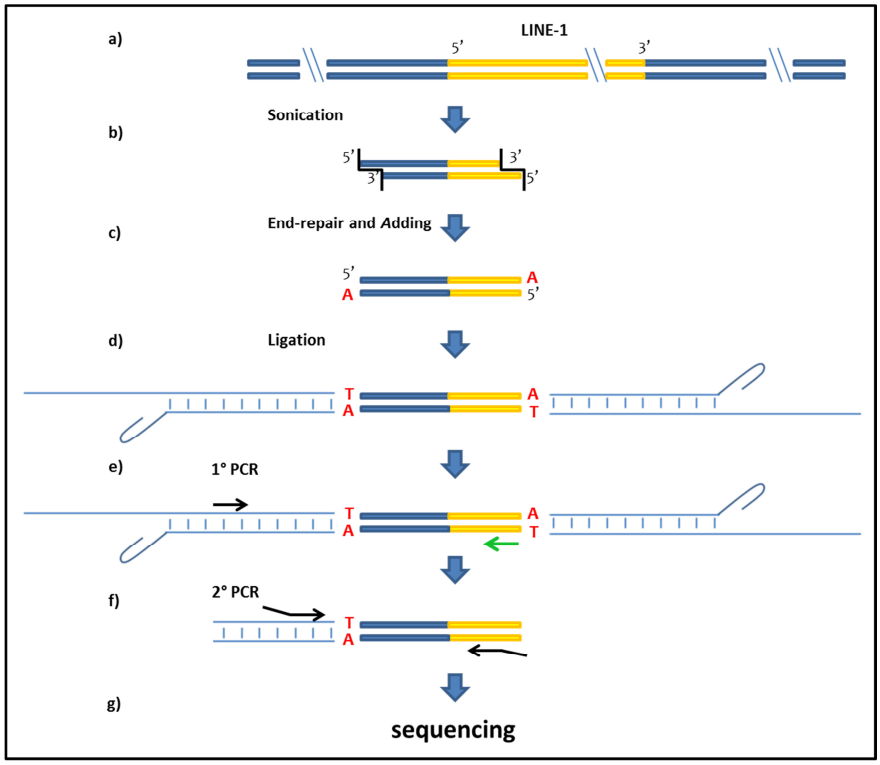


Figure 25: The SPAM technique. a) Representation of the genome with an L1 insertion. b) Sonication of gDNA to the average fragments size of 300 bp. c) End-repair and A-adding reactions, to make the fragments blunt and added with an extra-A. d) Ligation of gDNA fragments to the adaptor. e) First round of PCR to amplify the genomic region flanking the L1 insertion. f) Nested PCR to make the amplification more specific. g) MiSeq Illumina sequencing of the purified amplicons.

We performed a first test of the SPAM technique on samples of DNA extracted from the temporal cortexes of three healthy controls of the Italian cohort (used in the CNV experiments). While for the preliminary SPAM experiment on AD, we used samples of frontal cortex and kidney from 3 AD patients and three healthy controls of the Brazilian cohort.

Genomic DNA sonication

For each sample 2 µg of genomic DNA in 100 µL of nuclease-free water (Ambion®) were sheared to obtain an average fragments size of 300 bp with the program reported in Table 5, using the Bioruptor NGS at IGA (Udine).

Table 5: Sonication program for shearing genomic DNA to 300 bp fragments using the Bioruptor NGS.

7 cycles	1) 30 seconds	Sonication at max power
	2) 90 seconds	Stop

DNA fragmentation profiles (an example run is reported in Figure 26) were checked by Bioanalyzer run using a 2100 expert_High Sensitivity DNA Assay (Agilent Technologies).

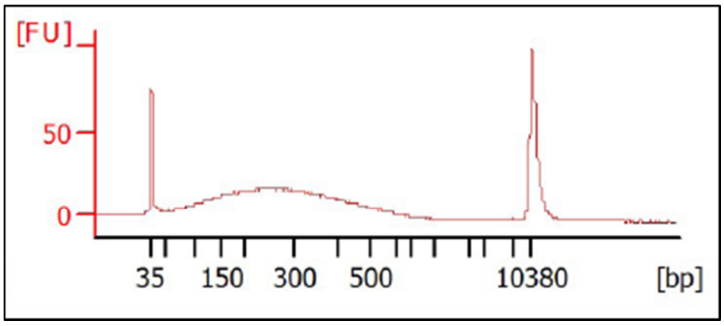


Figure 26: genomic DNA Fragmentation profile obtained with a Bioanalyzer run using a 2100 expert_High Sensitivity DNA Assay

The sonicated aliquots were then concentrated to a volume of 30 µL by Speedvac concentrator in order to proceed with the following end-repair step.

End-repair reaction

Since the gDNA fragments after sonication present irregular sticky ends, samples were processed with an end-repair reaction (inspired by Illumina protocols for sequencing libraries preparation) in order to get blunt-ended fragments. The reaction was performed as follows:

- 30 µL of sonicated gDNA
- 5 µL of 10X T4 Ligase buffer with 10 mM ATP (NEB)
- 2 µL of 10 mM dNTPs
- 1 µL of T4 DNA Polymerase (NEB)
- 1 µL of Klenow large fragment (NEB)
- 1 µL of Polynucleotide Kinase (NEB)
- Up to 50 µL with water

The reactions were incubated at 20°C for 30 minutes and then purified using the Qiaquick PCR purification kit (Qiagen) following manufacturer's instructions. End-repaired samples were eluted in 34 µL of elution buffer (Qiagen).

A-adding reaction

Before performing the ligation reaction to the adaptor, an extra-A at the 3'ends was added to the fragments, in order to make them compatible with the extra-T present at the 3'end of the adaptor.

The reaction was set as follows:

- 34 µL of end-repaired gDNA
- 5 µL of NEB buffer 2
- 1 µL of Klenow fragment (3'→5' exo$^-$)
- 10 µL of 1 mM dATPs

The reactions were incubated at 37°C for 30 minutes and then purified using the Qiaquick PCR purification kit (Qiagen) following manufacturer's instructions. End-repaired samples were eluted in 30 µL of water.

Adaptor ligation

In this protocol we used the adaptor sequence that was first reported by Mikkers and colleagues (Mikkers et al., 2002), but with a minor modification: since we fragmented the genomic DNA by sonication and not by enzymatic digestion, we eliminated the sticky end compatible with the enzymatic cut and added to the longest strand a protruding T, complementary to the extra-A of the genomic fragments (as previously done in the STA-PCR protocol by Yin and Largaespada, 2007) (Figure 27).

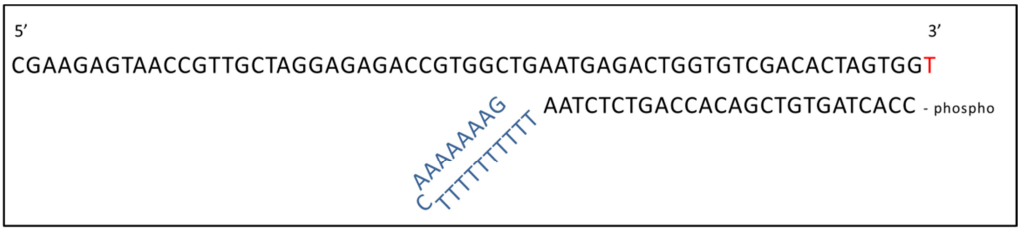


Figure 27: The SPAM adaptor structure.

The adaptor was assembled as follows: two oligonucleotides (Sigma) of the adaptor, corresponding to the separated adaptor strands, were resuspended in TE buffer (pH 7.5) to a final concentration of 100 µM and equal amounts of both were mixed together to reach a final concentration of 50 µM. The oligonucleotides were denatured at 95°C for 5 minutes and the annealing was performed by slow cooling to RT (the tube was kept in the switched off thermoblock until the solution reached the RT).

The ligations of each end-repaired sample were performed as follows:

- 8 µL of end-repaired and A-added gDNA
- 5 µL of 10X T4 Ligase buffer with 10 mM ATP (NEB)
- 2 µL of adaptor's solution
- 2.5 µL of T4 DNA Ligase (NEB)
- Up to 50 µL with water

The ligation reactions were performed at 16°C ON, plus 1 hour at 37°C after the addition of 0.5 µL more of enzyme, to make the ligation reaction more efficient.

Three ligation reactions were performed per sample, pooled together and purified using the Qiaquick PCR purification kit (Qiagen) following manufacturer's instructions. Ligated samples were eluted with 40 µL of 10 mM Tris-HCl pH 7.4, ready for PCR.

PCR reactions

In order to amplify the genomic region upstream to the 5'UTR of L1 insertions, we performed two sequential rounds of PCR, using forward primers specific to the adaptor sequence, and reverse primers specific to the very first part of L1's 5'UTR sequence.

Besides the specific portion, the primers used in the nested PCR contained also the tag sequences and the barcodes necessary for the Illumina MiSeq sequencing as reported in Figure 28.

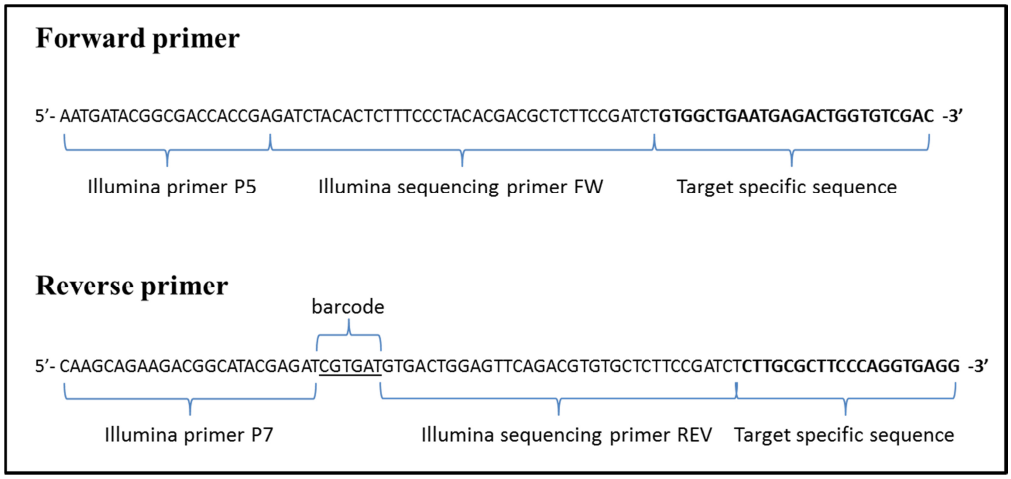


Figure 28: Example of the SPAM nested primers' structure

The portions of the two forward primers specific to the adaptor sequence were already published by Uren and colleagues (Uren et al., 2009), while the reverse primers specific to the human L1's 5'UTR were designed in our lab:

- Human reverse primer #1 (5'→3'):
  CGTCCGTCACCCCTTTCTTTGACTCG
- Human reverse primer #2 (5'→3'):
  CTTGCGCTTCCCAGGTGAGG

The primary PCR reactions for each sample were performed in a thermocycler (ABI) with the following conditions:

- 5 µL of purified ligated gDNA fragments
- 1.2 µL of 10 µM 1° forward primer specific to the adaptor
- 1.2 µL of 10 µM 1° reverse primer specific to L1's 5'UTR
- 1.25 µL of 10 mM dNTPs
- 5 µL of 10X High Fidelity PCR buffer (Invitrogen)
- 2 µL of 50 mM $MgSO_4$ (Invitrogen)
- 0.25 µL of Platinum® *Taq* High Fidelity (Invitrogen)
- Up to 50 µL with water

The protocol is reported in Table 6

Table 6: Primary SPAM PCR's thermocycler program

Cycle 1	Cycle 2 29X			Cycle 3
step 1	step 1	step 2	step 3	step 1
94.0° C	94.0° C	68.0° C	68.0° C	68.0° C
2 min	15 sec	30 sec	3 min	5 min

Three reactions of primary PCR for each sample were pooled together. The secondary nested PCRs were performed as follows with the protocol reported in Table 7:

- 1 µL of primary PCR product
- 1.2 µL of 10 µM 2° forward primer specific to the adaptor
- 1.2 µL of 10 µM 2° reverse primer specific to L1's 5'UTR
- 1.25 µL of 10 mM dNTPs
- 5 µL of 10X High Fidelity PCR buffer (Invitrogen)
- 2 µL of 50 mM $MgSO_4$ (Invitrogen)
- 0.25 µL of Platinum® *Taq* High Fidelity (Invitrogen)
- Up to 50 µL with water

Table 7: Secondary SPAM PCR's thermocycler program

Cycle 1	Cycle 2 25X			Cycle 3
step 1	step 1	step 2	step 3	step 1
94.0° C	94.0° C	60.0° C	68.0° C	68.0° C
2 min	15 sec	30 sec	5 min	5 min

For each primary PCR three secondary PCR reactions were performed, mixed and purified using the Microcon® DNA Fast Flow Centrifugal Filters (Millipore) with the following procedure: sample denaturation at 95°C for 10 minutes, adding of cooled water up to 500 µL, loading of the sample in the centrifugal filter, centrifugation at 4°C for 18 minutes at 500 g, inversion of the filter in a new tube and centrifugation at 4°C for 2 minutes at 500 g to transfer the purified and concentrated sample. Samples were finally quantified by Nanodrop (Thermoscientific): an average concentration of 190 ng/µL was obtained.

Control PCR

In order to check L1 sequences' enrichment after the nested PCR, before sequencing we performed a control PCR on serial dilutions of the nested product and of the corresponding genomic DNA using the following couple of primers, specific for the very first part of the human L1's 5'UTR region:

- Human forward control primer (5'→3'):
  AGGGAGTGCCAGACAGTG
- Human reverse control primer (5'→3'):
  AATGCCTCGCCCTGCTTC

For each sample the PCR was performed as follows with the protocol reported in Table 8:

- 1 µL of secondary PCR product
- 1 µL of 10 µM forward control primer
- 1 µL of 10 µM reverse control primer
- 1.6 µL of 2.5 mM dNTPs
- 2 µL of 10X Ex Taq buffer (Takara)
- 0.1 µL of Ex Taq™ (Takara)
- Up to 20 µL with water

Table 8: Control PCR's thermocycler program

Cycle 1	Cycle 2 40X			Cycle 3
step 1	step 1	step 2	step 3	step 1
98.0° C	98.0° C	57.0° C	72.0° C	72.0° C
2 min	10 sec	30 sec	30 sec	5 min

The PCR products were run on a 2.5% ethidium bromide agarose gel.

MiSeq Illumina sequencing

SPAM samples were sequenced at IGA (Udine) using the Illumina MiSeq technology (300bp paired-end set-up). Since different reverse nested primers with different barcodes were used to perform the secondary PCR, it was possible to sequence samples in multiplex, and repeating the runs allowed us to obtain as much reads as possible for each sample.

Bioinformatic analysis of SPAM-human samples

We established a bioinformatics pipeline with a precise nomenclature:

Reads: R1 and R2, single reads coming from paired end sequencing (forward and reverse sequences).

Fragment: assembled reads (R1 + R2).

MapFragment: fragment containing an L1 sequence and a mappable unique sequence.

MapCluster: at least three MapFragments in at least one sample.

Integration Site (IS): genomic region where one or more MapFragment have been mapped.

Annotated Integration site (AIS): integration site located at less than 1000 bp far from an L1 in the database.

Novel Integration site (NIS): integration site located at more than 1000 bp far from an L1 in the database.

The first step of the analysis was to align the corresponding R1 and R2 to form a Fragment, discarding those pairs that didn't align with at least 30bp and 80% identity. Then we selected Fragments containing the forward primer (specific for the splinker) and searched for the repetitive element's sequence, by alignment of the Fragments to the reference sequence. Once discarded the fragments that did not contain the repeat, we trimmed the portions of the fragments that overlapped with the forward primer and the repeat's sequence, and retained (after the trimming) only the sequences longer than 30 nucleotides. These sequences were run with blastn against the reference human genome (Hg19) in order to map them. MapFragments were defined only if they had a unique alignment result with an identity >50%.

These MapFragments could identify an AIS or a NIS, according to the presence or not of these insertion sites into the L1 database that we compiled.

Then a gene ontology study was performed in order to test the enrichment of IS associated genes involved in particular cell functions. The final step of the analysis was the validation by PCR of some of the AIS and NIS detected with the SPAM technique.

Validation PCRs

After the bioinformatics analysis, we deepened the investigation of some AIS and NIS. In particular, for the first test of the SPAM technique on the three human samples, we chose 10 integration sites for each sample, that we found to be present in exons, introns, or in brain specific gene loci. We analyzed each of these integration sites (annotated and

novel) and we decided to select a group of 8 to be validated by PCR. These 8 IS were: 3 different NIS defined by 5-10 MapFragments and found in only one sample, 3 NIS defined by 1 MapFragment and found in only one sample, 1 NIS defined by more than 100 MapFragment and found in only one sample, and finally 1 AIS defined by more than 100 MapFragment and found in all the three samples.

For each of these 8 IS we performed a first round of PCR and a nested PCR using two couples of primers with the forward primers designed on the genomic DNA at the insertion site, and the reverse primers designed against the very beginning of the L1 5'UTR:

- First reverse validation primer (5'→3'):
  CTTGCGCTTCCCAGGTGAGG
- Second reverse validation primer (5'→3'):
  AATGCCTCGCCCTGCTTC

Primary PCRs were performed on genomic DNA and the corresponding SPAM product as follows with the protocol reported in Table 9:

- 3 µL of genomic DNA or SPAM product
- 1.2 µL of 10 µM 1° forward primer specific to the genomic sequence
- 1.2 µL of 10 µM 1° reverse primer specific to L1's 5'UTR
- 4 µL of 2.5 mM dNTPs
- 5 µL of 10X Ex Taq buffer (Takara)
- 0.25 µL of Ex Taq™ (Takara)
- Up to 50 µL with water

Table 9: First validation PCR's thermocycler program

Cycle 1	Cycle 2 40X			Cycle 3
step 1	step 1	step 2	step 3	step 1
98.0° C	98.0° C	60.0° C	72.0° C	72.0° C
3 min	10 sec	30 sec	1 min	5 min

The nested PCR was performed as follows with the protocol reported in Table 10:

- 1 µL of primary validation PCR product
- 1 µL of 10 µM forward control primer
- 1 µL of 10 µM reverse control primer

- 1.6 µL of 2.5 mM dNTPs
- 2 µL of 10X Ex Taq buffer (Takara)
- 0.1 µL of Ex Taq™ (Takara)
- Up to 20 µL with water

Table 10: Second validation PCR's thermocycler program

Cycle 1	Cycle 2 40X			Cycle 3
step 1	step 1	step 2	step 3	step 1
98.0° C	98.0° C	57.0° C	72.0° C	72.0° C
2 min	10 sec	30 sec	30 sec	5 min

Both the PCR products were run on a 2.5% ethidium bromide agarose gel.

SPAM-technique adaptation to the *roo*-LTR in *Drosophila*

We decided to adapt the SPAM technique for identifying also the insertion sites of the LTR portion of the *roo* element, the most abundant family of LTR retrotransposons in the *Drosophila* genome.

The genomic DNA sample that we used was provided by Prof. Pimpinelli and Dr. Piacentini (Sapienza University, Rome), and it had already been extracted from a pool of adult carcasses. In order to understand if a less amount of *Drosophila* genomic DNA (<2µg, the quantity used in the protocol for human samples) could be used, we applied the SPAM technique on the same DNA sample, but starting from a different quantity: 2µg and 500ng.

Drosophila samples were treated as the human samples until the adaptor-ligation step, but since we were interested in the amplification of the genomic portion flanking the *roo*-LTR element, we designed two reverse primers specific for the *roo*'s LTR sequence that we used in the primary PCR reaction with the same conditions as for the human samples:

- Drosophila reverse primer #1 (5'→3'):
  CTAAGGGACTATTTTAGGAGGCGGGGAA
- Drosophila reverse primer #2 (5'→3'):
  CGGGGAACGATCTCAAGTGACTGACTC

We designed also a couple of primers specific for the very first part of the *roo* LTR sequence to be used in the secondary PCR reaction:

- Drosophila forward control primer (5'→3'):
  GACTTACAATTTTGGGCTCCGTTC
- Drosophila reverse control primer (5'→3'):
  CGGGGAACGATCTCAAGTGACTGACTC

The bioinformatics analysis of the two *Drosophila* samples was performed following the pipeline established for the human samples, but with few exceptions. In this case the good-quality Fragments were aligned to the *roo* LTR reference sequence, and, after the trimming step, they were stringently aligned to the dm3 *Drosophila melanogaster* reference genome in order to map them.

This page intentionally left blank

Results

L1 retrotransposition in human AD samples

In order to study L1 copy number variation in the human genome we decided to adopt the qPCR technique with Taqman probes. The Taqman qPCR assays used in this study were in part adapted from those published by Coufal and colleagues (Coufal et al., 2009), and in part designed in our laboratory.

In order to estimate the total amount of human L1 sequences present in the human genome, we employed a Taqman assay which detects the L1 ORF2 content (a portion of the L1 element that is present in both the full length and the 5'-truncated forms), and a probe that is specific for the invariant high copy number internal control SATA (both published by Coufal and colleagues).

In order to quantify the L1 full length form, we took advantage of the ORF1 Taqman assay published by Coufal and colleagues, but using as invariant control an assay that we designed in our laboratory and that we selected after an analysis of possible invariant genes at low copy number: the glycerhaldeyde 3-phosphate dehydrogenase or GAPDH.

Furthermore, to be sure to measure the relative content of the complete L1 sequence, we designed a new further Taqman assay at the very beginning of L1 5'UTR, to be used with the GAPDH assay. The new 5'UTR Taqman assay was designed by alignment in the L1Base, and considering the analysis performed by Lavie and colleagues on L1 promoter region, to define the very beginning of the 5'UTR region (Lavie et al., 2004; Penzkofer et al., 2005).

Each assay was performed three times on all the samples, and graphs for all the three replicas and for mean values across replicas are reported.

L1 retrotransposition in an Italian cohort

Thanks to the collaboration with Prof. Tagliavini (Istituto Neurologico Carlo Besta, Milan), we collected post mortem samples of temporal cortex taken from 11 patients affected by AD, and 13 healthy controls (Table 11).

Table 11: Italian cohort of post mortem temporal cortex samples from the Istituto Neurologico Carlo Besta, Milan.

Italian cohort	N.	Braak stage	Age	Gender (F:M)	PMD
CTRL	13	N.A.	65 ± 15	8:5	N.A.
AD	11	N.A.	66 ± 18	6:5	N.A.

We measured the total amount of L1 sequences (both truncated and full length) with the ORF2 assay (Figure 29), and we observed that no differences between AD samples and controls were present. Nevertheless a high variability was evident.

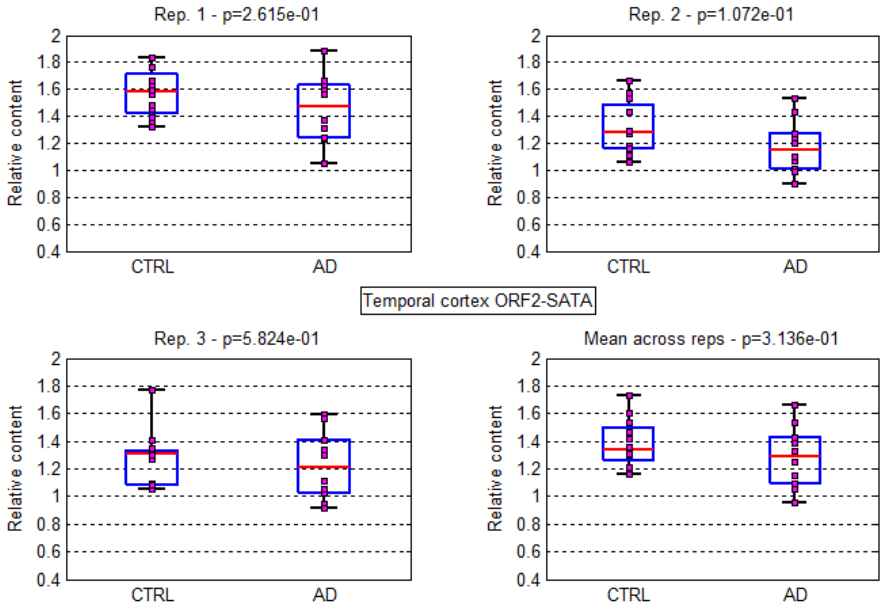


Figure 29: qPCR analysis of total L1 copy number in temporal cortexes of AD patients and controls. Relative quantification obtained by qPCR with the ORF2 Taqman probe and SATA. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (healthy controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

Also the relative content of full length L1s, measured using both the ORF1 (Figure 30) and the 5'UTR (Figure 31) assays, in both cases did not vary between AD samples and controls.

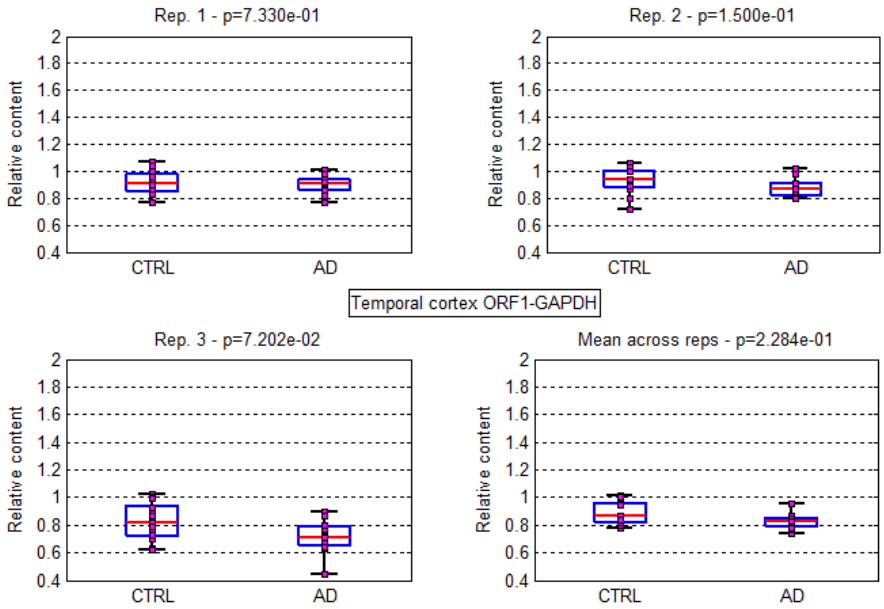


Figure 30. qPCR analysis of L1 full length copy number in temporal cortexes of AD patients and controls. Relative quantification obtained by qPCR with the ORF1 Taqman probe and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (healthy controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

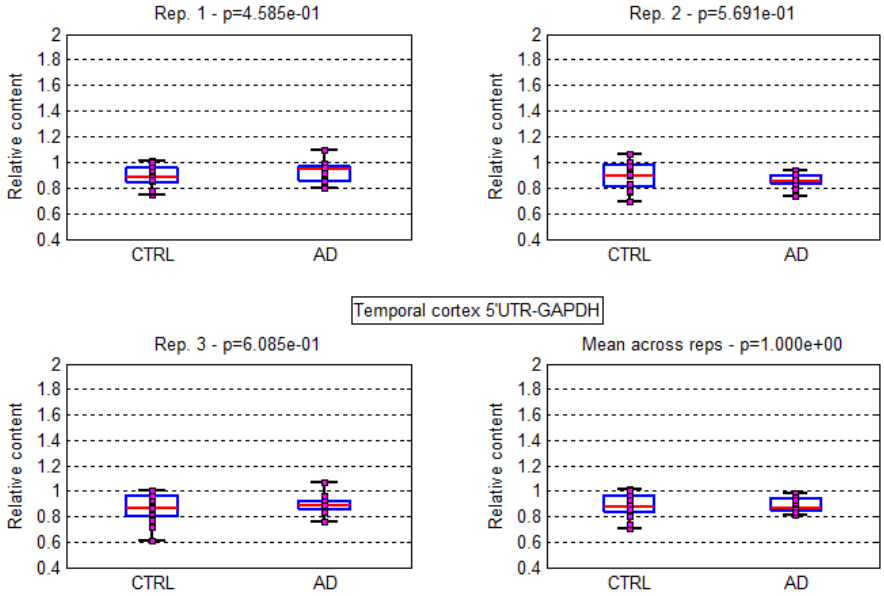


Figure 31: qPCR analysis of L1 full length copy number in temporal cortexes of AD patients and controls. Relative quantification obtained by qPCR with the 5'UTR Taqman probe and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (healthy controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

L1 retrotransposition in a Spanish cohort

We performed the qPCR with Taqman probes also on a Spanish cohort of samples that were collected thanks to the collaboration with Prof. Isidro Ferrer (Bellvitge Neuropathology Institute, Barcelona). They were post mortem samples of frontal cortex from 10 AD patients at the final AD Braak stages V-VI (late AD), 10 patients at Braak stages I-II (early AD), and 7 healthy controls (Table 12).

Table 12: Spanish cohort of post mortem frontal cortex samples from the Bellvitge Neuropathology Institute, Barcelona.

Spanish cohort	N.	Braak stage	Age	Gender (F:M)	PMD
CTRL	7	0	70 ± 8	2:5	3 ± 1
Early AD	10	I-II	75 ± 5	5:5	8 ± 6
Late AD	10	V-VI	80 ± 4	5:5	10 ± 5

In this case we decided to compare the relative amount of L1's full length forms present in the three groups of samples, using both the ORF1 and the 5'UTR assays. By using the ORF1 assay we observed a significant decrease (~8%) of full length L1 content in AD patients affected by early AD and late AD compared to healthy controls, with no differences detectable between the early and the late groups (Figure 32).

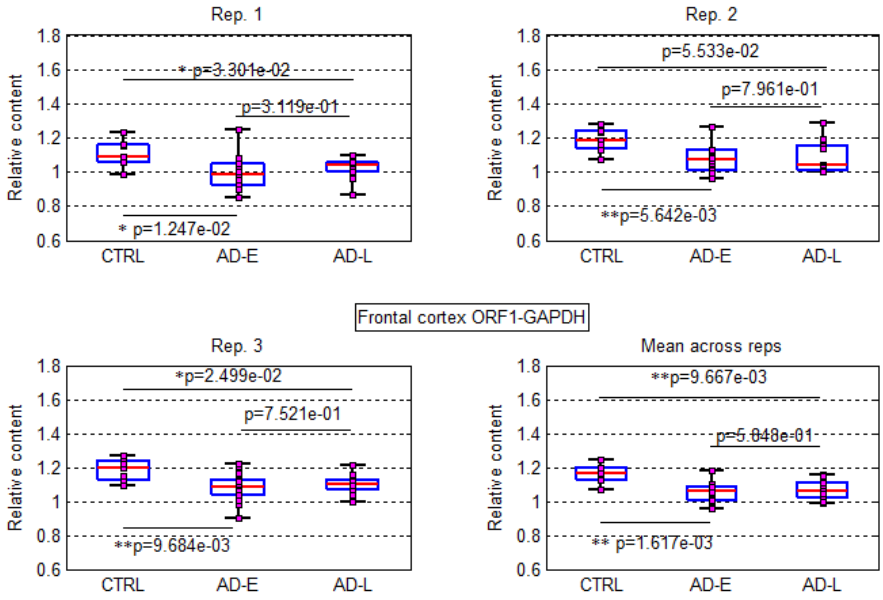


Figure 32: qPCR analysis of L1 full length copy number in temporal cortexes AD patients at different stages of the disease and controls. Relative quantification obtained by qPCR with the ORF1 Taqman probe and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (healthy controls), AD-E (patients affected by early AD) and AD-L (patients affected by late AD). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05; **p<0.01).

On the other hand, analyzing L1's full length forms using the 5'UTR assay, we detected a progressive and always significant decrease in the content of L1 full length sequences starting from the healthy controls group to the group of patients affected by late AD (Figure 33). In particular we observed a decrease of ~14% between controls and early AD patients, ~20% between controls and late AD patients, and finally ~7% between early and late AD patients.

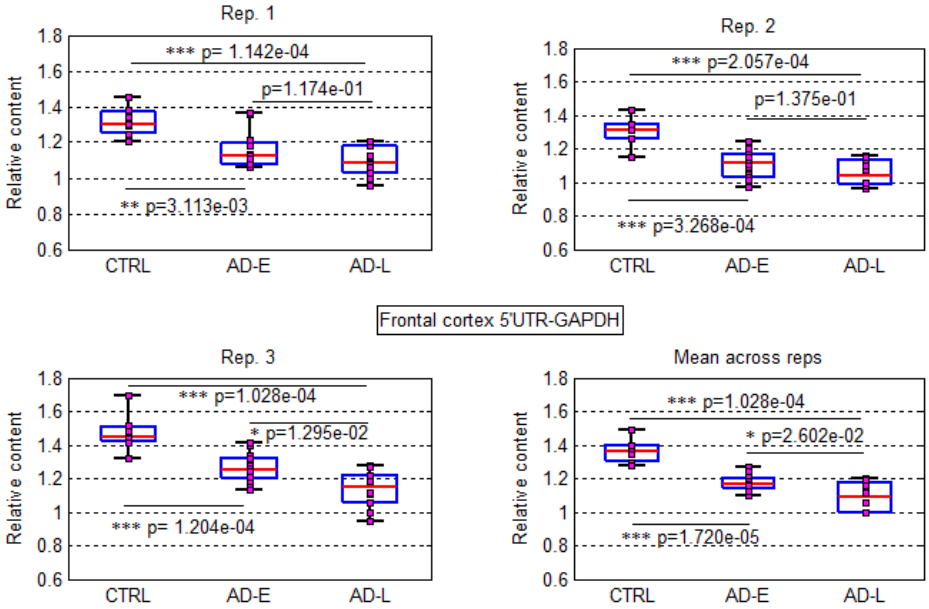


Figure 33: qPCR analysis of L1 full length copy number in temporal cortexes of AD patients at different stages of the disease and controls. Relative quantification obtained by qPCR with the 5'UTR Taqman probe and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (healthy controls), AD-E (patients affected by early AD) and AD-L (patients affected by late AD). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05; **p<0.01; ***p<0.001).

L1 retrotransposition in a Brazilian cohort

We collected genomic DNA extracted from samples of frontal cortex, temporal cortex, hippocampus, cerebellum and an extra-nervous tissue: the kidney. The tissues, stored at the Brain Bank of Sao Paulo (provided by Prof. Lea Grinberg) and belonging to a Brazilian cohort, had been taken from 10 AD patients at Braak stages IV-VI and 10 patients at Braak stages 0-II (Table 13).

Table 13: Brazilian cohort of post mortem tissue samples from the Brain Bank of Sao Paulo. For each individual we collected genomic DNA of frontal cortex, temporal cortex, hippocampus, cerebellum and kidney.

Brazilian cohort	N.	Braak stage	Age	Gender (F:M)	PMD
CTRL	10	0-II	80 ± 15	6:4	N.A.
AD	10	IV-VI	84 ± 10	6:4	N.A.

At first we measured the total amount of L1 sequences (both truncated and full length) with the ORF2 assay, and in all the analyzed tissues we couldn't be able to see any significant difference in the number of total L1 elements between patients and controls, probably because too abundant to allow the detection of small differences (Figure 34, Figure 35, Figure 36, Figure 37 and Figure 38). Nevertheless a high variability was evident.

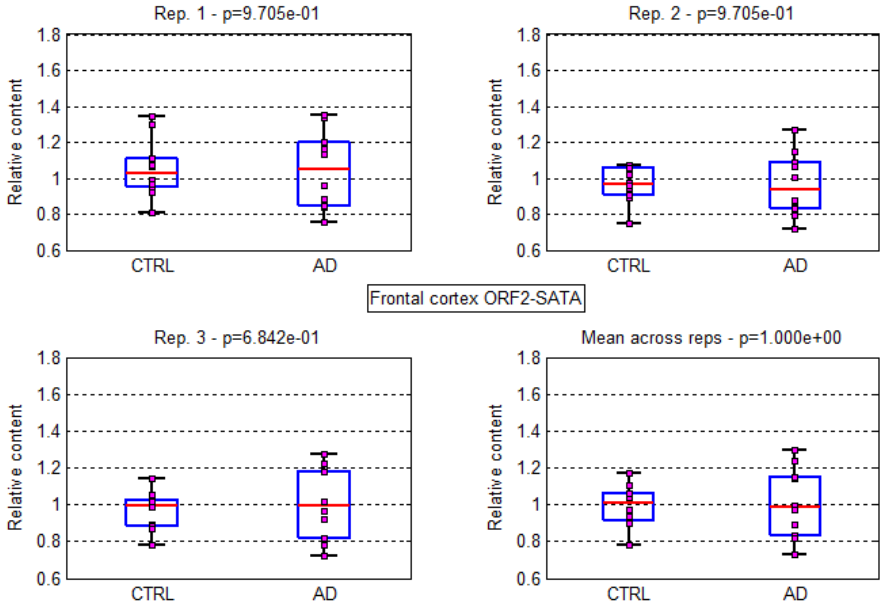


Figure 34: qPCR analysis of total L1 copy number in frontal cortexes of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe ORF2 and SATA. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

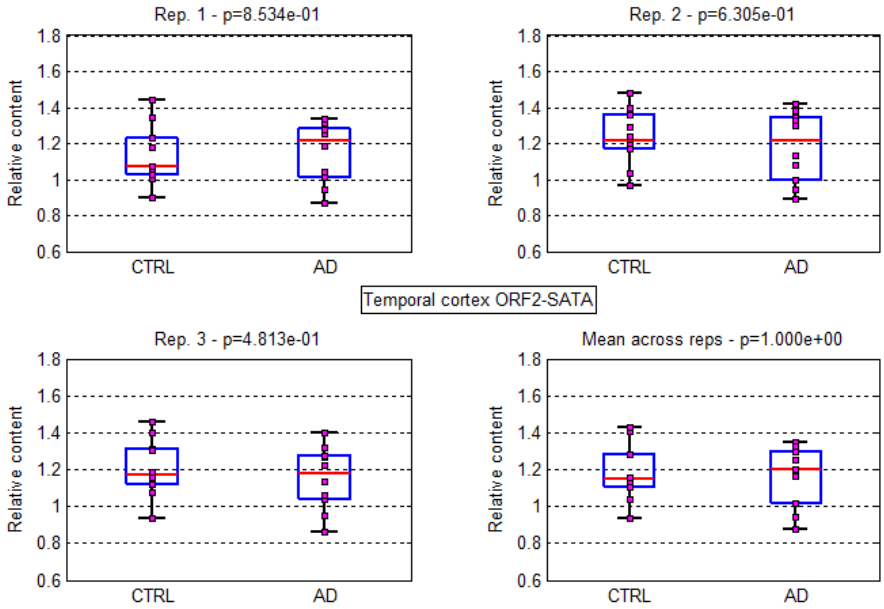


Figure 35: qPCR analysis of total L1 copy number in temporal cortexes of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe ORF2 and SATA. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

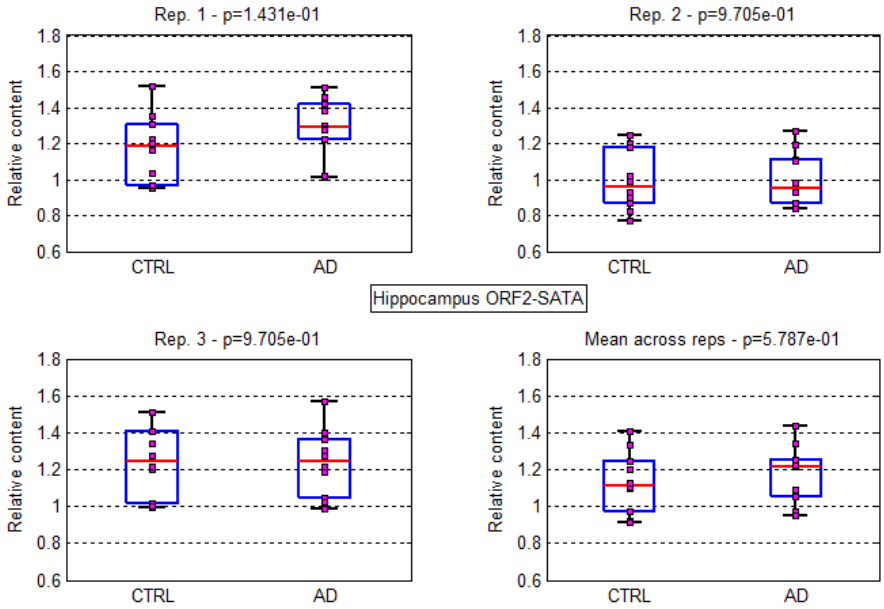


Figure 36: qPCR analysis of total L1 copy number in hippocampus of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe ORF2 and SATA. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

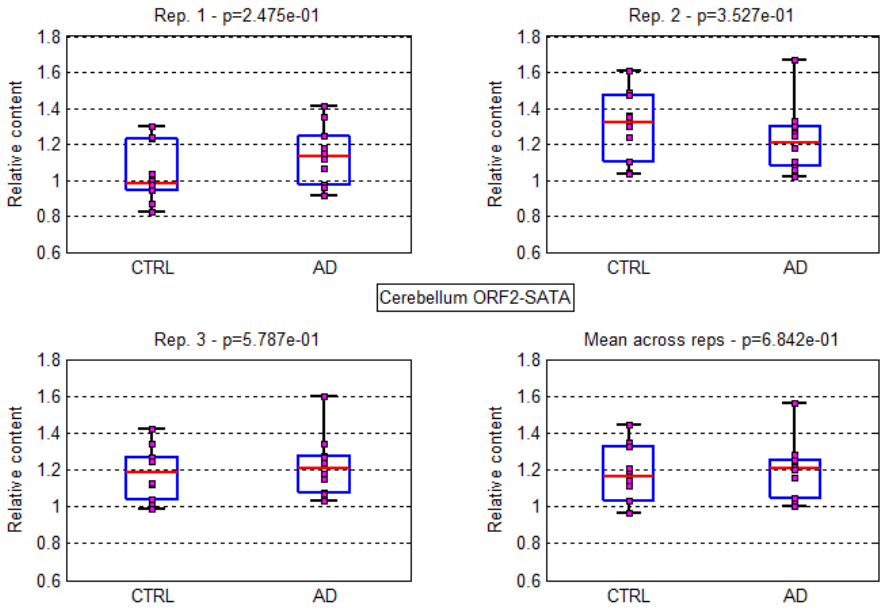


Figure 37: qPCR analysis of total L1 copy number in cerebellum of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe ORF2 and SATA. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

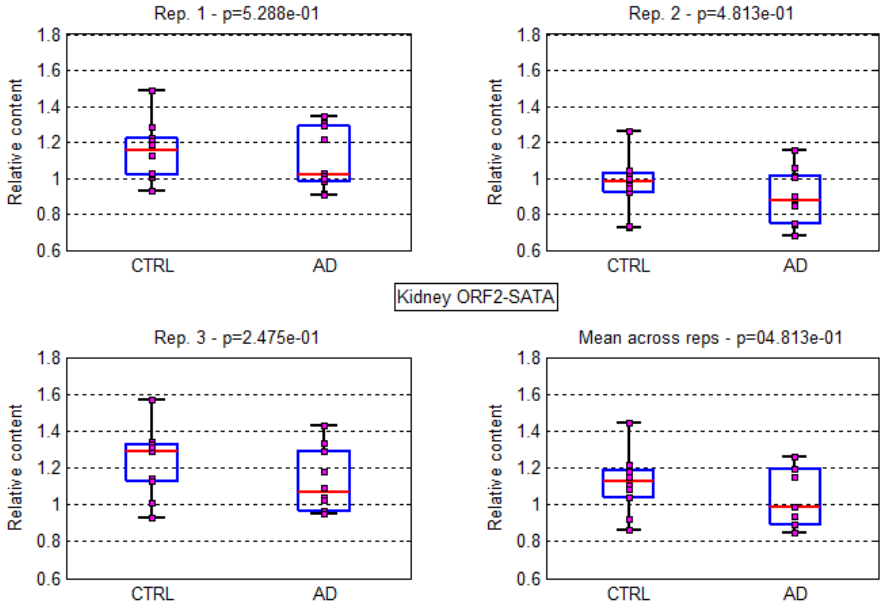


Figure 38: qPCR analysis of total L1 copy number in kidneys of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe ORF2 and SATA. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

Then we compared the relative amount of L1's full length forms present in all tissues using both the ORF1 and the 5'UTR assays. By using the ORF1 assay we observed a

significant decrease of full length L1 content in AD patients compared to controls in both the cortical tissues (frontal and temporal cortexes) (Figure 39 and Figure 40). Surprisingly, no differences were detected in the hippocampus (Figure 41), which is the most affected tissue in AD together with cerebral cortex, while a strong decrease was detected in the cerebellum (Figure 42), a tissue that is not primarily affected in AD. As expected, no significant differences were detected in the kidney (Figure 43).

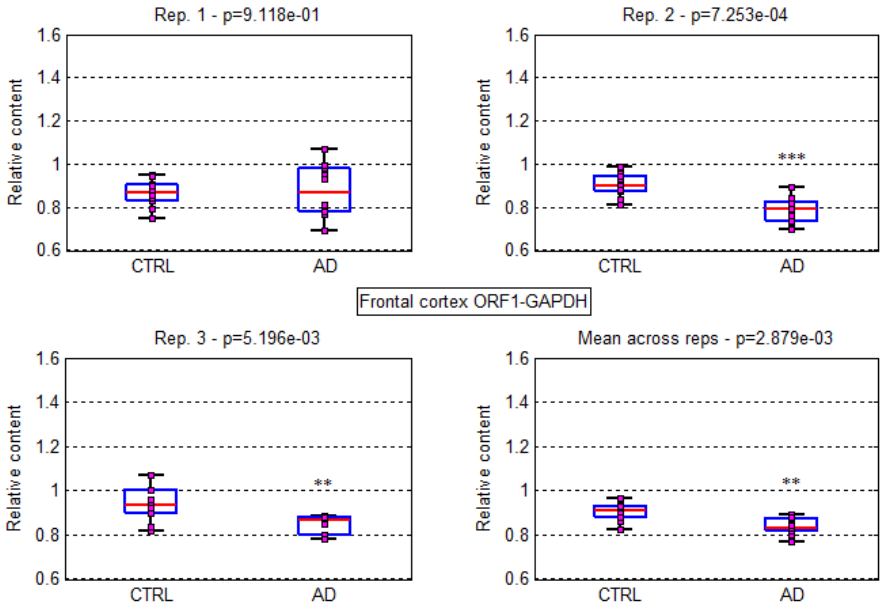


Figure 39: qPCR analysis of L1 full length copy number in frontal cortexes of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe ORF1 and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (p<0.01; ***p<0.001).**

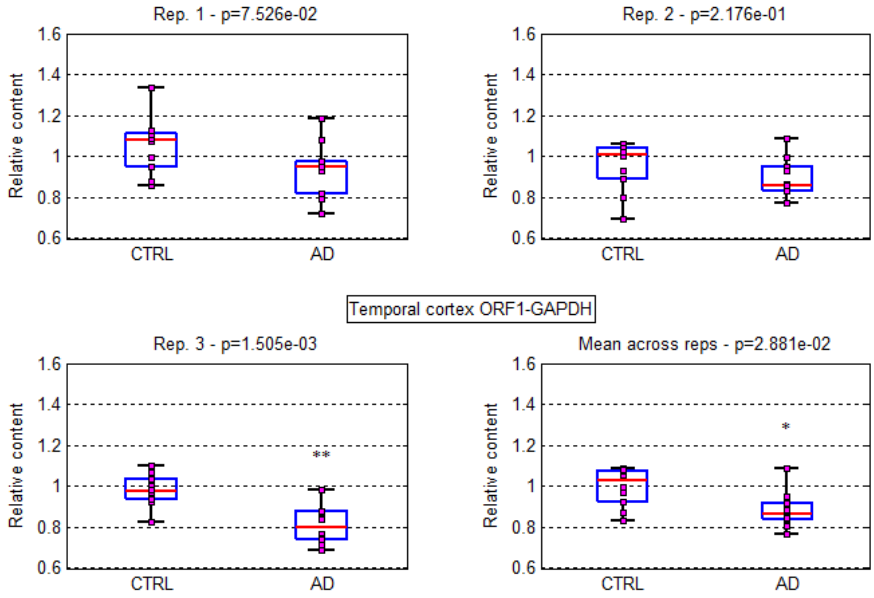


Figure 40: qPCR analysis of L1 full length copy number in temporal cortexes of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe ORF1 and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05; **p<0.01).

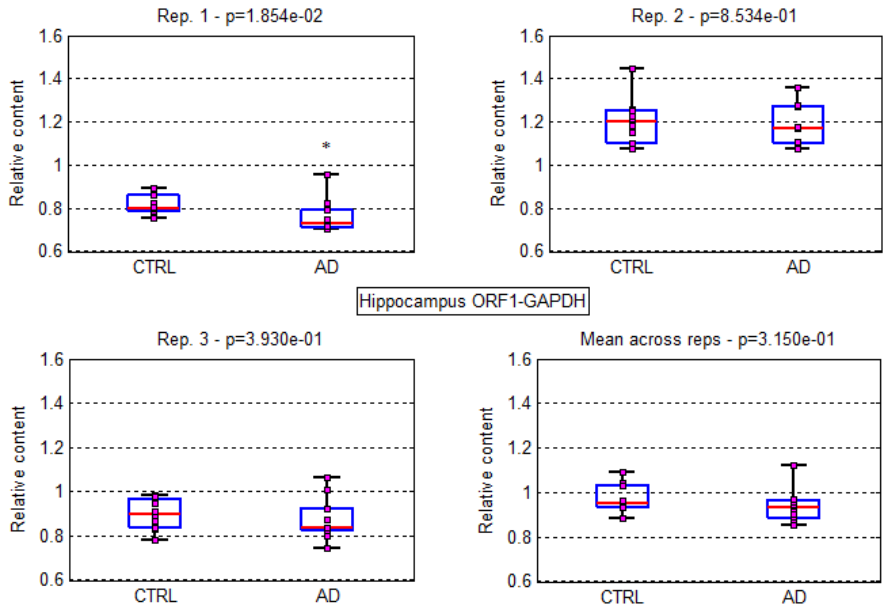


Figure 41: qPCR analysis of L1 full length copy number in hippocampus of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe ORF1 and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05).

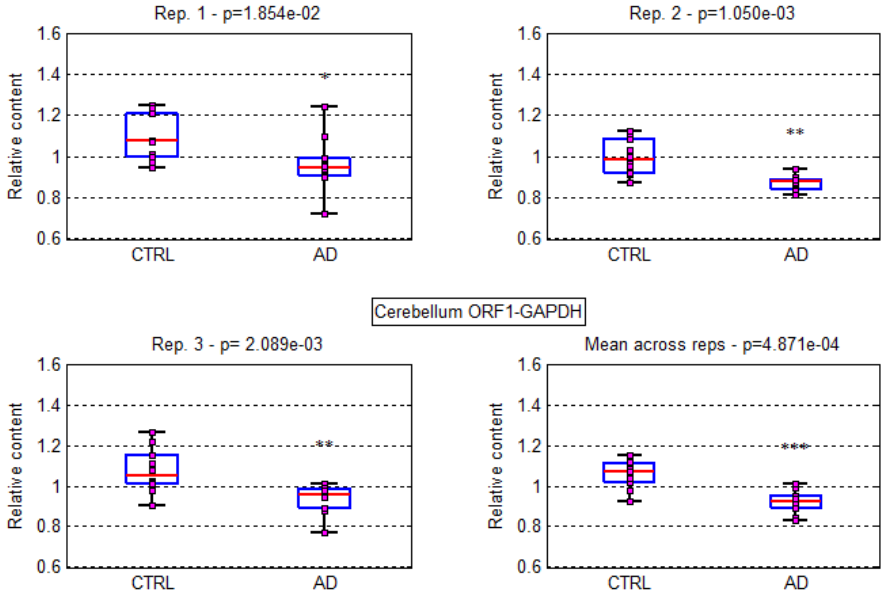


Figure 42: qPCR analysis of L1 full length copy number in cerebellum of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe ORF1 and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05; **p<0.01; ***p<0.001).

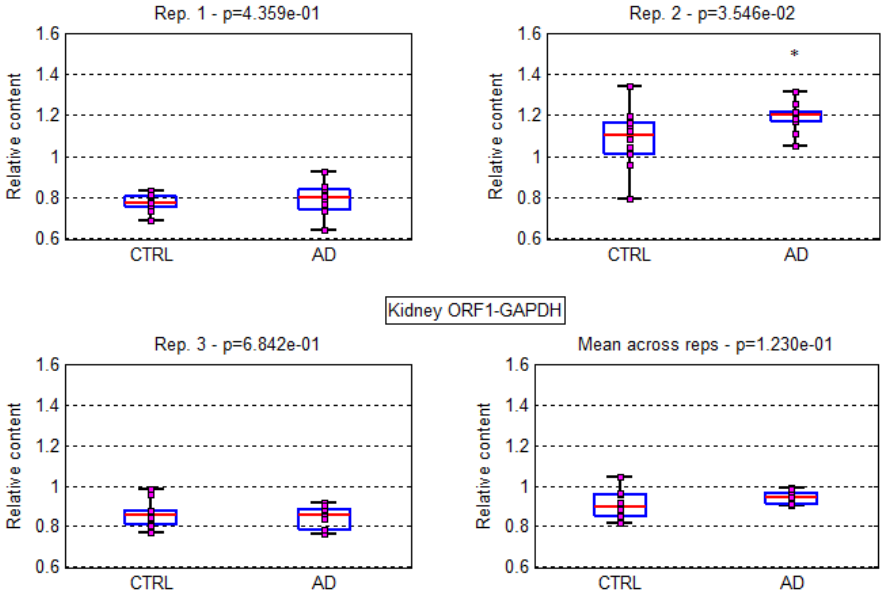


Figure 43: qPCR analysis of L1 full length copy number in kidneys of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe ORF1 and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05; **p<0.01; ***p<0.001).

When we performed the analysis of the full length L1 content with the 5'UTR Taqman probe, we observed a significant decrease in AD patients compared to controls at the level of frontal cortex (as for the ORF1 assay) (Figure 44), while in temporal cortex the difference was no longer present (Figure 45). No differences were detected in hippocampus (Figure 46), as with the ORF1 assay, while a significant difference was observed for both cerebellum and kidney (Figure 47 and Figure 48).

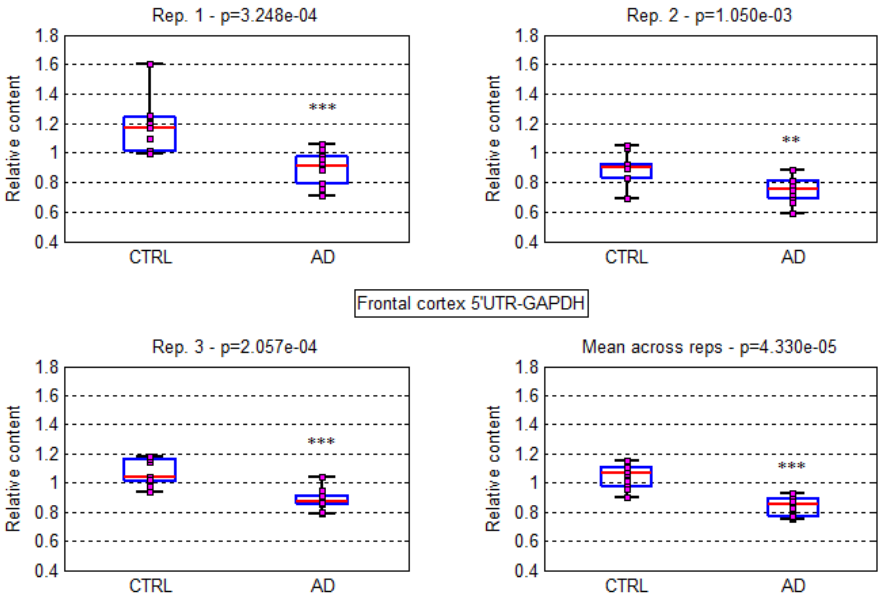


Figure 44: qPCR analysis of L1 full length copy number in frontal cortexes of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe 5'UTR and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (p<0.01; ***p<0.001).**

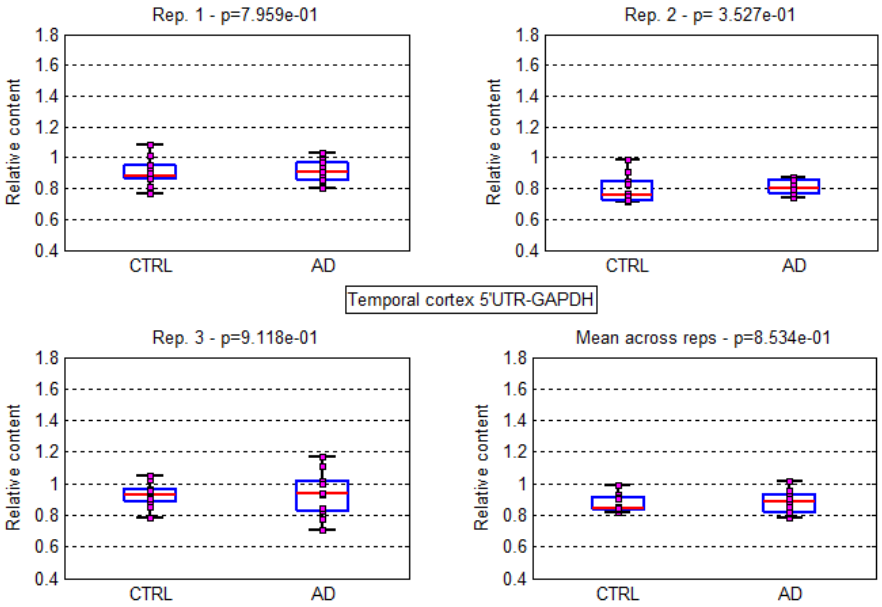


Figure 45: qPCR analysis of L1 full length copy number in temporal cortexes of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe 5'UTR and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

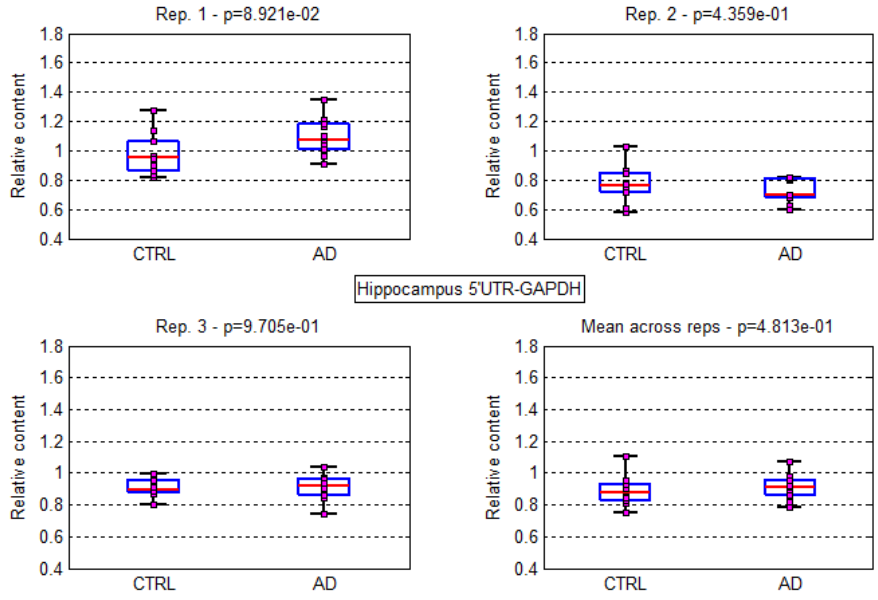


Figure 46: qPCR analysis of L1 full length copy number in hippocampus of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe 5'UTR and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

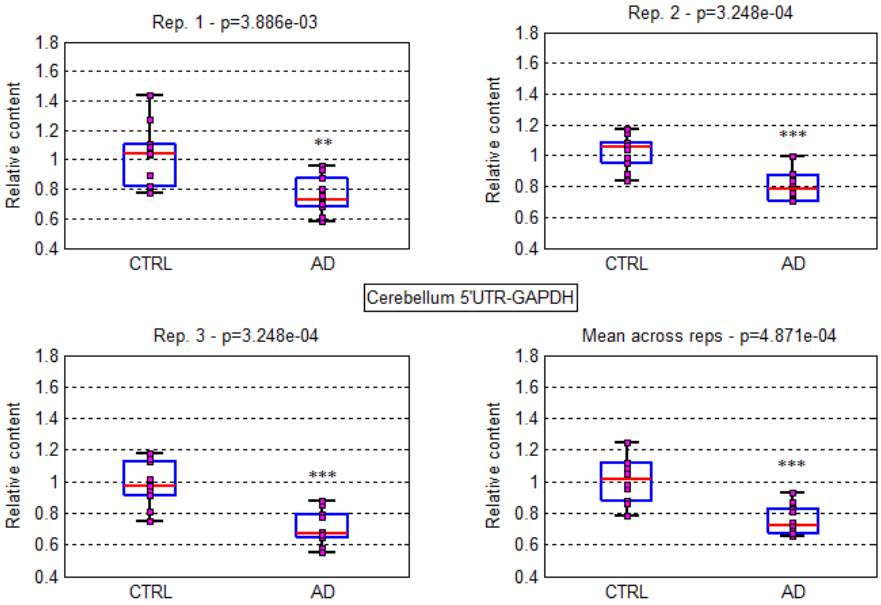


Figure 47: qPCR analysis of L1 full length copy number in cerebellum of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe 5'UTR and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (p<0.01; ***p<0.001).**

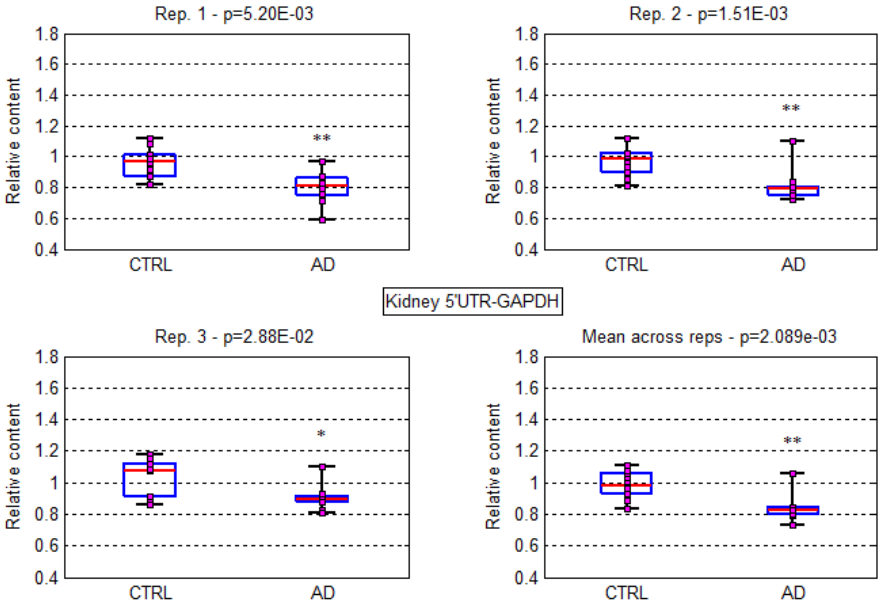


Figure 48: qPCR analysis of L1 full length copy number in kidneys of AD patients and controls. Relative quantification obtained by qPCR with the Taqman probe 5'UTR and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on CTRL (controls) and AD (affected patients). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05; **p<0.01).

L1 retrotransposition in a transgenic AD mouse model

By performing the copy number variation assay on human post mortem samples we observed a reduced L1 full length content in samples of AD patients compared to healthy controls. This decreased amount of L1 full length sequences in AD patients could be due to a deficient rate of retrotransposition during the embryonic development of these individuals, or could be an inherited trait, both able to make these individuals prone to the development of AD. Another cause of this decrease in L1 sequences could be simply cell death in neurons.

In order to investigate whether a change in the L1 rate of retrotransposition occurs at the first stage of the disease and triggers cell death, we developed new Taqman assays based on those one used for the human genome, to study L1 copy number variation in a severe and early onset transgenic AD mouse model: the TgCRND8 mouse.

Since the murine L1 5'UTR region contains different repetitive monomers which characterize different L1 subfamilies, we designed specific probes for the 5'UTR of each active mouse L1 type (A, Tf and Gf), allowing us to discriminate between the different mouse L1 subfamilies. The same approach was used to design assays specific for the 3'UTR region of each L1 subfamily, taking advantage of a polymorphic sequence present in the 3'UTR sequence. On the other hand, we also designed probes and primers complementary to the ORF2 region of the L1 sequence able to detect the total amount of L1 sequences, without discriminating between the different L1 families present in the mouse genome

In order to perform a relative quantification of the truncated and the full length L1 copies, as internal controls we used Taqman probes designed on non-mobile repetitive genomic sequences with high or low repeats number. In particular we designed an assay for a high copy number invariant sequence, the centromeric microsatellite sequence or MICSAT (about 500000 copies), and an assay for a low copy number control, the glyceraldehyde 3-phosphate dehydrogenase or GAPDH (362 copies).

The qPCR experiments were performed on samples from TgCRND8 and control mice, provided by Prof. Scarpa and Dr. Fuso (Sapienza University, Rome) (Table 14).

Recent works demonstrated that the bulk of L1 retrotransposition occurs during embryogenesis (Kano et al., 2009). To assess if in transgenic mice a perturbation in L1 copy number occurs before birth or during aging together with amyloid deposition, we tested P0 mice and mice at two stages of the adulthood (3 and 8 months).

Table 14: Mice samples used in the CNV experiment. We collected cortex, hippocampus and kidney samples from wt and TgCRND8 mice at various ages: P0, 3 months and 8 months.

Mice samples	P0	3 months	8 months
WT mice (cortex-hippocampus-kidney)	16	12	6
TgCRND8 mice (cortex-hippocampus-kidney)	16	12	4

Each assay was performed three times on all the samples, and graphs for all the three replicas and for mean values across replicas are reported.

L1 retrotransposition at P0

First of all we analyzed L1 copy number variation in mice at P0: in particular we analyzed cortex, hippocampus and kidney from 16 WT and 16 TgCRND8 mice. We first measured the content of all L1 sequences present in the mouse genome using the ORF2 assay, normalizing on the high copy number control MICSAT.

In the case of cortex samples (Figure 49) we were able to see a statistically significant higher amount of ~7% of L1 sequences in TgCRND8 mice compared to controls, while in hippocampus and kidney we didn't see any difference (Figure 50 and Figure 51).

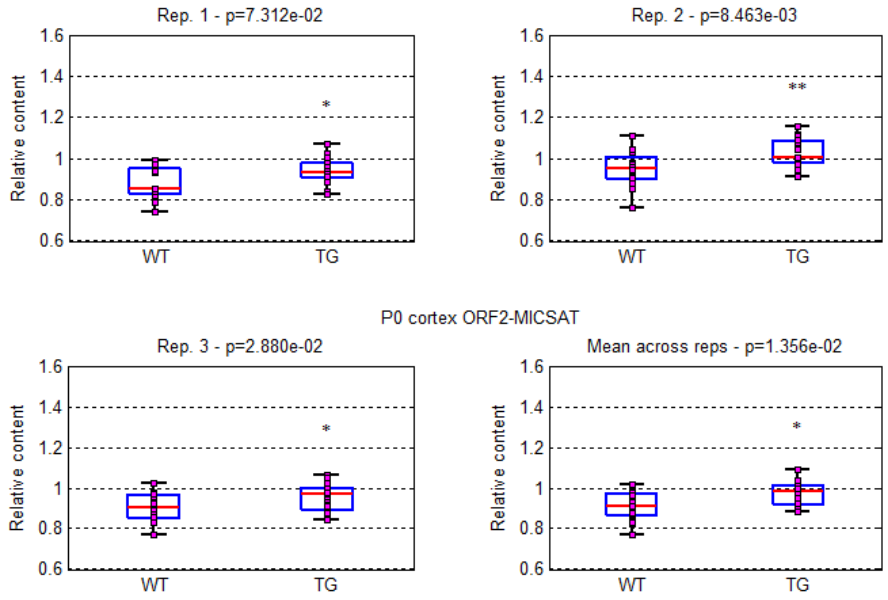


Figure 49: qPCR analysis of total L1 copy number in cortexes of TgCRND8 and control mice at P0. Relative quantification obtained by qPCR with the Taqman probe ORF2 and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05; **p<0.01).

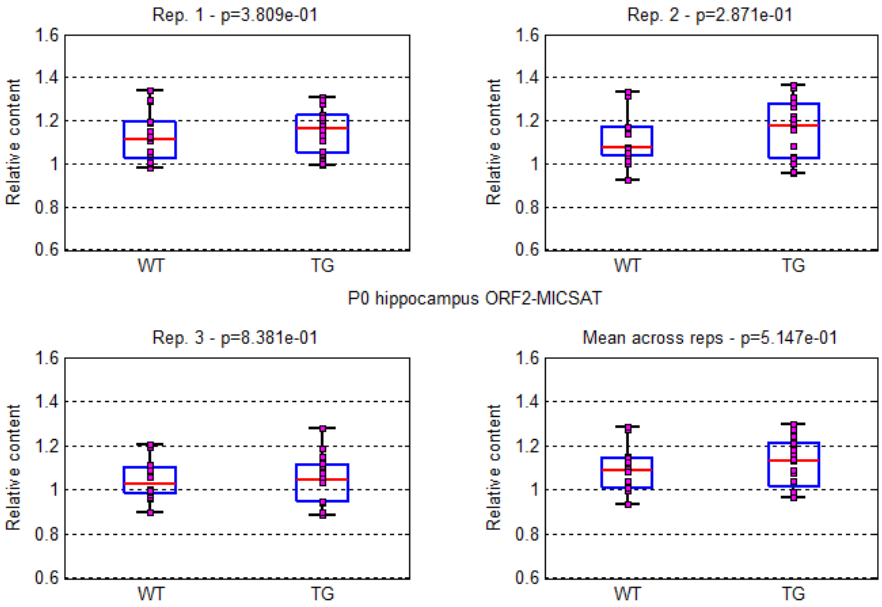


Figure 50: qPCR analysis of total L1 copy number in hippocampus of TgCRND8 and control mice at P0. Relative quantification obtained by qPCR with the Taqman probe ORF2 and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

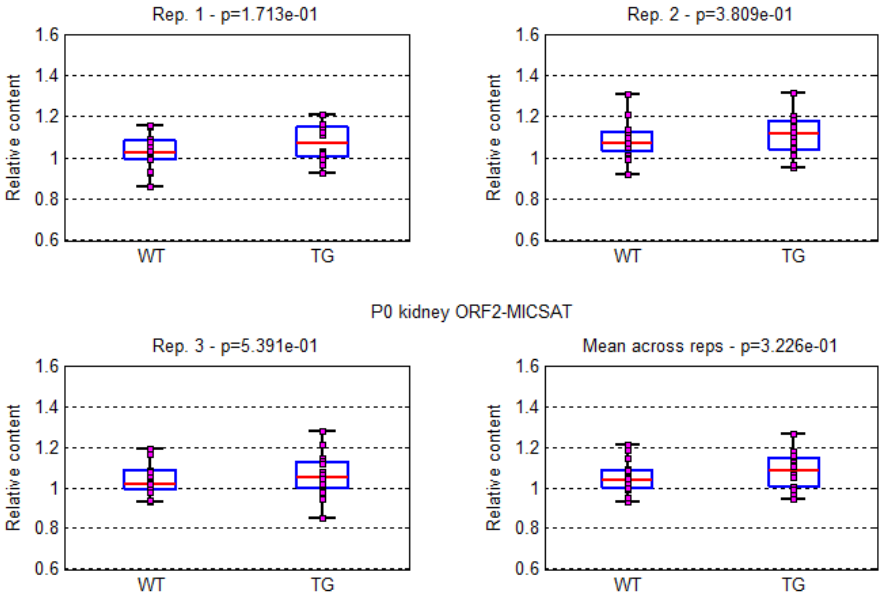


Figure 51: qPCR analysis of total L1 copy number in kidneys of TgCRND8 and control mice at P0. Relative quantification obtained by qPCR with the Taqman probe ORF2 and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

Since we observed a higher amount of L1s in the cortex of transgenic mice, we decided to go deeper in the study of L1 sequences, analyzing separately the three principal L1

families in the mouse cortex. The relative content of L1 elements from A, Tf and Gf families was measured using the 3'UTR-MICSAT/GAPDH assay.

In accordance with our first results, we observed a significant increase of L1 sequences for each L1 family in transgenic mice compared to controls. In particular we detected an increase of ~12% of L1-A elements, ~13% of L1-Tf elements and ~6% of L1-Gf elements (Figure 52, Figure 53 and Figure 54).

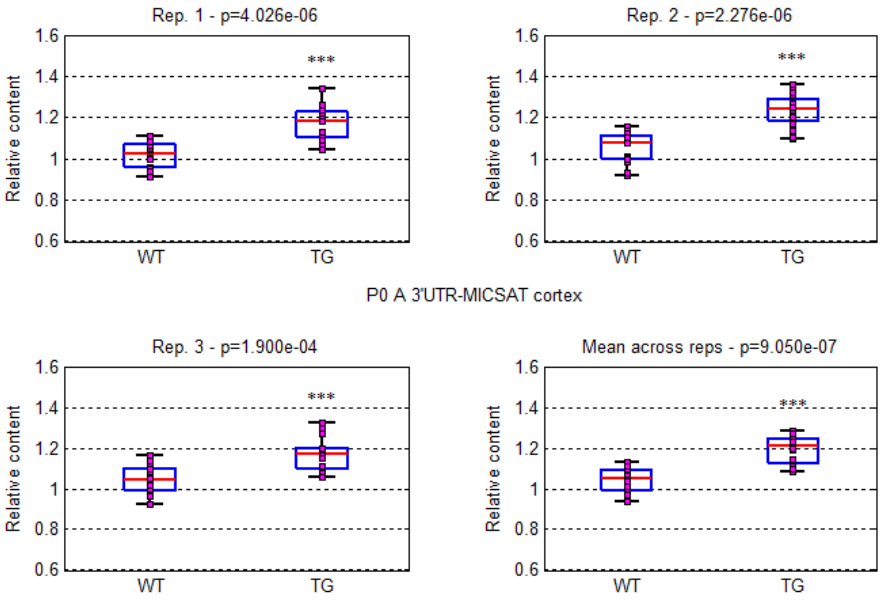


Figure 52: qPCR analysis of total L1-A copy number in cortexes of TgCRND8 and control mice at P0. Relative quantification obtained by qPCR with the Taqman probe 3'UTR-A and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.001).**

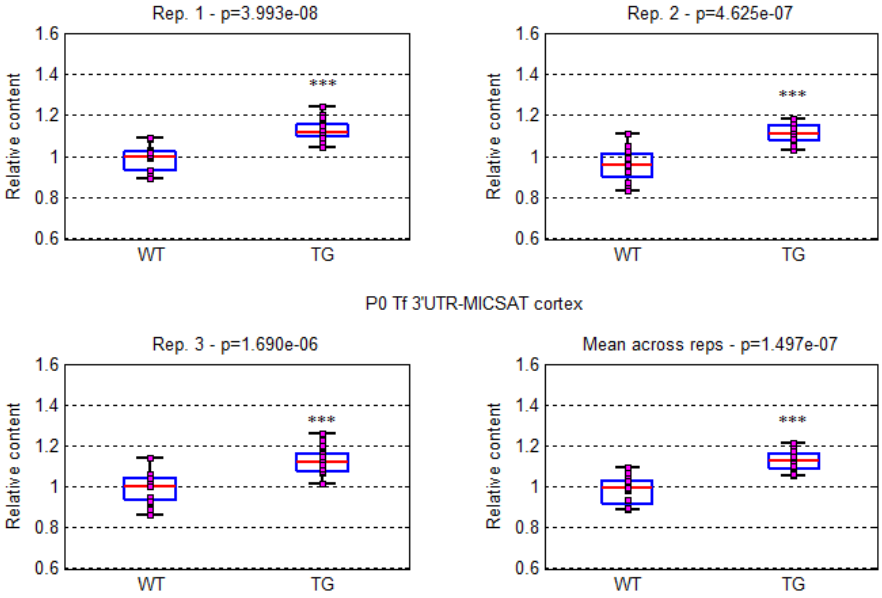


Figure 53: qPCR analysis of total L1-Tf copy number in cortexes of TgCRND8 and control mice at P0. Relative quantification obtained by qPCR with the Taqman probe 3'UTR-Tf and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.001).**

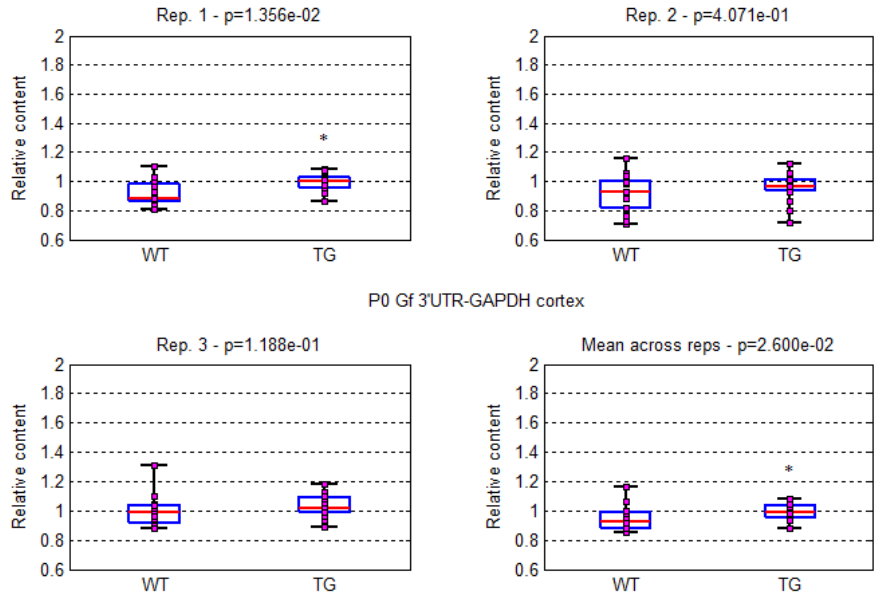


Figure 54: qPCR analysis of total L1-Gf copy number in cortexes of TgCRND8 and control mice at P0. Relative quantification obtained by qPCR with the Taqman probe 3'UTR-Gf and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05).

In order to understand if the higher amount of L1 elements detected in transgenic mice at P0 could be due to an increase of retrotransposition during embryogenesis, we

measured on these samples also the rate of retrotransposition with the 3'UTR-5'UTR assays for each single family (see chapter "Materials and methods").

Surprisingly, besides a strong increase in the rate of retrotranpsosition for the A family in transgenic cortexes, in the case of Tf family we observed even a dicrease in rate of retrotransposition, while in Gf family no differences were detectable (Figure 55, Figure 56 and Figure 57).

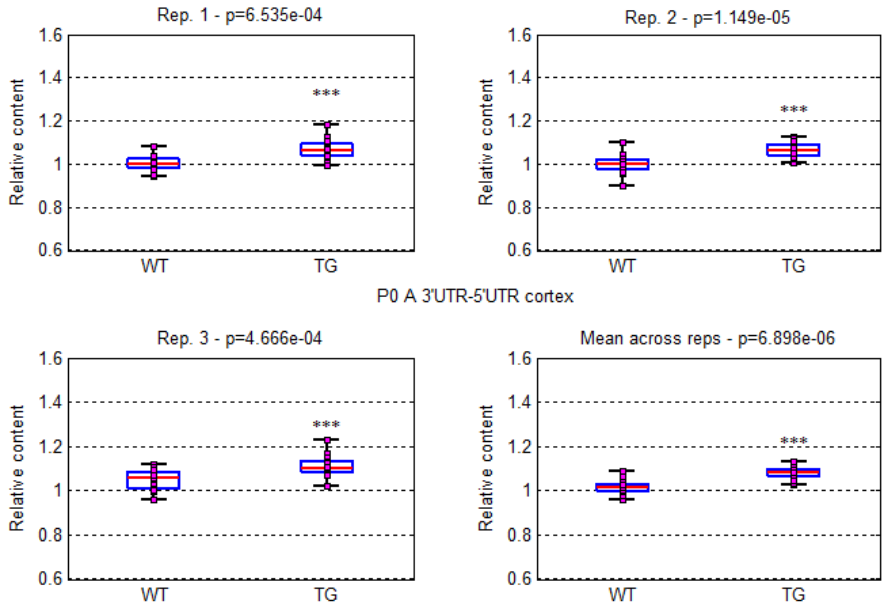


Figure 55: qPCR analysis of L1-A rate of retrotransposition in cortexes of TgCRND8 and control mice at P0. Relative quantification obtained by qPCR with the Taqman probes 3'UTR-A and 5'UTR-A. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.001).**

Figure 54

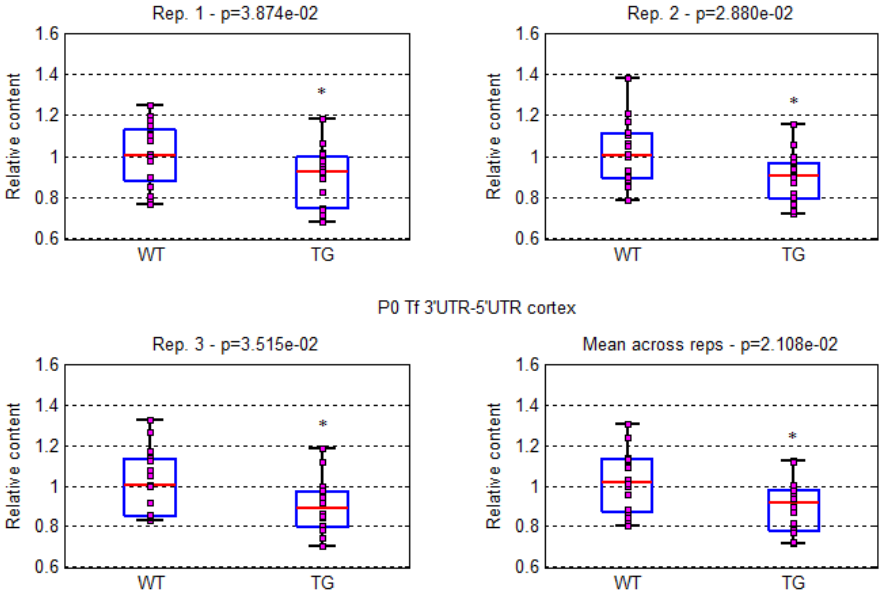


Figure 56: qPCR analysis of L1-Tf rate of retrotransposition in cortexes of TgCRND8 and control mice at P0. Relative quantification obtained by qPCR with the Taqman probes 3'UTR-Tf and 5'UTR-Tf. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (p<0.05).

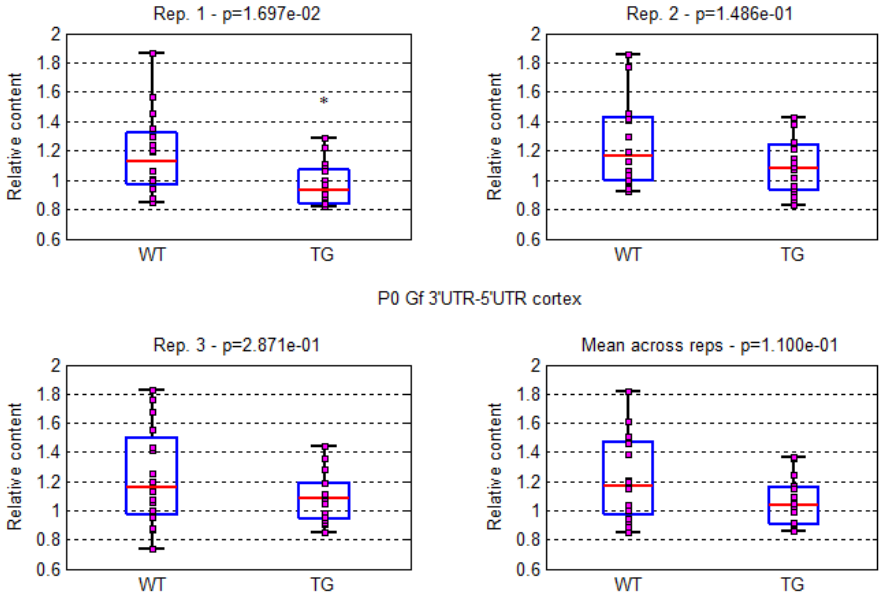


Figure 57: qPCR analysis of L1-Gf rate of retrotransposition in cortexes of TgCRND8 and control mice at P0. Relative quantification obtained by qPCR with the Taqman probes 3'UTR-Gf and 5'UTR-Gf. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (p<0.05).

L1 retrotransposition at 3 months of age

We observed a perturbation of L1 retrotransposition before P0. In order to verify if this variation was maintained or rescued during postnatal development, we analyzed the L1 content in cortex, hippocampus and kidney of 12 WT + 12 TgCRND8 mice at 3 months of age.

The total amount of L1 was evaluated by qPCR with the ORF2-MICSAT assay. Surprisingly we couldn't detect any difference in the total L1 content in cortex and kidney, but we observed a slight but significant increase (~7%) in the hippocampus of transgenic mice (Figure 58, Figure 59 and Figure 60).

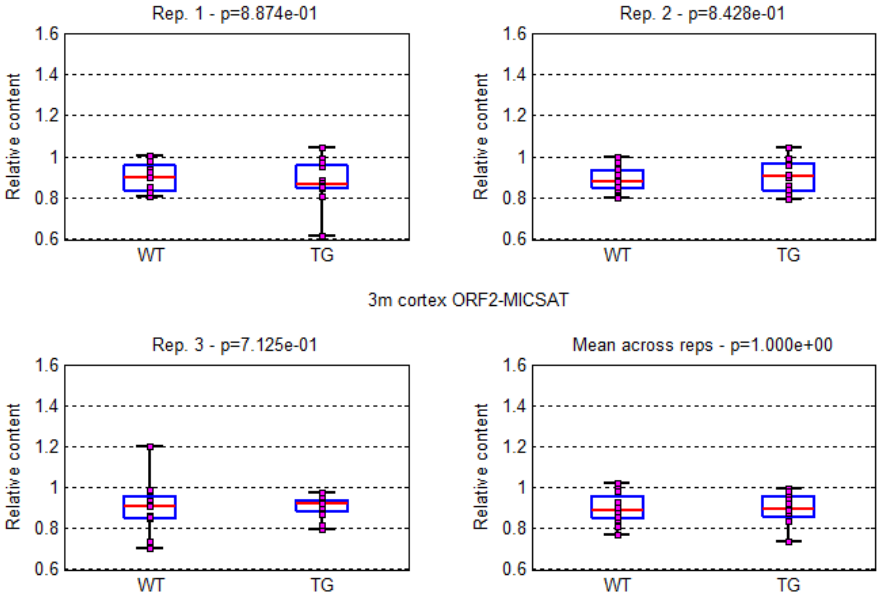


Figure 58: qPCR analysis of total L1 copy number in cortexes of TgCRND8 and control mice at 3 months. Relative quantification obtained by qPCR with the Taqman probe ORF2 and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

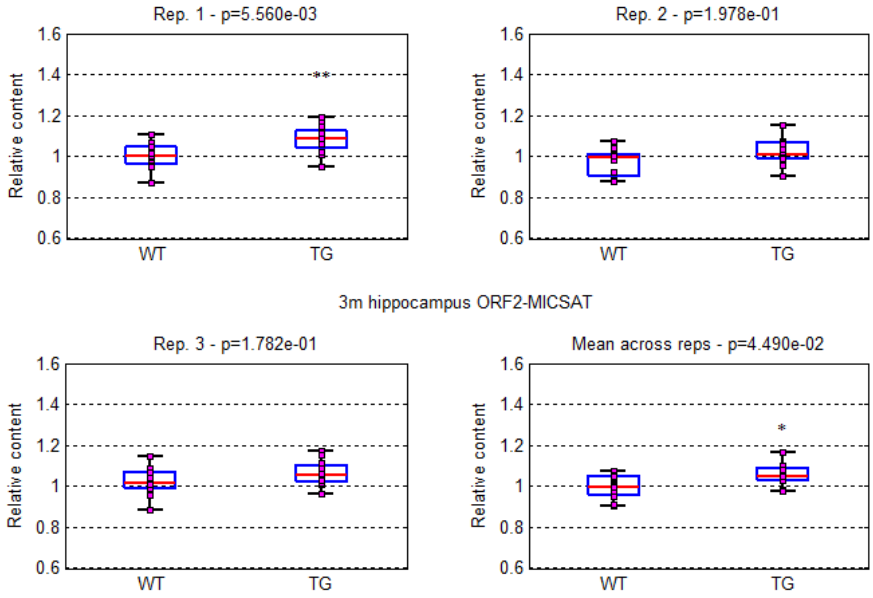


Figure 59: qPCR analysis of total L1 copy number in hippocampus of TgCRND8 and control mice at 3 months. Relative quantification obtained by qPCR with the Taqman probe ORF2 and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05; **p<0.01).

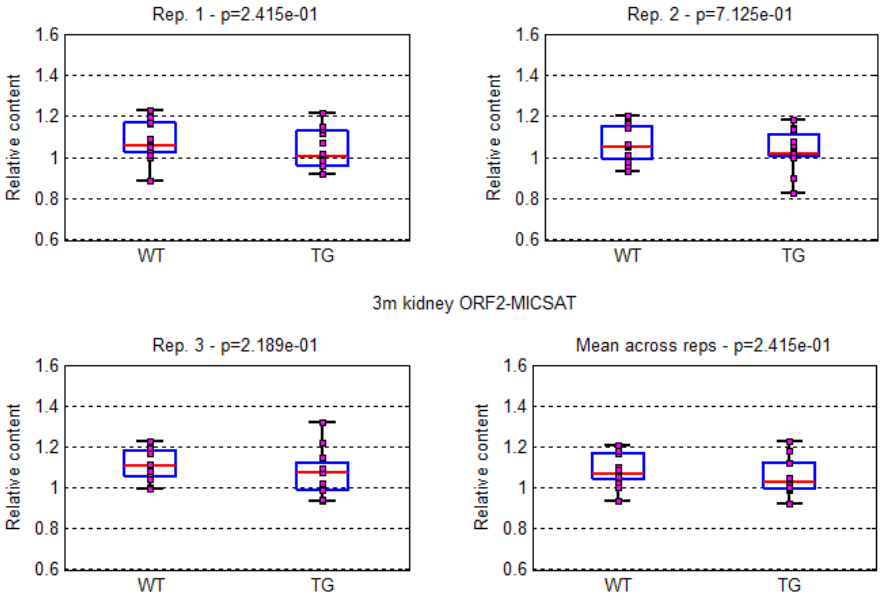


Figure 60: qPCR analysis of total L1 copy number in kidneys of TgCRND8 and control mice at 3 months. Relative quantification obtained by qPCR with the Taqman probe ORF2 and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

To further analyze L1 copy number variation in the transgenic hippocampus, we measured the three L1 families content only in this cerebral region. The relative content of L1 elements from A, Tf and Gf families was measured using the 3'UTR-MICSAT/GAPDH assay.

In accordance with our first results, we observed a significant increase (~9%) of L1 sequences for the A family in transgenic mice compared to controls (Figure 61), but no differences were detected for the Tf and the Gf families (Figure 62 and Figure 63).

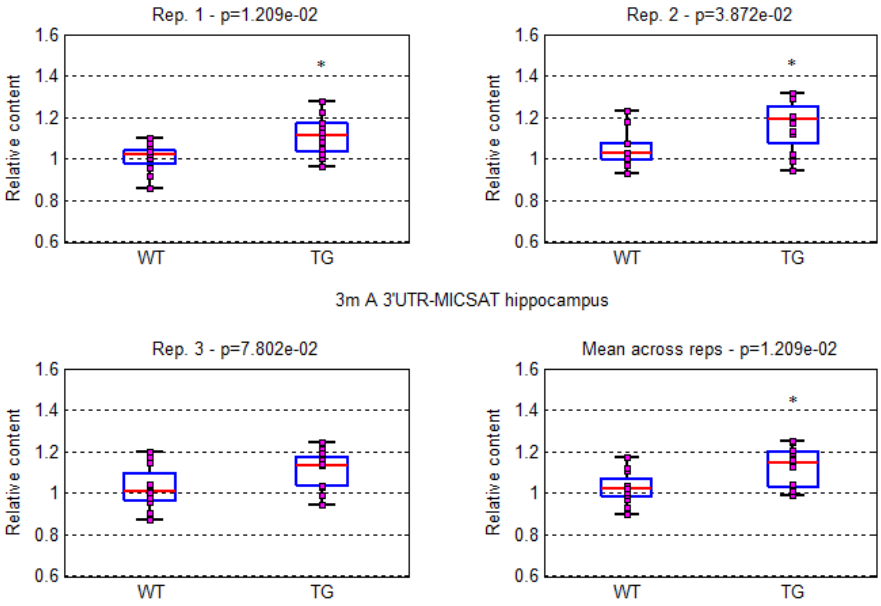


Figure 61: qPCR analysis of total L1-A copy number in hippocampus of TgCRND8 and control mice at 3 months. Relative quantification obtained by qPCR with the Taqman probe 3'UTR and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05).

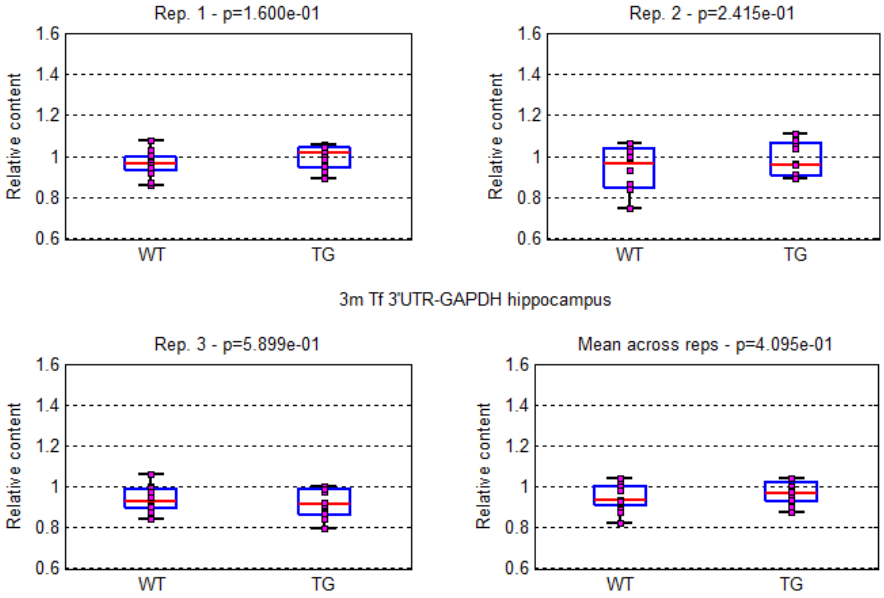


Figure 62: qPCR analysis of total L1-Tf copy number in hippocampus of TgCRND8 and control mice at 3 months. Relative quantification obtained by qPCR with the Taqman probe 3'UTR and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

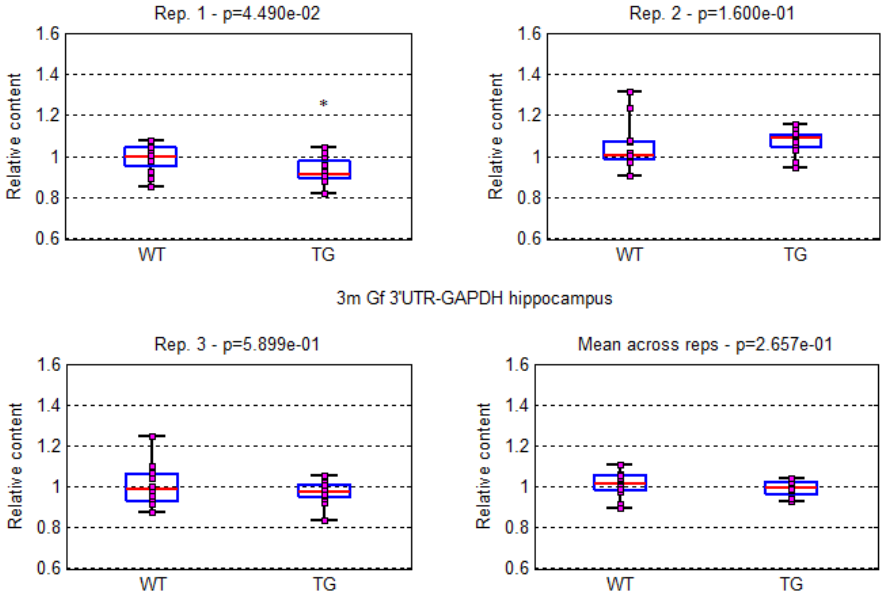


Figure 63: qPCR analysis of total L1-Gf copy number in hippocampus of TgCRND8 and control mice at 3 months. Relative quantification obtained by qPCR with the Taqman probe 3'UTR and GAPDH. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05).

In order to understand if the higher amount of L1-A elements detected in transgenic mice at 3 months could be due to an increase of retrotransposition, we measured on these samples also the rate of retrotransposition with the 3'UTR-5'UTR assays for each single family (see chapter "Materials and methods").

As expected, considering the previous results, we observed a strong increase in the rate of retrotranpsosition for the A family in transgenic hippocampus (Figure 64), while for Tf and Gf families no differences were detectable (Figure 65 and Figure 66).

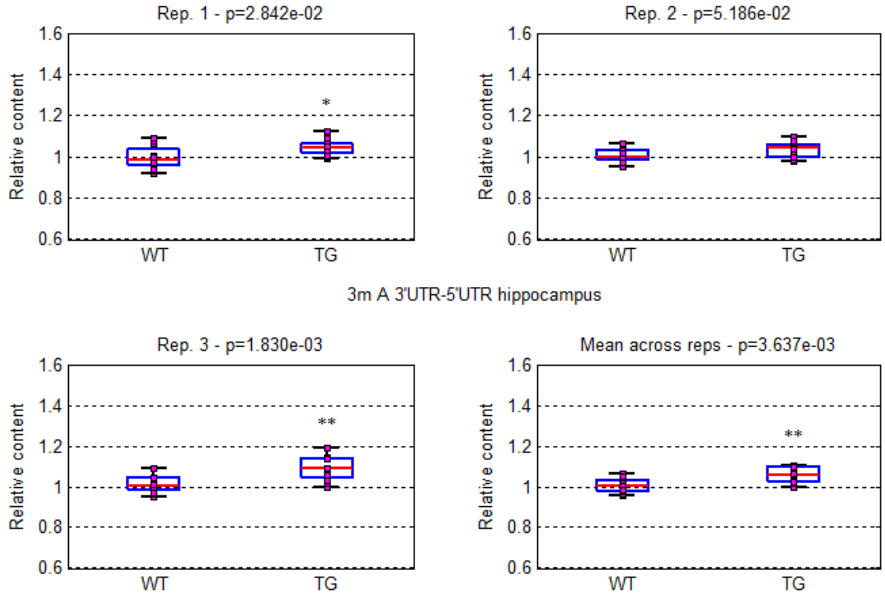


Figure 64: qPCR analysis of L1-A rate of retrotransposition in hippocampus of TgCRND8 and control mice at 3 months. Relative quantification obtained by qPCR with the Taqman probes 3'UTR-A and 5'UTR-A. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values. (*p<0.05; **p<0.01).

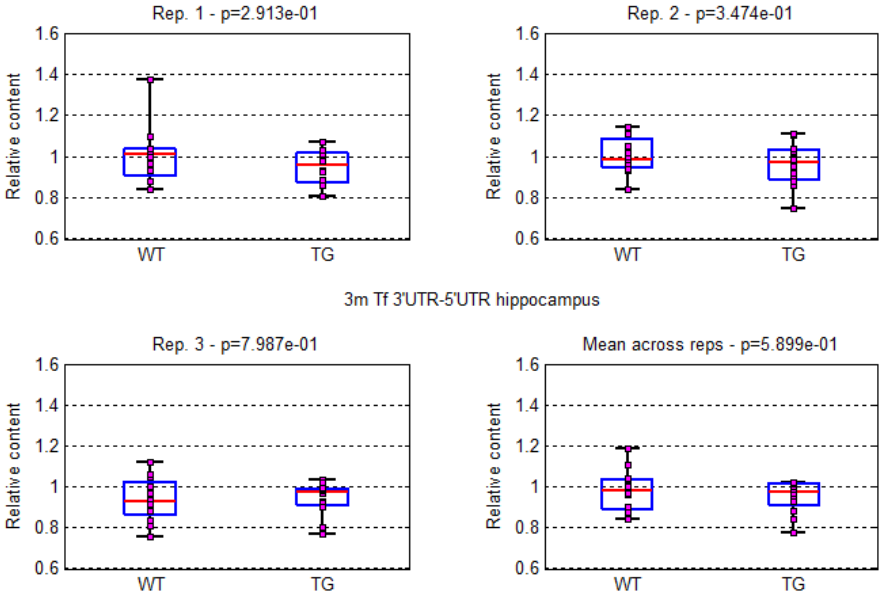


Figure 65: qPCR analysis of L1-Tf rate of retrotransposition in hippocampus of TgCRND8 and control mice at 3 months. Relative quantification obtained by qPCR with the Taqman probes 3'UTR-Tf and 5'UTR-Tf. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

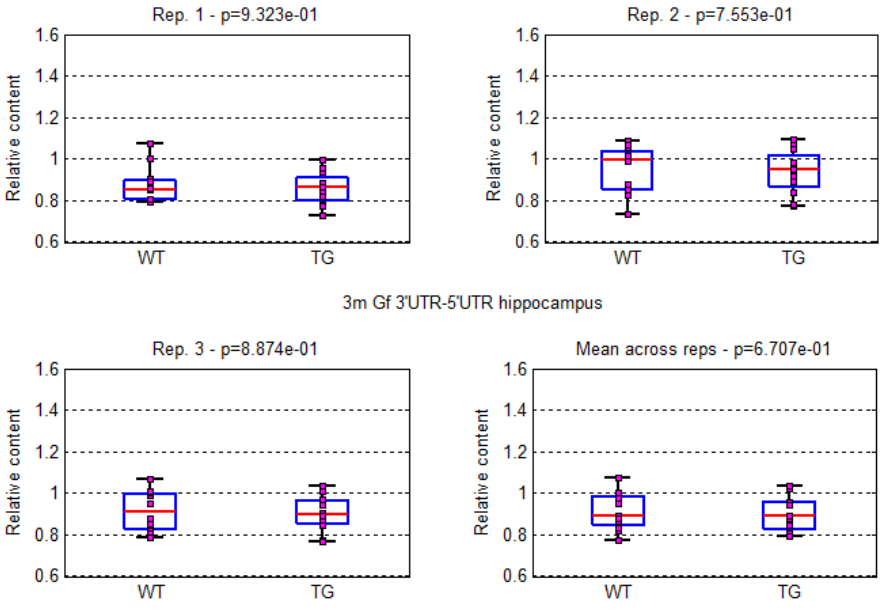


Figure 66: qPCR analysis of L1-Gf rate of retrotransposition in hippocampus of TgCRND8 and control mice at 3 months. Relative quantification obtained by qPCR with the Taqman probes 3'UTR-Gf and 5'UTR-Gf. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

L1 retrotransposition at 8 months of age

Our experiments showed the presence of a higher amount of L1 elements in the cortex of transgenic mice at P0, but this increase was not present at 3 months. On the other hand at 3 months we observed an increase in L1 content at the level of the hippocampus. The increase seen in the cortex at P0 could be linked to an abnormal L1 retrotransposition occurred in the process of cortical neural differentiation during the embryonic development. This accumulation at P0 could have led to cell death and only those cells in which L1 was less active survived to adulthood. At the same time, the accumulation of L1 sequences in the hippocampus at 3 months could be due to an abnormal L1 retrotransposition occurred in the process of hippocampal neurogenesis during adulthood.

In order to understand if neurodegeneration could be one of the possible mechanisms responsible for the lack of difference in L1 content in cortexes at 3 months, we decided to test aging mice at 8 months. TgCRND8 mice exhibit marked premature mortality, with only approximately 50% of mice reaching 9–10 months. At 8 months transgenic mice present cerebral amyloid depositions in the hippocampus, neocortex, but also in the cerebellum and brainstem, together with severe gliosis, neuritic dystrophy and vascular depositions.

The total amount of L1 sequences was measured in cortex, hippocampus and kidney of 6 WT + 4 TgCRND8 mice at 8 months of age. In all the three tissues we did not observe any difference between transgenic and control mice (Figure 67, Figure 68 and Figure 69). This result may support the hypothesis that neurodegeneration, parallel to aging and progression of the disease, depletes L1 content in the hippocampus of transgenic mice, preventing us from detecting differences between TgCRND8 mice and controls at 8 months.

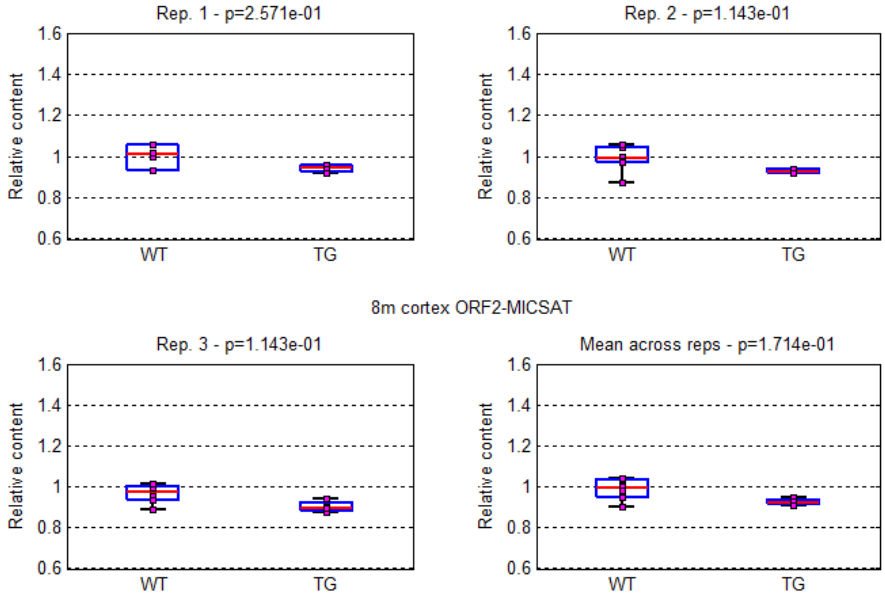


Figure 67: qPCR analysis of total L1 copy number in cortexes of TgCRND8 and control mice at 8 months. Relative quantification obtained by qPCR with the Taqman probe ORF2 and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

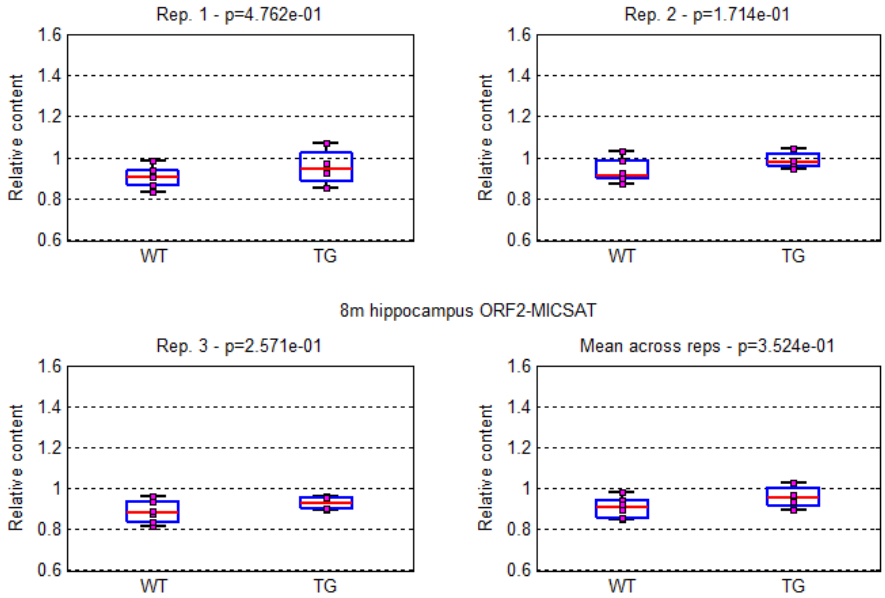


Figure 68: qPCR analysis of total L1 copy number in hippocampus of TgCRND8 and control mice at 8 months. Relative quantification obtained by qPCR with the Taqman probe ORF2 and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

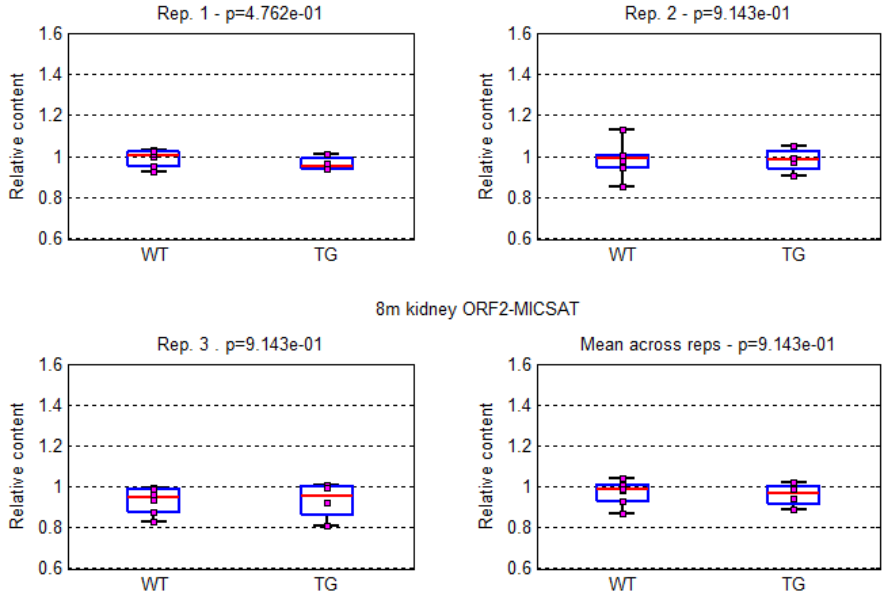


Figure 69: qPCR analysis of total L1 copy number in kidneys of TgCRND8 and control mice at 8 months. Relative quantification obtained by qPCR with the Taqman probe ORF2 and MICSAT. Boxplots for all the three replicas and the mean across them are reported, with the p-value deriving from the Wilcoxon rank sum statistical test performed on WT (control mice) and TG (TgCRND8 mice). Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles and the p-values.

SPlinkerette Analysis of Mobile elements (SPAM)

To study L1 insertion sites we developed a new technique based on splinkerette PCR (spPCR), a widely used PCR approach to clone flanking DNA regions of repetitive elements in the genome. In this technique, genomic DNA is fragmented (usually by enzymatic digestion) and ligated to a synthetic double stranded adaptor (splinkerette). Two rounds of PCR are then performed to amplify the genomic sequence between the repetitive element and the splinkerette, followed by a sequencing reaction.

In order to avoid amplification biases due to the genomic fragmentation by enzymatic digestion, we substituted it with genomic DNA shearing by sonication (see chapter "Materials and methods" for the detailed protocol).

At first we tested the SPAM technique on three DNA samples extracted from the temporal cortex of three human healthy donors. DNA samples were treated as reported in chapter "Materials and methods", and run in a MiSeq sequencing 2x300bp in multiplex.

In our laboratory the bioinformaticians developed a precise and reproducible bioinformatics pipeline with a specific nomenclature:

Reads: R1 and R2, single reads coming from paired end sequencing (forward and reverse sequences).

Fragment: assembled reads (R1 + R2).

MapFragment: fragment containing an L1 sequence and a mappable unique sequence.

MapCluster: at least three MapFragments in at least one sample.

Integration Site (IS): genomic region where one or more MapFragment have been mapped.

Annotated Integration site (AIS): integration site located at less than 1000 bp far from an L1 in the database.

Novel Integration site (NIS): integration site located at more than 1000 bp far from an L1 in the database.

We performed a quality filtering (Phred score for each nucleotide) of the reads (R1 and R2) and we aligned the corresponding R1 and R2 to form a fragment, discarding those pairs that didn't align with at least 30bp and 80% identity. We selected the fragments containing the forward primer (specific for the adaptor) and the human L1 sequence using blastn. Once discarded the fragments that did not contain the repeat, we trimmed the portions of the fragments that overlapped with the forward primer and the L1 sequence, and retained only the genomic sequences longer than 30 nucleotides. These

sequences were aligned with blastn against the reference human genome (Hg19) and defined as MapFragments only if they had a unique alignment with an identity >50% (all data reported in Table 15).

Table 15: Step-by-step analysis of the human SPAM samples. Reads are the total forward and reverse sequences obtained by sequencing. Fragments are the total assembled reads (100%) that underwent the following trimming, based on alignment to the L1 reference sequence and to the reference human genome (Hg19). Mapclusters are defined by at least three MapFragments.

HUMAN samples	Reads	Fragments	Fragments aligned to L1 Ref seq	MapFragments aligned to Hg19	MapClusters (at least 3 mapfragments)
Sample 1	R1: 2812142 R2: 2812142	1101323 (100%)	1050782 (95%)	188592 (17 %)	1715
Sample 2	R1: 2583156 R2: 2583156	976218 (100%)	928415 (95 %)	126822 (12 %)	1599
Sample 3	R1: 2880251 R2: 2880251	1043915 (100%)	990220 (95%)	137412 (13 %)	1599

In order to identify an IS as annotated or novel, we created an L1-database, collecting all the L1 genomic coordinates reported in literature, Repeat Masker and db RIP, and this allowed us to obtain a list of 8930 full length L1 sequences (length ≥6000bp) and 850898 total L1 sequences (Baillie et al., 2011; Bundo et al., 2014; Evrony et al., 2012; Ewing et al., 2010; Huang et al., 2010; Iskow et al., 2010; Shukla et al., 2013; Stewart et al., 2011; Xing et al., 2009).

Then we decided to define an AIS as an IS present at less than 1000 bp far from an L1 in the database, and a NIS as an IS located at more than 1000 bp far from an L1 in the database, both determined by at least 3 MapFragments (Table 16).

Table 16: Total number of AIS and NIS for each sample, defined by at least 3 MapFragment. The two values are reported also normalized by the initial total amount of MapFragments. MF, MapFragment. The ratio of NIS and AIS is reported.

HUMAN samples	AIS Defined by at least 3 MapFragment	NIS Defined by at least 3 MapFragment	Normalized AIS (AIS/MF)*10000	Normalized NIS (NIS/MF)*10000	NIS/AIS
Sample 1	1504	211	79.7	11.2	0.1
Sample 2	1422	177	112.1	14.0	0.1
Sample 3	1452	147	105.7	10.7	0.1

For both AIS and NIS we obtained IS specific for each single patient, and IS common between the three (Figure 70). In the case of AIS, ~56% of integration sites were common, while only the ~2% was common for NIS.

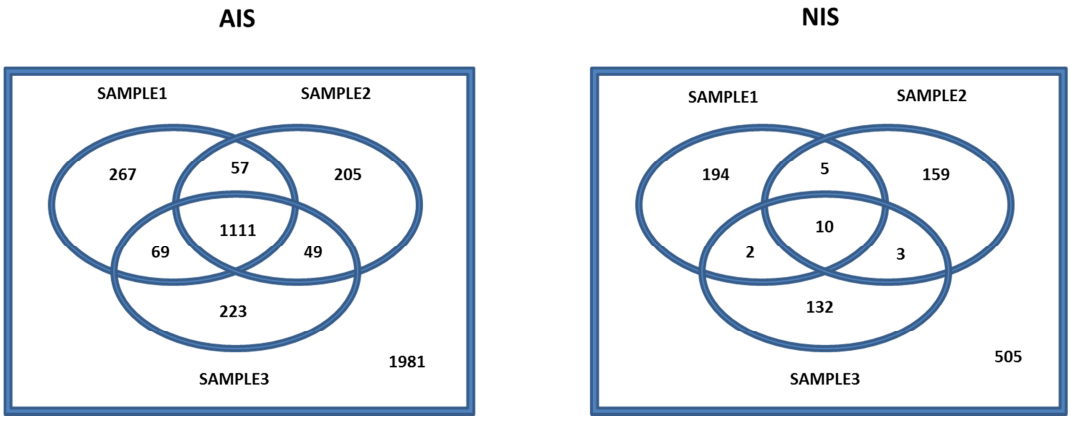


Figure 70: Venn diagram of common and exclusive AIS and NIS obtained for each sample.

Then a gene ontology study of the total AIS and NIS detected in the three samples was performed using the Gene Functional Classification Tool DAVID: from this analysis it came out an enrichment of IS associated genes involved in brain functions, such as neuron development and differentiation. Surprisingly, in the case of genes involved in some brain functions such as axon development and guidance, we observed the presence of only NIS, suggesting that L1 mobilization may have a role in neuron and axon plasticity (Figure 71).

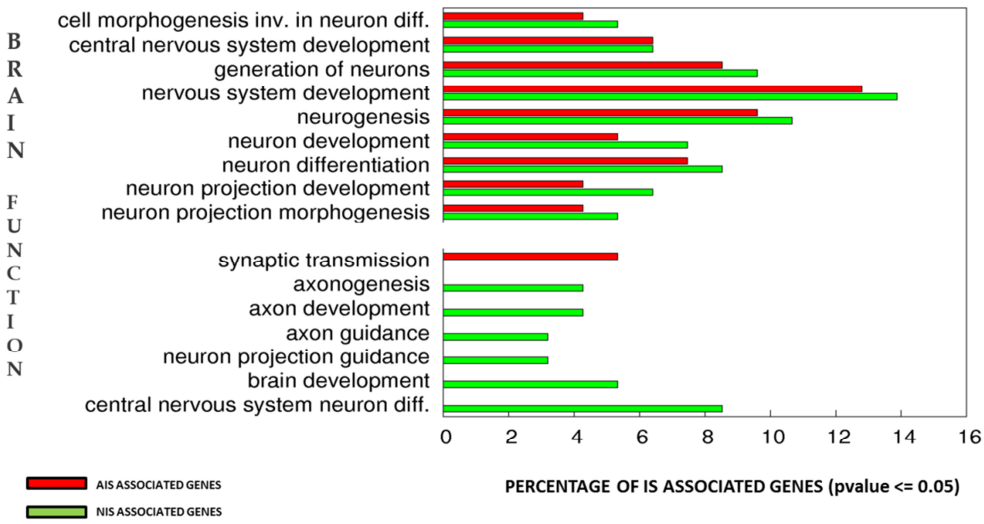


Figure 71: Bar diagram of Gene Ontology analysis performed with DAVID on detected AIS and NIS.

After the bioinformatics analysis, we deepened the investigation of some AIS and NIS. In particular, we chose 10 integration sites for each sample that we found to be present in exons, introns, or in brain specific gene loci. We analyzed each of these integration sites (annotated and novel) looking for the presence of the adaptor sequence, the L1 5'UTR sequence, and mapping the genomic portion in the human reference sequence.

We decided to select a group of 8 IS to be validated by PCR. These 8 IS were: 3 different NIS defined by 5-10 MapFragments and found in only one sample, 3 NIS defined by 1 MapFragment and found in only one sample, 1 NIS defined by more than 100 MapFragment and found in only one sample, and finally 1 AIS defined by more than 100 MapFragment and found in all the three samples.

For each of these 8 IS we performed a first round of PCR and a nested PCR using two couples of primers with the forward primers designed on the genomic DNA at the insertion site, and the reverse primers designed against the very beginning of the L1 5'UTR (example of an analyzed sequence with designed primers is reported in Figure 72).

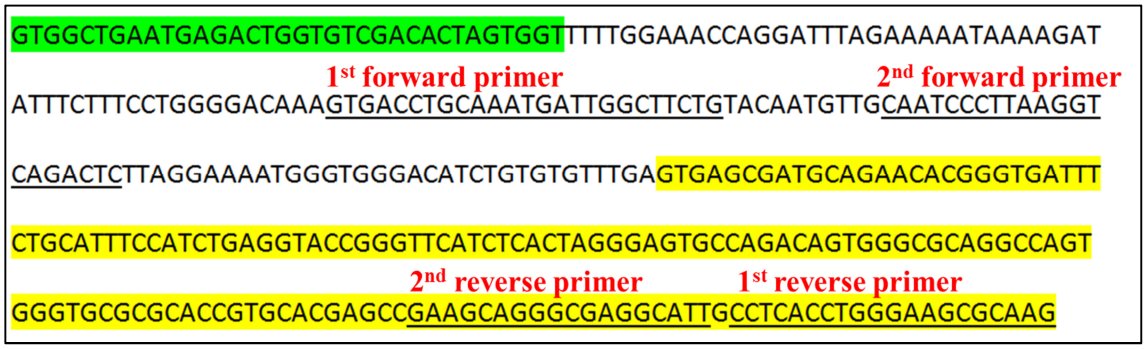


Figure 72: Analysis of 1 NIS defined by 5-10 MapFragments present only in sample 1. The sequence highlighted in green represents the final part of the ligated adaptor. The sequence highlighted in yellow represents the very first part of the L1 5'UTR. The not-highlighted sequence corresponds to the genomic part in which the integration occurred. The underlined portions are the sequences on which the two couples of primers for the validation PCRs were designed.

Here we report the validations of one NIS with 5-10 MapFragments detected only in sample 1 (amplicon length: 190bp), one NIS with 1 MapFragment detected only in sample 3 (amplicon length: 105bp), and one AIS with 100 MapFragments detected in all the samples (amplicon length: 242bp) (Figure 73).

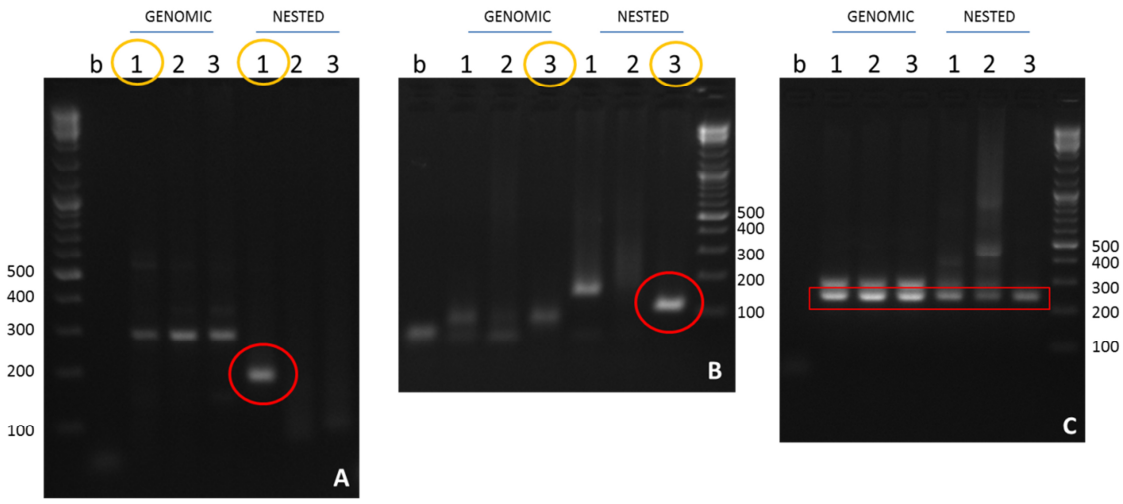


Figure 73: IS validations. Electrophoretic run on agarose gel of the validation PCRs of two NIS and one AIS detected by the bioinformatic analysis of the SPAM experiment. A: NIS with 5-10 MapFragments detected only in sample 1. B: NIS with 1 MapFragment detected only in sample3. C: AIS with 100 MapFragment detected in all the three samples. Specific bands are indicated in red.

The two NIS were confirmed with a specific band present in the SPAM product of the corresponding DNA sample, but no bands, except for some aspecific products, were detected in the PCR performed on the genomic DNA, probably because these NIS were poorly represented and not detectable in a pool of genomes. On the other hand, the band specific for the AIS was, as expected, present in all the three samples and in both the PCRs performed on the genomic DNA and on the SPAM product.

L1 insertion sites in frontal cortex and kidney of AD patients and controls

In order to verify whether in AD patients there is a different pattern of distribution of L1 elements, we applied the SPAM technique to samples of frontal cortex and kidney of 3 AD patients and 3 controls (from the Brazilian cohort).

DNA samples were treated as reported in the chapter "Materials and methods", and underwent a first preliminary run with Illumina MiSeq sequencing 2x300bp in multiplex (further sequencing runs will be necessary in the future to complete the analysis with a deeper coverage and with 2 more samples for each group).

We applied on these samples the same pipeline established for the first test of SPAM on human samples, aligning R1 with R2 and filtering by quality and complementarity for the L1 sequence and the human genome (Table 17).

Table 17: Step-by-step analysis of the human AD and CTRL SPAM samples. FC, frontal cortex, K, kidney.

	Sample	Tissue	Reads	Fragments	Fragments aligned over L1 Ref seq	MapFragments aligned over Hg19
AD	5248	FC	R1: 1565238 R2: 1565238	635730 (100 %)	548281(86.24 %)	36122 (5.68 %)
		K	R1: 1347169 R2: 1347169	631604 (100%)	545266 (86 %)	91025 (14.41%)
	7466	FC	R1: 1017622 R2: 1017622	393910 (100%)	329672 (83 %)	16363 (4.15 %)
		K	R1: 2854513 R2: 2854513	1024634 (100%)	778558(76 %)	26034 (2.54%)
	9345	FC	R1: 2251827 R2: 2251827	738608(100%)	549465 (74 %)	10702(1.45%)
		K	R1: 1890426 R2: 1890426	729034 (100%)	608133 (83 %)	41415(5.68 %)
CTRL	6868	FC	R1: 3361839 R2: 3361839	1550023 (100%)	1270944 (82 %)	76232 (4.92 %)
		K	R1: 4001288 R2: 4001288	1729861 (100 %)	1374564 (79 %)	85778 (4.96 %)
	9269	FC	R1: 1843873 R2: 1843873	739446 (100 %)	624957 (84 %)	32876 (4.45 %)
		K	R1: 1845406 R2: 1845406	701756 (100%)	560397(80%)	30812 (4.39 %)
	929	FC	R1: 3051473 R2: 3051473	1307753 (100 %)	1063881 (81 %)	59025 (4.51 %)
		K	R1: 2927667 R2: 2927667	1183817 (100%)	886506 (75 %)	32242 (2.72 %)

In this case we obtained a smaller amount of aligned MapFragments (compared to the first samples analyzed) probably because during the preparation of these samples they were not gel-purified as the previous one, and this caused the presence, during the sequencing, of short sequences that were preferentially sequenced, but were not suitable for the following analysis. Moreover, we obtained on average shorter Fragments than

those obtained in the first test, and this forced us to discard a lot of Fragments containing a genomic portion not long enough to be mapped in the reference genome.
As shown in Figure 74, for each sample analysed, the MapFragments fell at 211 bp from the beginning of the L1's 5'UTR region, according to the designing of the primers.

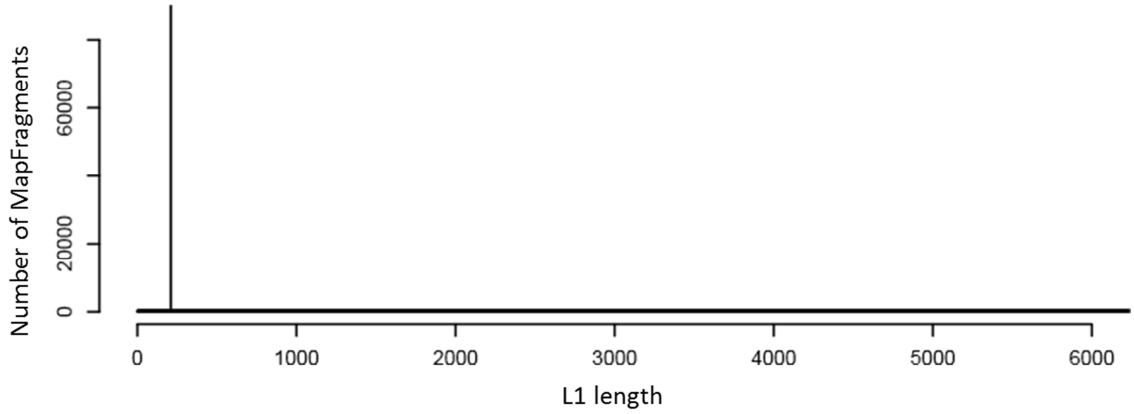


Figure 74: Example of MapFragments' mapping on the L1 sequence. The MapFragments for each sample mapped at 211bp from the beginning of L1's 5'UTR as expected, according to the designing of primers used in the SPAM protocol.

Based on the distance (> or <1000 bp) from an L1 annotated in our database, we defined AIS and NIS, determined by at least 1 MapFragments. 1 MapFragment was chosen, instead of 3 (as for the previous experiment), since much less reads and MapFragments were obtained for this preliminary sequencing. When further sequencing runs will be performed, more stringent parameters will be used.
The total amount of AIS and NIS are reported in Table 18, and values were also normalized on the initial number of MapFragments obtained for each sample. We decided to do this normalization taking into account the unbalanced number of MapFragments that we obtained for each sample.

Table 18: Total number of AIS and NIS for each sample, defined by at least 1 MapFragment. The two values are reported also normalized by the initial total amount of MapFragments. MF, MapFragment. The ratio of NIS and AIS is reported.

	Sample	Tissue	AIS Defined by at least 1 MapFragment	NIS Defined by at least 1 MapFragment	Normalized AIS (AIS/MF)*10000	Normalized NIS (NIS/MF)*10000	NIS/AIS
AD	5248	FC	2659	1663	736.12	460.38	0.6
		K	21893	23969	2405.16	2633.23	1.1
	7466	FC	2332	1447	1425.17	884.31	0.6
		K	2763	2284	1061.30	877.31	0.8
	9345	FC	1275	261	1191.37	243.88	0.2
		K	9381	10025	2265.12	2420.62	1.1
CTRL	6868	FC	3537	2396	463.98	314.30	0.7
		K	3763	2649	438.69	308.82	0.7
	9269	FC	2185	1085	664.62	330.03	0.5
		K	9012	9700	2924.83	3148.12	1.1
	929	FC	3814	2824	646.17	478.44	0.7
		K	5448	5144	1689.72	1595.43	0.9

The ratio between NIS and AIS can be considered a measure of somatic mobilization of full length L1s. Lower the ratio, lower the number of NIS respect to the annotated ones; higher the ratio, higher the number of NIS respect to the AIS detected.
By plotting the average values of these ratios for AD and control tissues (Figure 75), we observed that the ratio is higher in kidney than in frontal cortex, suggesting a high rate of retrotransposition in the kidney, and that in AD frontal cortexes there is a lower ratio compared to the controls, in accordance with the qPCR data.

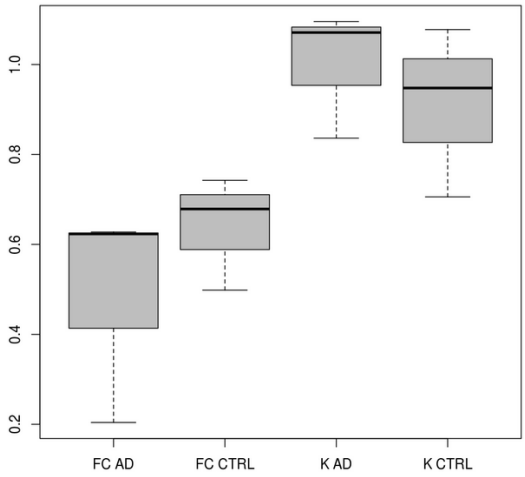


Figure 75: Average values of NIS/AIS ratios for AD and control tissues (K, kidney; FC, frontal cortex). These ratios allow us to avoid the bias of sequencing depth and differences in chromatin accessibility. Boxplots report the maximum and minimum values, the median, 25% and 75% percentiles.

We checked for common and exclusive AIS and NIS between frontal cortex and kidney of each samples, and, as seen for the test samples, common IS were more abundant for AIS than for NIS (Figure 76).

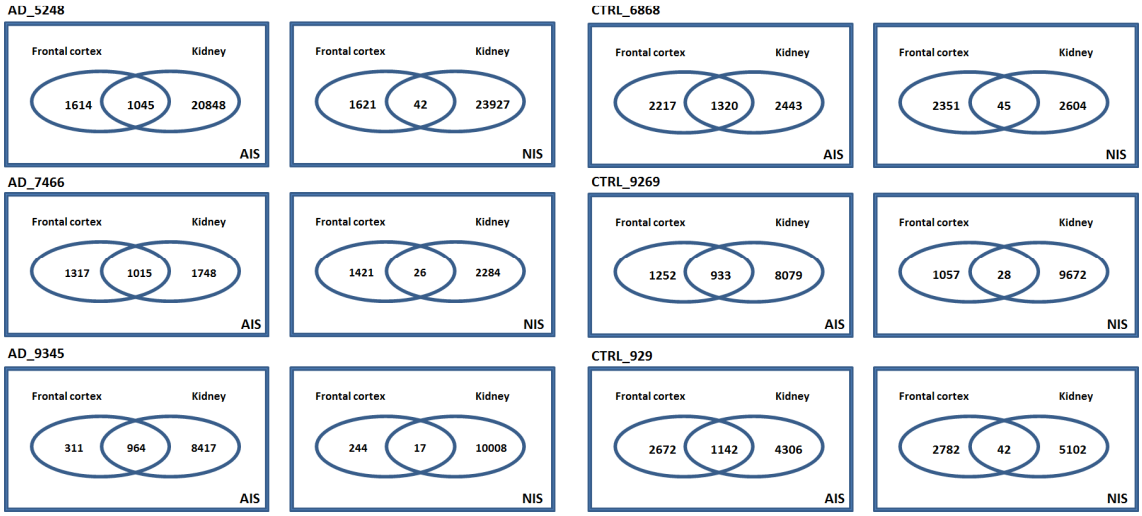


Figure 76: Venn diagram of common and exclusive AIS and NIS between frontal cortex and kidney of each sample.

We then calculated a value for each IS that we called MPM (MapFragments per million), to normalize the number of MapFragments defining each IS with the total number of MapFragments detected in the corresponding sample.

We chose a common cutoff of MPM for all the NIS, and we performed a Gene Ontology analysis on NIS associated genes considering samples above the cutoff (cutoff = 25 MPM).

In, Table 19 and Table 20 are reported the gene ontology results with the lowest p-value scores obtained for each sample.

Table 19: Gene ontology analysis of NIS associated genes for AD patients with a cutoff of MPM above 25. Background frequency corresponds to the total number of genes associated to the term in the human genome; sample frequency corresponds to the AIS associated genes found for the term. The first 10 results with the smallest p-value are shown. p<0.05

	Biological function	Background frequency	Sample frequency	P-value
AD_5248_frontal cortex	biological_process (GO:0008150)	15920	956	2.23E-24
	single-organism process (GO:0044699)	12164	787	4.67E-23
	cellular process (GO:0009987)	13436	839	4.86E-21
	single-organism cellular process (GO:0044763)	10784	705	7.62E-19
	nervous system development (GO:0007399)	1922	192	2.43E-16
	anatomical structure development (GO:0048856)	4236	337	1.48E-15
	multicellular organismal development (GO:0007275)	4225	336	1.82E-15
	single-organism developmental process (GO:0044767)	4773	368	2.37E-15
	developmental process (GO:0032502)	4825	371	2.42E-15
	system development (GO:0048731)	3692	303	2.87E-15
AD_5248 kidney	positive regulation of telomere maintenance (GO:0032206)	7	2	1.35E-03
	regulation of telomere maintenance (GO:0032204)	16	2	7.03E-03
	nucleotide-binding oligomerization domain containing signalin	28	2	2.13E-02
	cytoplasmic pattern recognition receptor signaling pathway (G	34	2	3.13E-02
	epithelial cell development (GO:0002064)	175	3	4.58E-02
AD_7466_frontal cortex	biological_process (GO:0008150)	15920	840	7.47E-22
	biological regulation (GO:0065007)	10349	614	2.36E-20
	regulation of cellular process (GO:0050794)	9486	574	7.68E-20
	regulation of biological process (GO:0050789)	9875	591	9.50E-20
	single-organism process (GO:0044699)	12164	685	1.38E-18
	cellular process (GO:0009987)	13436	730	1.21E-16
	single-organism cellular process (GO:0044763)	10784	614	2.66E-15
	multicellular organismal process (GO:0032501)	5984	388	2.19E-14
	single-multicellular organism process (GO:0044707)	5739	375	3.55E-14
	system development (GO:0048731)	3692	266	2.57E-13
AD_7466_kidney	single-organism process (GO:0044699)	12164	1076	2.38E-33
	biological_process (GO:0008150)	15920	1296	6.54E-32
	cellular process (GO:0009987)	13436	1146	2.12E-30
	single-organism developmental process (GO:0044767)	4773	528	1.19E-28
	multicellular organismal development (GO:0007275)	4225	483	1.25E-28
	single-organism cellular process (GO:0044763)	10784	967	4.32E-28
	developmental process (GO:0032502)	4825	530	5.10E-28
	system development (GO:0048731)	3692	430	3.61E-26
	anatomical structure development (GO:0048856)	4236	475	4.80E-26
	single-multicellular organism process (GO:0044707)	5739	595	5.35E-26
AD_9345_frontal cortex	biological_process (GO:0008150)	15920	164	3.55E-07
	multicellular organismal process (GO:0032501)	5984	82	4.58E-05
	regulation of multicellular organismal process (GO:0051239)	2121	41	5.93E-05
	single-organism process (GO:0044699)	12164	133	1.13E-04
	positive regulation of biological process (GO:0048518)	4686	68	1.40E-04
	single-multicellular organism process (GO:0044707)	5739	78	1.74E-04
	positive regulation of cellular process (GO:0048522)	4232	60	2.10E-03
	positive regulation of macromolecule metabolic process (GO:0	2307	39	3.24E-03
	developmental process (GO:0032502)	4825	65	3.97E-03
	cell adhesion (GO:0007155)	823	20	5.21E-03
AD_9345_kidney	cellular process (GO:0009987)	13436	232	5.19E-06
	biological_process (GO:0008150)	15920	261	5.80E-06
	single-organism process (GO:0044699)	12164	208	8.41E-04
	multicellular organismal process (GO:0032501)	5984	120	9.89E-04
	biological regulation (GO:0065007)	10349	183	1.20E-03
	single-multicellular organism process (GO:0044707)	5739	114	3.45E-03
	regulation of cellular process (GO:0050794)	9486	169	3.51E-03
	regulation of metabolic process (GO:0019222)	6113	118	8.68E-03
	generation of neurons (GO:0048699)	1237	36	9.56E-03
	regulation of biological process (GO:0050789)	9875	172	1.20E-02

Table 20: Gene ontology analysis of NIS associated genes for control patients with a cutoff of MPM above 25. Background frequency corresponds to the total number of genes associated to the term in the human genome; sample frequency corresponds to the AIS associated genes found for the term. The first 10 results with the smallest p-value are shown. p<0.05.

	Biological function	Background frequency	Sample frequency	P-value
CTRL_6868_frontal cortex	cellular component organization or biogenesis (GO:0071840)	4569	59	5.05E-05
	nervous system development (GO:0007399)	1922	34	5.59E-05
	anatomical structure development (GO:0048856)	4236	56	5.69E-05
	system development (GO:0048731)	3692	51	6.79E-05
	cellular component organization (GO:0016043)	4449	57	1.22E-04
	single-organism developmental process (GO:0044767)	4773	59	2.33E-04
	single-organism process (GO:0044699)	12164	112	2.86E-04
	developmental process (GO:0032502)	4825	59	3.38E-04
	multicellular organismal development (GO:0007275)	4225	53	7.47E-04
	cell adhesion (GO:0007155)	823	19	9.55E-04
CTRL_6868_kidney	regulation of focal adhesion assembly (GO:0051893)	37	3	1.50E-03
	regulation of cell-substrate junction assembly (GO:0090109)	37	3	1.50E-03
	regulation of adherens junction organization (GO:1903391)	38	3	1.62E-03
	regulation of cell junction assembly (GO:1901888)	50	3	3.64E-03
	regulation of cell-matrix adhesion (GO:0001952)	78	3	1.34E-02
	cardiac right ventricle morphogenesis (GO:0003215)	16	2	1.41E-02
	positive regulation of cell junction assembly (GO:1901890)	18	2	1.79E-02
	positive regulation of focal adhesion assembly (GO:0051894)	18	2	1.79E-02
	positive regulation of adherens junction organization (GO:190	19	2	1.99E-02
	regeneration (GO:0031099)	107	3	3.33E-02
CTRL_9269_frontal cortex	biological_process (GO:0008150)	15920	642	2.90E-14
	nervous system development (GO:0007399)	1922	139	5.99E-14
	neurogenesis (GO:0022008)	1312	103	8.92E-12
	generation of neurons (GO:0048699)	1237	99	9.99E-12
	system development (GO:0048731)	3692	211	1.26E-11
	cellular process (GO:0009987)	13436	558	4.23E-11
	neuron differentiation (GO:0030182)	914	80	5.47E-11
	anatomical structure development (GO:0048856)	4236	231	6.10E-11
	multicellular organismal development (GO:0007275)	4225	228	2.95E-10
	cell development (GO:0048468)	1458	106	4.17E-10
CTRL_9269_kidney	biological_process (GO:0008150)	15920	3949	1.74E-92
	single-organism process (GO:0044699)	12164	3192	1.59E-75
	cellular process (GO:0009987)	13436	3409	2.08E-67
	single-organism cellular process (GO:0044763)	10784	2862	1.10E-63
	biological regulation (GO:0065007)	10349	2745	6.58E-58
	regulation of biological process (GO:0050789)	9875	2636	5.48E-56
	regulation of cellular process (GO:0050794)	9486	2546	1.71E-54
	single-organism developmental process (GO:0044767)	4773	1478	3.82E-54
	developmental process (GO:0032502)	4825	1486	4.04E-53
	multicellular organismal development (GO:0007275)	4225	1336	3.82E-52
CTRL_929_frontal cortex	negative regulation of cellular process (GO:0048523)	3608	63	2.42E-05
	negative regulation of biological process (GO:0048519)	3932	65	1.06E-04
	regulation of cellular process (GO:0050794)	9486	121	1.22E-04
	tube development (GO:0035295)	650	21	1.62E-04
	single-organism cellular process (GO:0044763)	10784	131	3.94E-04
	cell adhesion (GO:0007155)	823	23	5.36E-04
	biological adhesion (GO:0022610)	826	23	5.69E-04
	regulation of biological process (GO:0050789)	9875	122	7.29E-04
	single-organism process (GO:0044699)	12164	142	7.39E-04
	biological regulation (GO:0065007)	10349	126	8.71E-04
CTRL_929_kidney	biological_process (GO:0008150)	15920	2481	1.97E-63
	single-organism process (GO:0044699)	12164	2030	5.61E-57
	cellular process (GO:0009987)	13436	2155	5.97E-49
	single-organism cellular process (GO:0044763)	10784	1810	5.59E-45
	multicellular organismal development (GO:0007275)	4225	856	1.84E-37
	single-organism developmental process (GO:0044767)	4773	936	1.49E-36
	anatomical structure development (GO:0048856)	4236	854	1.51E-36
	developmental process (GO:0032502)	4825	943	2.38E-36
	system development (GO:0048731)	3692	764	3.05E-35
	multicellular organismal process (GO:0032501)	5984	1099	6.61E-33

We also performed a further GO analysis on NIS associated genes that were common between samples of the same group (AD or controls). In Table 21 are reported the gene ontology results with the lowest p-value scores obtained for each group of sample (complete tables are reported in Appendix A).

Table 21: Gene ontology of NIS associated genes, found common between AD samples (frontal cortex and kidney) and healthy controls. Background frequency corresponds to the total number of genes associated to the term in the human genome; sample frequency corresponds to the AIS associated genes found for the term. The first 10 results with the smallest p-value are shown. p<0.05.

	Biological function	Background frequency	Sample frequency	P-value
AD common NIS genes	regulation of focal adhesion assembly (GO:0051893)	37	2	7.21E-04
	regulation of cell-substrate junction assembly (GO:0090109)	37	2	7.21E-04
	regulation of adherens junction organization (GO:1903391)	38	2	7.60E-04
	regulation of cell junction assembly (GO:1901888)	50	2	1.31E-03
	regulation of cell-matrix adhesion (GO:0001952)	78	2	3.18E-03
	negative regulation of ripoptosome assembly involved in necroptotic process (GO:1902443)	1	1	3.30E-03
	regulation of ripoptosome assembly involved in necroptotic process (GO:1902442)	1	1	3.30E-03
	regulation of cellular component biogenesis (GO:0044087)	520	3	6.25E-03
	fasciculation of motor neuron axon (GO:0097156)	2	1	6.60E-03
	fasciculation of sensory neuron axon (GO:0097155)	3	1	9.90E-03
CTRL common NIS genes	cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742)	185	8	1.87E-06
	cell-cell adhesion (GO:0098609)	186	8	1.95E-06
	homophilic cell adhesion via plasma membrane adhesion molecules (GO:0007156)	137	7	4.74E-06
	cell adhesion (GO:0007155)	823	12	7.38E-05
	biological adhesion (GO:0022610)	826	12	7.66E 05
	neuron differentiation (GO:0030182)	914	12	2.17E-04
	central nervous system neuron differentiation (GO:0021953)	170	6	3.54E-04
	generation of neurons (GO:0048699)	1237	13	8.85E-04
	neuron development (GO:0048666)	743	10	1.28E-03
	nervous system development (GO:0007399)	1922	16	1.38E-03

The next step was to check whether the AIS and NIS that we discovered were inserted in genomic loci involved in the pathogenesis of early-onset familial AD (EO-FAD) or late-onset AD (LOAD). We used as reference the list of functional classes of AD genes reported in the paper published by Karch and colleagues (Karch et al., 2014), and, on a total of 42 AD-associated genes, we found on average a greater number of both AIS and NIS in the kidney of AD patients compared to the healthy controls, and a lower amount of AIS and NIS in the frontal cortex of AD patients compared to healthy controls (Table 22).

Table 22: Total number of AIS and NIS discovered in AD-associated genes in frontal cortex and kidney of both AD patients and healthy controls.

	Sample	Tissue	AIS Defined by at least 1 MapFragment	NIS Defined by at least 1 MapFragment	Normalized AIS (AIS/MF)*10000	Normalized NIS (NIS/MF)*10000	NIS/AIS
AD	5248	FC	6	13	1.66	3.60	2.2
		K	51	62	5.60	6.81	1.2
	7466	FC	4	4	2.44	2.44	1.0
		K	4	11	1.54	4.23	2.8
	9345	FC	1	1	0.93	0.93	1.0
		K	16	21	3.86	5.07	1.3
CTRL	6868	FC	6	8	0.79	1.05	1.3
		K	5	8	0.58	0.93	1.6
	9269	FC	1	4	0.30	1.22	4.0
		K	18	24	5.84	7.79	1.3
	929	FC	9	9	1.52	1.52	1.0
		K	7	13	2.17	4.03	1.9

We decided to focus our attention on one of the most important genes linked to AD: the amyloid beta precursor protein gene APP, and we observed the presence of 14 total L1 insertions (AIS and NIS) for the AD patients (frontal cortex and kidney) (Figure 77), and 8 insertions for the healthy controls (Figure 78).

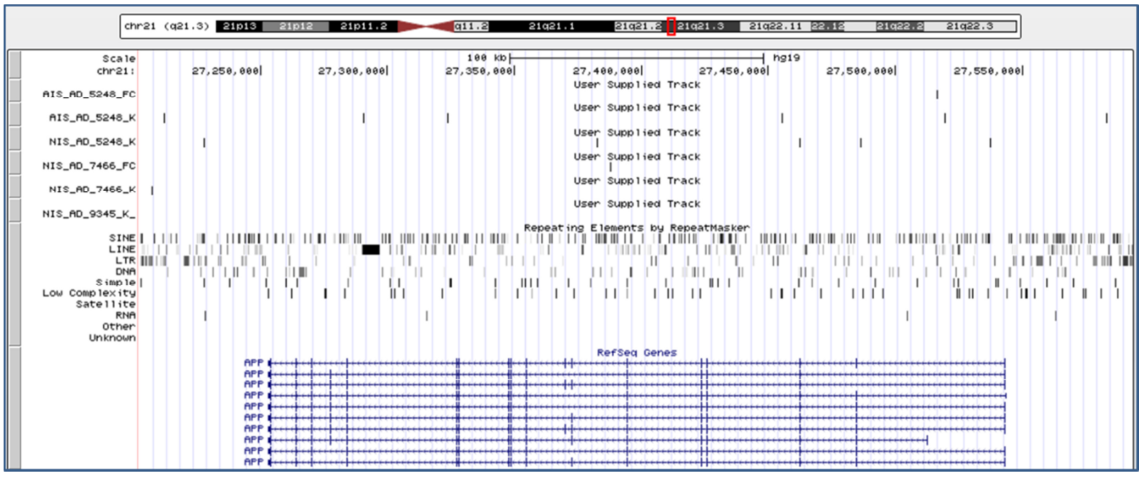


Figure 77: Genome Browser view of AIS and NIS falling in the APP gene ± 50 kb. Data are reported for both the frontal cortex and kidney of AD patients.

The insertions detected for the control patients were all intronic, on an annotated L1 full length or outside the gene (we considered a genomic area from 50 kb upstream and downstream the APP gene). The insertions detected for the AD patients were intronic, outside the gene, in the annotated full length L1 and one was also partially overlapping with an intron and an exon.

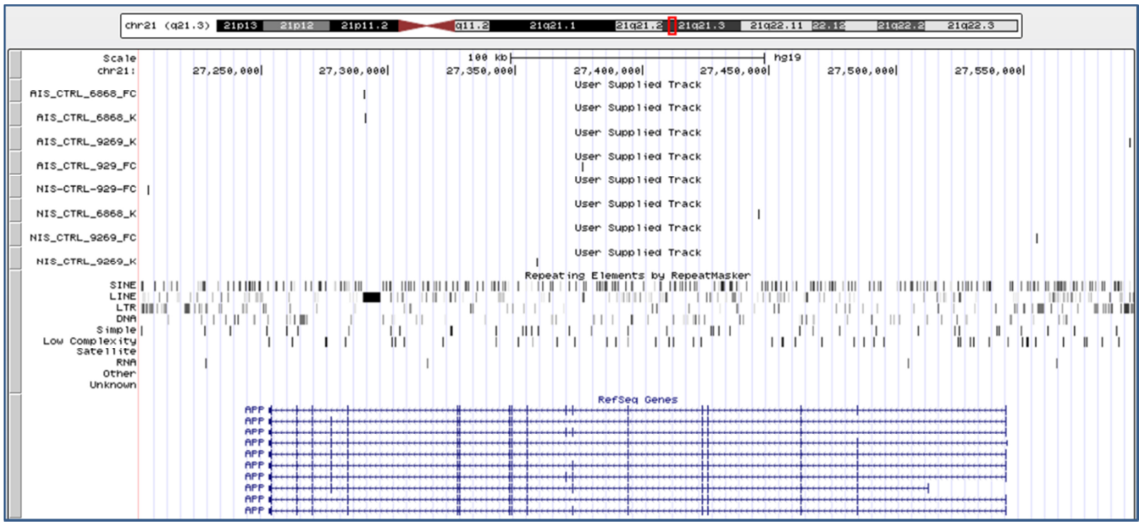


Figure 78: Genome Browser view of AIS and NIS falling in the APP gene ± 50 kb. Data are reported for both the frontal cortex and kidney of healthy controls.

Scaling the SPAM protocol to the *Drosophila melanogaster*'s *roo* element

To prove the scalability of this new SPAM technique, we tried to apply it also to a different organism and a different repetitive element. In particular we adapted it for identifying the insertion sites of the *roo* element, the most abundant family of LTR retrotransposons in the *Drosophila* genome.

In order to understand if a less amount of *Drosophila* genomic DNA (<2µg, the quantity used in the protocol for human samples) could be used, we applied the SPAM technique on the same DNA sample (extracted from a pool of adult *Drosophila* carcasses), but starting from a different quantity: 2µg and 500ng.

The protocol that we used was the same of human samples, except for the primers specific for *roo*-LTR, used in the two final steps of PCR. The two PCR products were finally purified and run with an Illumina MiSeq 2x300bp sequencing in 2-plex.

After the bioinformatics quality filtering (Phred score for each nucleotide), only the ~45% of the reads were selected. We aligned the corresponding R1 and R2 to form a fragment, discarding those pairs that didn't align with at least 30bp and 80% identity. We checked for the presence in the fragment of the forward primer (specific for the splinker) and we aligned it to the *roo*-LTR reference sequence. Once discarded the fragments that did not contain the repeat (~5% per sample), we trimmed the portions of the fragments that overlapped with the forward primer and the *roo*-LTR sequence, and retained only the sequences longer than 30 nucleotides. These sequences were stringently aligned over the dm3 *Drosophila melanogaster* reference genome in order to map them, and defined as MapFragments only if they gave a unique alignment result with an identity >50%. The major loss of fragments (~82 %) was due to the fact that they did not contain a genomic part long enough to allow the mapping (Table 23).

Table 23: Results of the bioinformatic pipeline on Drosophila samples. Reads were filtered by quality, overlapping on each other, alignment over the roo-LTR reference sequence and on the dm3 Drosophila melanogaster reference genome.

DROSOPHILA samples	Reads	Fragments	Fragments aligned over roo-LTR reference sequence	MapF aligned over dm3 Drosophila melanogaster reference genome
2 µg sample	R1: 9167269 R2: 9167269	4008244 (100%)	3598549 (90%)	153579 (3.83%)
500 ng sample	R1: 7546761 R2: 7546761	3241271 (100%)	2871372 (89%)	105558 (3.26%)

The next step was to identify AIS and NIS, defined by one or more than one MapFragment (Table 24, Table 25, Table 26, Table 27 and Table 28). We decided to increase the stringency of this parameter because with 1 MapFragment we obtained an unexpectedly high number of NIS, and we observed a progressive decrease in the total number of both AIS and NIS.

This allowed us to verify that there was a substantial difference in the number of AIS and NIS detected in the two samples if defined by at least 1 MapFragment, while with more stringent cutoffs these differences between the two samples became flat.

In the *Drosophila* genome there are 541 annotated *roo*-LTR insertions, and even with AIS defined by at least 1 MapFragment we didn't detect them all, probably because the system was saturated by the high number of NIS that we found.

Table 24: Total number of AIS and NIS in the two samples, defined by at least 1 MapFragment. The two values are reported also normalized by the initial total amount of MapFragments. MF, MapFragment.

DROSOPHILA samples	AIS Defined by at least 1 MapFragment	NIS Defined by at least 1 MapFragment	Normalized AIS (AIS/MF)*10000	Normalized NIS (NIS/MF)*10000	NIS/AIS
2 µg sample	227	4530	14.8	295.0	20.0
500 ng sample	180	2275	17.1	215.5	12.7

Table 25: Total number of AIS and NIS in the two samples, defined by at least 2 MapFragments. The two values are reported also normalized by the initial total amount of MapFragments. MF, MapFragment.

DROSOPHILA samples	AIS Defined by at least 2 MapFragment	NIS Defined by at least 2 MapFragment	Normalized AIS (AIS/MF)*10000	Normalized NIS (NIS/MF)*10000	NIS/AIS
2 µg sample	129	1270	8.4	82.7	9.8
500 ng sample	114	1124	10.8	106.5	9.9

Table 26: Total number of AIS and NIS in the two samples, defined by at least 3 MapFragments. The two values are reported also normalized by the initial total amount of MapFragments. MF, MapFragment.

DROSOPHILA samples	AIS Defined by at least 3 MapFragment	NIS Defined by at least 3 MapFragment	Normalized AIS (AIS/MF)*10000	Normalized NIS (NIS/MF)*10000	NIS/AIS
2 µg sample	97	477	6.3	31.1	4.9
500 ng sample	80	608	7.6	57.6	7.6

Table 27: Total number of AIS and NIS in the two samples, defined by at least 5 MapFragments. The two values are reported also normalized by the initial total amount of MapFragments. MF, MapFragment.

DROSOPHILA samples	AIS Defined by at least 5 MapFragment	NIS Defined by at least 5 MapFragment	Normalized AIS (AIS/MF)*10000	Normalized NIS (NIS/MF)*10000	NIS/AIS
2 µg sample	67	221	4.4	14.4	3.3
500 ng sample	60	262	5.7	24.8	4.4

Table 28: Total number of AIS and NIS in the two samples, defined by at least 10 MapFragments. The two values are reported also normalized by the initial total amount of MapFragments. MF, MapFragment.

DROSOPHILA samples	AIS Defined by at least 10 MapFragment	NIS Defined by at least 10 MapFragment	Normalized AIS (AIS/MF)*10000	Normalized NIS (NIS/MF)*10000	NIS/AIS
2 µg sample	46	179	3.0	11.7	3.9
500 ng sample	42	167	4.0	15.8	4.0

We also checked for common and exclusive AIS and NIS obtained for the two samples, considering all the 5 cutoffs: with a cutoff of 1 MapFragment almost half of the AIS detected for the two samples was in common, while a very small percentage of NIS was found to be present in both the samples (Figure 79, Figure 80, Figure 81, Figure 82 and Figure 83). With more stringent cutoffs we observed the percentage of common IS becoming higher and reach the ~90% for those AIS and NIS defined by at least 10 MapFragments.

Figure 79: Venn diagram of common and exclusive AIS and NIS defined by at least 1 MapFragment for the 2µg and the 500ng samples.

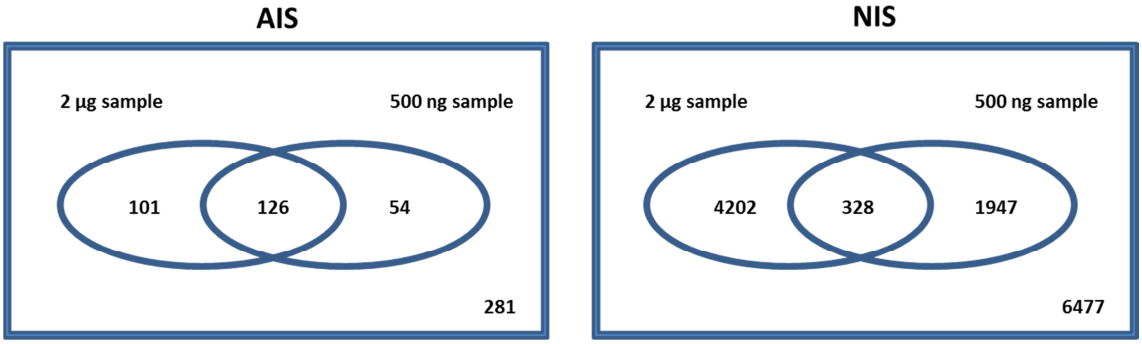


Figure 80: Venn diagram of common and exclusive AIS and NIS defined by at least 2 MapFragments for the 2µg and the 500ng samples.

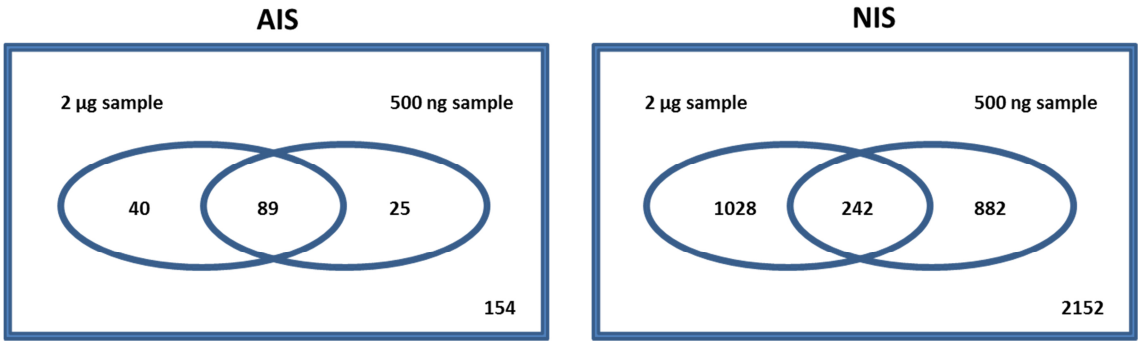


Figure 81: Venn diagram of common and exclusive AIS and NIS defined by at least 3 MapFragments for the 2µg and the 500ng samples.

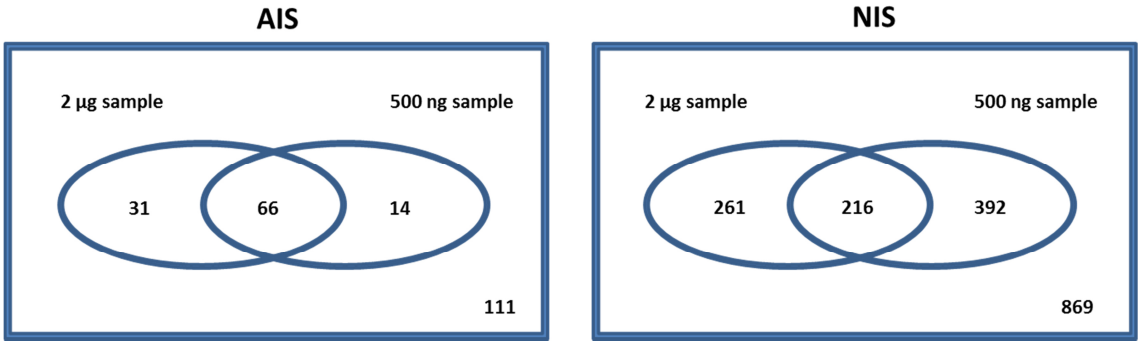


Figure 82: Venn diagram of common and exclusive AIS and NIS defined by at least 5 MapFragments for the 2µg and the 500ng samples.

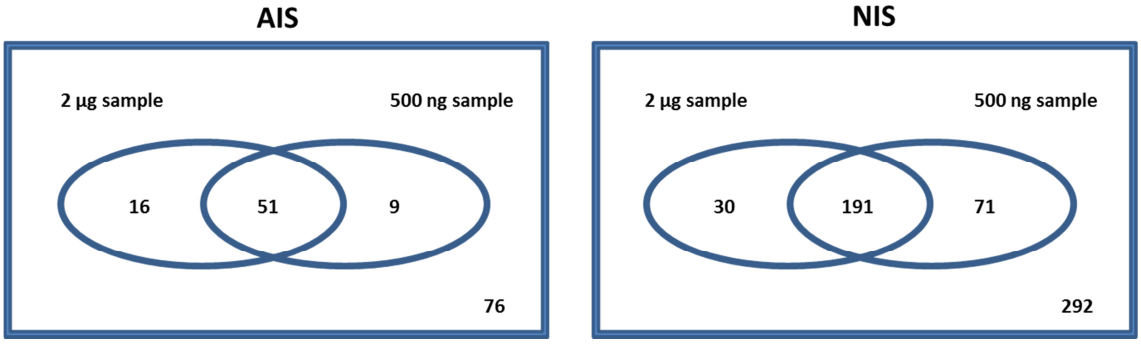
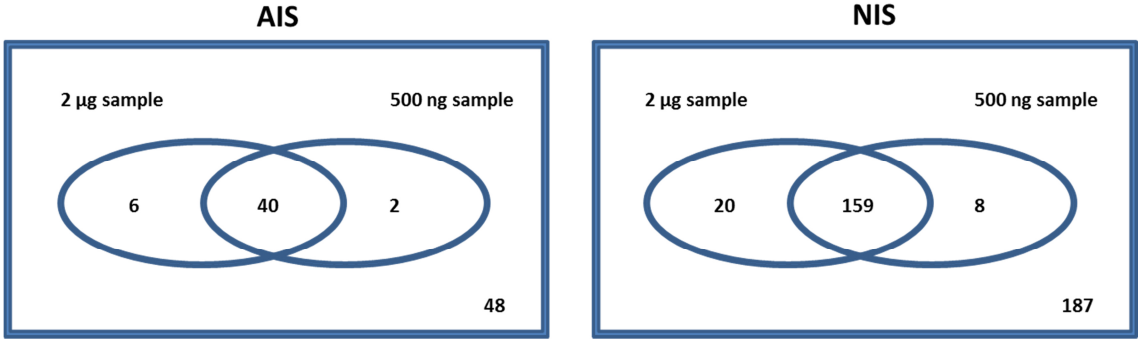


Figure 83: Venn diagram of common and exclusive AIS and NIS defined by at least 10 MapFragments for the 2µg and the 500ng samples.



Combining together overlapping or "book-ended" MapFragments, 4757 total IS (defined by 1 or more MapFragments) were obtained for the 2 µg sample and 2455 for the 500 ng sample. The 70% of all these IS fell inside a gene. The MapFragments of the 2 µg sample covered 2307 Drosophila genes while the MapFragments of the 500 ng sample covered 1315 genes.

We performed a gene ontology (GO) analysis of the NIS associated genes (found in the 2µg sample), and we did it for NIS defined by 1,2,3,5 and 10 MapFragments (the complete GO analysis are reported in the chapter "Supplementary"). The number of biological functions (terms) that we found, varied with the number of MapFragments defining the NIS: for NIS defined by at least 1 MapFragments we found 406 terms, and 30 for NIS defined by at least 10 MapFragments (Table 29).

Table 29: Gene ontology of NIS associated genes (defined by at least 1 and 10 MapFragments). Background frequency corresponds to the total number of genes associated to the term in Drosophila; sample frequency corresponds to the NIS associated genes found for the term. The first 10 results with the smallest p-value are shown. p<0.05

Biological function (NIS defined by at least 1 MF)	Background frequency	Sample frequency	P-value
anatomical structure morphogenesis (GO:0009653)	1634	503	3.08E-45
biological regulation (GO:0065007)	3307	829	6.00E-45
regulation of cellular process (GO:0050794)	2829	733	7.58E-43
regulation of biological process (GO:0050789)	3047	772	2.48E-42
organ development (GO:0048513)	1220	400	1.77E-40
system development (GO:0048731)	2172	597	4.54E-40
single-multicellular organism process (GO:0044707)	3217	792	3.68E-39
organ morphogenesis (GO:0009887)	792	292	7.90E-37
anatomical structure development (GO:0048856)	2870	712	8.23E-35
single-organism developmental process (GO:0044767)	3155	764	1.06E-34

Biological function (NIS defined by at least 10 MF)	Background frequency	Sample frequency	P-value
response to stimulus (GO:0050896)	2415	45	1.31E-05
single-multicellular organism process (GO:0044707)	3217	51	2.11E-04
cell communication (GO:0007154)	1384	28	1.39E-03
regulation of biological process (GO:0050789)	3047	47	1.47E-03
system process (GO:0003008)	548	16	1.81E-03
single-organism process (GO:0044699)	6262	77	2.19E-03
regulation of cellular process (GO:0050794)	2829	44	2.62E-03
axonogenesis (GO:0007409)	284	11	3.19E-03
axon development (GO:0061564)	291	11	3.96E-03
regulation of response to DNA damage stimulus (GO:2001020)	24	4	4.81E-03

We performed also a GO analysis of the AIS associated genes that we found in the 2 µg sample, doing it for AIS defined by 1,2,3,5 and 10 MapFragments (complete GO analysis are reported in the chapter "Supplementary") (Table 30).

Table 30: Gene ontology of AIS associated genes (defined by at least 1 and 10 MapFragments). Background frequency corresponds to the total number of genes associated to the term in Drosophila; sample frequency corresponds to the AIS associated genes found for the term. The first 10 results with the smallest p-value are shown. p<0.05

Biological function (AIS defined by at least 1 MF)	Background frequency	Sample frequency	P-value
regulation of response to stimulus (GO:0048583)	855	16	1.25E-03
neuron projection extension (GO:1990138)	29	4	1.37E-03
axon extension (GO:0048675)	29	4	1.37E-03
regulation of signaling (GO:0023051)	701	13	8.86E-03
regulation of cell communication (GO:0010646)	710	13	1.00E-02
developmental cell growth (GO:0048588)	50	4	1.10E-02
regulation of signal transduction (GO:0009966)	629	12	1.25E-02
anatomical structure morphogenesis (GO:0009653)	1634	21	1.38E-02
regulation of cellular process (GO:0050794)	2829	30	1.40E-02
synapse assembly (GO:0007416)	57	4	1.81E-02

Biological function (AIS defined by at least 10 MF)	Background frequency	Sample frequency	P-value
sperm individualization (GO:0007291)	47	3	3.61E-04
spermatid development (GO:0007286)	99	3	3.20E-03
cellularization (GO:0007349)	99	3	3.20E-03
positive regulation of JAK-STAT cascade (GO:0046427)	24	2	4.20E-03
spermatid differentiation (GO:0048515)	111	3	4.45E-03
histone H3-K4 dimethylation (GO:0044648)	1	1	9.68E-03
regulation of JAK-STAT cascade (GO:0046425)	44	2	1.39E-02
peptidyl-lysine dimethylation (GO:0018027)	2	1	1.94E-02
histone H3-K4 trimethylation (GO:0080182)	2	1	1.94E-02
spermatogenesis (GO:0007283)	204	3	2.51E-02

We finally performed a GO analysis of the genes associated to the *roo*-LTR annotated in our database that we didn't detect with our technique (Table 31). Some of these genes

were found to be involved in biological functions important during development. Since the DNA used in this experiment came from a pool of adult *Drosophila* carcasses, it could be that *roo*-LTR inserted in genomic loci silent during adulthood were inaccessible to the technique and therefore not detectable.

Table 31: Gene ontology of genes associated to the *roo*-LTR annotated in our database that we didn't detect. Background frequency corresponds to the total number of genes associated to the term in Drosophila; sample frequency corresponds to the AIS associated genes found for the term. The first 10 results with the smallest p-value are shown. p<0.05

Biological function	Background frequency	Sample frequency	P-value
heart contraction (GO:0060047)	6	2	4.72E-03
blood circulation (GO:0008015)	6	2	4.72E-03
single-organism process (GO:0044699)	6262	35	5.75E-03
anatomical structure development (GO:0048856)	2870	21	1.01E-02
leg disc development (GO:0035218)	101	4	1.03E-02
single-multicellular organism process (GO:0044707)	3217	22	1.80E-02
regionalization (GO:0003002)	462	7	2.66E-02
cell morphogenesis involved in differentiation (GO:0000904)	464	7	2.73E-02
axon target recognition (GO:0007412)	15	2	2.89E-02
single-organism cellular process (GO:0044763)	4829	28	3.29E-02

This page intentionally left blank

Discussion

Since their discovery in maize by Barbara McClintock in the 40s, TEs have been identified in every eukaryotic genome analysed (Akagi et al., 2013).

TEs are a class of repetitive DNA sequences able to mobilize and change location in the genome, and were demonstrated to make up almost 50% of the human genome and slightly less in the mouse (Lander et al., 2001; Waterston et al., 2002).

According to their structure and mechanism of mobilization, different types of TEs have been identified, but the autonomous non-LTR retrotransposon Long Interspersed Nuclear Element-1 (L1) is for sure the most impactful and still active TE in the human and mouse genome (Richardson et al., 2014).

Historically thought to transpose only during gametogenesis and in tumours, L1s were recently demonstrated to be active in the mouse, rat and human neural progenitor cells (NPCs) (Thomas and Muotri, 2012). Moreover, they were shown to be more abundant in the human hippocampus, than in other non-nervous tissues such as liver and heart, challenging the dogma that neurons are genetically stable entities, and introducing the concept of neuronal mosaicism (Coufal et al., 2009; Reilly et al., 2013).

Besides human diseases caused by the direct effect of L1 insertion, some neurological diseases (Rett syndrome and ataxia telangiectasia) were demonstrated to misregulate L1 retrotransposition, which could contribute to some pathological aspects of the diseases (Muotri et al., 2010; Coufal et al., 2011; Thomas et al., 2012), however it is still unknown how or even if L1 retrotransposition can directly cause neurological disorders (Erwin et al., 2014).

Alzheimer's disease (AD), the leading cause of dementia in the elderly, is a progressive neurodegenerative disorder characterized by progressive loss of memory and other cognitive functions. Since his first description in 1907 by the German neuropathologist Alois Alzheimer, great steps forward have been made towards the comprehension of this complex disease, but no effective drug treatments able to revert or prevent AD have been discovered yet (Kosik, 2013).

It has been recently demonstrated that AD patients suffer from vitamin B12 deficiency and high homocystein content in blood, which contribute to the disregulation of S-adenosylmethionine synthesis and DNA methylation (Scarpa et al., 2006), and that the level of methylation markers such as 5-methylcytidine are lower in AD postmortem

hippocampus compared to controls (Chouliaras et al., 2013). Since altered methylation at the genomic level can cause modifications in the expression and mobilization of L1 elements (Muotri et al., 2010), we asked ourselves if there could be any detectable change in L1 content in AD.

We collected human post mortem tissues of AD patients and healthy controls from three cohorts of different nationalities: Italian, Spanish and Brazilian, and we measured the content of L1 sequences by qPCR with Taqman probes, using assays in part adapted from those published by Coufal and colleagues (Coufal et al., 2009), and in part designed in our laboratory. In order to estimate the total amount of human L1 sequences present in the human genome, we employed a Taqman assay which detects the L1 ORF2 content, and an assay specific for the invariant high copy number internal control SATA (both published by Coufal and colleagues).

While, in order to quantify the L1 full length forms, we took advantage of the ORF1 Taqman assay published by Coufal and colleagues, using as invariant control an assay at low copy number designed in our laboratory: the glycerhaldeyde 3-phosphate dehydrogenase or GAPDH. In order to be sure to measure the relative content of the complete L1 sequence, we designed a further Taqman assay at the very beginning of L1 5'UTR, to be used with the GAPDH assay.

With the qPCR analysis performed on the Italian cohort (composed of samples of temporal cortex from AD patients and controls) we couldn't be able to see any significant difference neither in the content of the full length L1s (with both ORF1 and 5'UTR assays), nor in the total content of L1 sequences (with the ORF2 assay). In all the three experiments we noticed a high variability, which was probably linked to the young and variable age inside each group, which possibly prevented us from detecting any small difference. Furthermore, it was not clear at which stages were both the controls and patients. For a meaningful analysis of this group an increased number of individuals will be needed. This will allow a better stratification of patients and a more uniform age.

We then moved to the Spanish cohort: in this case we wanted to measure the content of full length L1 sequences (with the ORF1 and 5'UTR assays), taking into account the staging of AD. In particular we analyzed samples of frontal cortex from patients affected by severe AD (Braak stages V-VI), patients affected by mild AD (Braak stages I-II), and healthy controls. With this analysis we observed a progressive and significant decrease in the content of L1 full length sequences from stage to stage, starting from the

healthy controls group to the group of patients affected by severe AD. Notably, we observed a decrease of ~14% between the control group and the mild AD patients, and a decrease of ~7% between the mild AD patients and the severe AD patients.

Since a decrease in the global level of DNA methylation had been previously demonstrated in AD patients, we actually expected to detect a higher amount of L1 sequences in AD tissues compared to controls, deriving from a higher mobilization of L1s thanks to a lower methylation of L1's promoter. At the same time, a higher mobilization of L1 elements in AD patients could have been toxic to neuronal cells, causing cell death, and therefore the detection of a lower amount of L1 elements compared to the controls.

Since a high inter-individual variability in the content of L1 sequences inside the same group of samples had been observed, and extra-nervous tissues such as liver and heart had been previously demonstrated to contain quite stable amounts of L1s (Coufal et al., 2009), we thought that analyzing an extra-brain tissue would have allowed us to use it as an intra-individual normalizer, in order to better detect differences between AD patients and controls in the brain.

The Brazilian cohort was composed of samples from patients affected by severe AD (Braak stages IV-VI) and 10 controls at Braak stages 0-II. The DNA was extracted from samples of frontal cortex, temporal cortex, hippocampus, cerebellum and an extra-nervous tissue: the kidney. By qPCR with Taqman probes we analyzed both the amount of total L1 sequences and the amount of only the full length L1s.

Also in this case, we couldn't observe any difference in the total amount of L1 sequences between AD patients and controls (with the ORF2 assay), but differences were detectable for the full length: by using the ORF1 assay we observed a significant decrease of full length L1 content in AD patients compared to controls in both the cortical tissues (frontal and temporal cortexes). Surprisingly, no differences were detected in the hippocampus, which is the most affected tissue in AD together with cerebral cortex, while a strong decrease was detected in the cerebellum, a tissue that is not primarily affected in AD. As expected, no significant differences were detected in the kidney.

When we performed the analysis of the full length L1 content with the 5'UTR assay, we observed a significant decrease in AD patients at the level of frontal cortex (as for the ORF1 assay), while in temporal cortex the difference was no longer present. Unfortunately, replicas of the ORF1 assay on this tissue were highly variable and the

statistical significance very weak, so that a further analysis will be needed. No differences were detected in hippocampus, as with the ORF1 assay, while a significant difference was observed for both cerebellum and kidney.

Summarizing the qPCR data obtained on human samples, we demonstrated that both the Spanish and the Brazilian cohort presented a less amount of full length L1 sequences in the frontal cortex of AD patients compared to controls. The same strong decrease of full length L1s was detected also in the cerebellum of AD patients from the Brazilian cohort. This latter result was quite unexpected, since the cerebellum is not primarily affected in AD.

The data obtained for the temporal cortex were statistically unclear in the case of the Brazilian cohort, and not significant for the Italian cohort, leading to the conclusion that probably there are not detectable differences in this tissue, even if it is one of the most damaged cortical structures together with the frontal cortex.

The most unexpected result actually regarded the kidney: this tissue was meant to be our intra-individual control, and instead turned out to contain a less amount of L1 sequences, as if it was affected like the brain.

These data can be accounted for two alternative models.

1. If we assume that in AD patients there is a lower degree of DNA methylation, we can speculate that a higher retrotransposition of L1 elements in certain neurons of frontal cortex and cerebellum caused the death of these cells, eventually leading to the detection of a lower amount of L1 sequences in AD patients. To explain the surprising fact that no differences were detected in the hippocampus, we may speculate that it may had been compensated by a higher neurogenesis, as demonstrated to be present in the hippocampus of AD patients (Martinez-Canabal, 2014). This would have led to the detection of an equal amount of L1 sequences in AD patients and controls.
2. Alternatively, AD patients may present less L1 sequences due to a deficit of retrotransposition during embryonic development. According to this hypothesis, AD patients would have been genetically prone to the development of the disease from birth, and the severity of the disease could have been linked directly to the amount of L1 sequences.

In order to investigate which one of the two models is valid, we resorted to the TgCRND8 mouse model. This mouse expresses the human APP protein with Swedish and Indiana mutations under control of the PrP promoter, it is a very aggressive model

of amyloidosis, exhibiting marked premature mortality, with only approximately 50% of mice reaching 9–10 months of age.

First of all, we designed specific assays for the 5'UTR and the 3'UTR regions of each active mouse L1 family (A, Tf and Gf), and an assay for the ORF2 region of the L1, able to detect the total amount of L1 sequences, without discriminating between the different L1 families. Moreover we designed assays for high (MICSAT) and low (GAPDH) copy number internal controls in order to perform the relative quantification of truncated and full length L1 copies.

To assess if in transgenic mice a perturbation in L1 copy number occurs before birth or during aging together with amyloid deposition, we tested samples of cortex, hippocampus and kidney from P0 mice and mice at two stages of the adulthood (3 and 8 months).

We first measured the content of all L1 sequences present in the tissues of P0 mice using the ORF2-MICSAT assay: in the case of cortex samples we were able to see a statistically significant higher amount of ~7% of L1 sequences in TgCRND8 mice compared to controls, while in hippocampus and kidney we didn't see any difference.

We then decided to deepen the study of L1 sequences in the P0 cortex, analyzing separately the three principal L1 families with the 3'UTR-MICSAT/GAPDH assay, which measures the relative content of L1 elements. In accordance with our first results, we observed a significant increase of L1 sequences for each L1 family in transgenic mice compared to controls. In particular, we detected an increase of ~12% of L1-A elements, ~13% of L1-Tf elements and ~6% of L1-Gf elements.

In order to understand if the higher amount of L1 elements detected in transgenic mice could be due to an increase of retrotransposition during embryogenesis, we measured on these samples also the rate of retrotransposition with the 3'UTR-5'UTR assays for each single family. Surprisingly, besides a strong increase in the rate of retrotransposition for the A family in transgenic cortexes, in the case of Tf family we observed even a decrease in rate of retrotransposition, while in Gf family no differences were detectable.

In order to verify if the variations observed in P0 mice were maintained or rescued during postnatal development, we analysed the L1 content in cortex, hippocampus and kidney mice at 3 months of age.

Surprisingly, by qPCR with the ORF2-MICSAT assay we couldn't detect any difference in the total L1 content in cortex and kidney, but we observed a slight but significant increase (~7%) in the hippocampus of transgenic mice. We then decided to

analyse in the hippocampus the relative content of each single L1 family using the 3'UTR-MICSAT/GAPDH assay. In accordance with our first results, we observed a significant increase (~9%) of L1 sequences for the A family in transgenic mice compared to controls, but no differences were detected for the Tf and the Gf families.

As for the P0 mice, we measured on these samples also the rate of retrotransposition with the 3'UTR-5'UTR assay, and, as expected, we observed a strong increase in the rate of retrotranpsosition for the A family in transgenic hippocampus, while for Tf and Gf families no differences were detected.

These experiments confirmed the first hypothesis that we formulated according to the results obtained for the human samples. Indeed, in the cortex of transgenic mice at P0 we observed the presence of a higher amount of L1 elements, but this increase was not present at 3 months of age. On the other hand, at 3 months we observed an increase in L1 content at the level of the hippocampus. The increase seen in the cortex at P0 could have been linked to an abnormal L1 retrotransposition occurred in the process of cortical neural differentiation during the embryonic development. This accumulation of L1s at P0 could have led to cell death and only those cells in which L1 was less active survived to adulthood. At the same time, the accumulation of L1 sequences in the hippocampus at 3 months could have been linked to an abnormal L1 retrotransposition occurred in the process of hippocampal neurogenesis during adulthood.

In order to understand if neurodegeneration is one of the possible mechanisms responsible for the lack of difference in L1 content in cortexes at 3 months, we decided to test aging mice at 8 months. The total amount of L1 sequences (with the ORF2-MICSAT assay) was measured in cortex, hippocampus and kidney of mice at 8 months of age. In all the three tissues we did not observe any difference between transgenic and control mice. This result may support the hypothesis that neurodegeneration, parallel to aging and progression of the disease, depleted L1 content in the hippocampus of transgenic mice, preventing us from detecting differences between TgCRND8 mice and controls at 8 months. Further experiments, such as tests of L1 retrotransposition on single cells in this mouse model of AD, will be necessary to confirm our hypothesis.

The observation of an altered L1 content in both human sporadic AD samples and a transgenic mouse model of the disease brought out the question if there could be any differential pattern of distribution of L1 sequences in the AD pathologic context, compared to controls.

To study L1 insertion sites we decided to develop a new technique based on splinkerette PCR (spPCR), that we called SPlinkerette Analysis of Mobile elements (SPAM).

The protocol comprises a step of genomic DNA fragmentation by sonication, the end-repair of these fragments and their ligation to a synthetic double stranded adaptor (splinkerette), followed by two rounds of PCR, to amplify the genomic sequence between the repetitive element and the splinkerette, and finally sequencing.

At first we tested the SPAM technique on three DNA samples of temporal cortex from three human healthy donors. The sequencing output was analysed using the bioinformatics pipeline that we developed, and according to the position in respect of the annotated L1 sequences (collected in our L1 database), we defined a total of (on average) 1450 AIS (annotated integration sites) and 180 NIS (novel integration sites) for each sample. In the case of AIS, ~56% of the total integration sites detected were in common between the three samples, while for NIS only the ~2% was in common, indicating a high somatic variability in the integration sites of newly retrotransposed L1s.

A gene ontology (GO) study of the total AIS and NIS detected in the three samples revealed an enrichment of IS associated genes involved in brain functions, such as neuron development and differentiation. Surprisingly, in the case of genes involved in some brain functions such as axon development and guidance, we observed the presence of only NIS, suggesting that L1 mobilization may have a role in neuron and axon plasticity. The AIS and NIS that we selected were all validated by PCR: AIS were detected in both the SPAM product and the genomic DNA as expected, while NIS were detected only in the SPAM product, probably because present in only one or few cells, and therefore too poorly represented to be detected.

A preliminary SPAM experiment was performed on genomic DNA from frontal cortex and kidney of 3 AD patients and 3 controls. The low deepness of the sequencing and the high multiplex degree (sequencing runs will be repeated in order to obtain balanced and comparable results for all the samples) prevented us from obtaining a quantitative result, but interesting issues emerged.

We first normalized each amount of AIS and NIS on the initial number of MapFragments obtained for each sample, in order to avoid the bias of sequencing depth. Then we considered the ratio between NIS and AIS as a measure of somatic mobilization of full length L1s. Lower the ratio, lower the number of NIS respect to the

annotated ones; higher the ratio, higher the number of NIS respect to the AIS detected. This ratio allowed us to abolish differences in chromatin accessibility between samples. By plotting the average values of these ratios for AD and control tissues, we observed that the ratio is higher in kidney than in frontal cortex, suggesting a high rate of retrotransposition in the kidney, and that in AD frontal cortexes there is a lower ratio compared to the controls, in accordance with the qPCR data.

When we focused our attention on a list of AD-related genes (comprising genes involved in genetic and sporadic forms of the disease) we observed that in these genes the NIS/AIS ratio is high, suggesting that AD genes may be hotspots for L1 mobilization. Finally, when we looked in particular at the APP gene, we noticed that AD patients (frontal cortex and kidney) presented a higher amount of IS than controls.

This analysis will be clearly integrated with further sequencing experiments in order to obtain more reliable and quantitative data. Validations by PCR will be performed in order to validate AIS and NIS discovered in AD associated genes (not only in APP). The analysis will be applied on a great number of AD samples and controls in order to establish whether AD patients present a real differential pattern of distribution of L1 sequences. The SPAM technique seems to be suitable for integration sites discovery analysis, and these results, although preliminary, give us an idea of the unexpected complexity in terms of AIS and NIS integration sites present in the human genome.

The final part of this work regarded the adaptation and the application of the SPAM technique to a different repetitive element, the *roo* element, in a different organism: the *Drosophila melanogaster*.

The primers used in the human SPAM PCRs were substituted with primers specific for the *roo*-LTR sequence, while the bioinformatics pipeline remained constant, except for the alignment, which was performed on the *roo*-LTR reference sequence and the dm3 *Drosophila melanogaster* reference genome. We applied this protocol to two different quantities of the same genomic DNA (from adult carcasses), in order to understand if it was possible to reduce the starting input, and we analyzed the detected integration sites using in parallel different cutoffs of MapFragments. We observed that increasing the stringency of the parameters, the number of total IS decreased for both samples, until no relevant differences were detectable between the two samples. What we found surprising, was that we detected a huge amount of NIS, that probably saturated the system, preventing us from detecting all the annotated *roo*-LTR sequences present in the *Drosophila* genome.

Performing a gene ontology study on the genes associated to the *roo*-LTR annotated in our database that we didn't detect, we noticed that some of these genes were found to be involved in biological functions important during development. Since the DNA used in this experiment came from a pool of adult carcasses, it could be that *roo*-LTRs inserted in genomic loci silent during adulthood were inaccessible to the technique and therefore not detectable.

The SPAM technique therefore can be considered a scalable approach, suitable for the integration sites discovery of different kinds of repetitive elements in different organisms.

This page intentionally left blank

Bibliography

Akagi, K., Li, J., Symer, D.E. (2013). How do mammalian transposons induce genetic variation? A conceptual framework: the age, structure, allele frequency, and genome context of transposable elements may define their wide-ranging biological impacts. Bioessays 35(4):397-407.

Alagiakrishnan, K., Gill, S.S., Fagarasanu, A. (2012). Genetics and epigenetics of Alzheimer's disease. Postgrad Med J 88(1043):522-9.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. 23(2):185–188.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 31(6):785-99.

Awano, H., Malueka, R.G., Yagi, M., Okizuka, Y., Takeshima, Y., Matsuo, M. (2010). Contemporary retrotransposition of a novel non-coding gene induces exon-skipping in dystrophin mRNA. J Hum Genet 55(12):785-90.

Babushok, D.V., Kazazian, H.H. Jr (2007). Progress in understanding the biology of the human mutagen LINE-1. Hum Mutat 28(6):527-39.

Babushok, D.V., Ostertag, E.M., Courtney, C.E., Choi, J.M., Kazazian, H.H. Jr (2006). L1 integration in a transgenic mouse model. Genome Res. 16(2):240-50.

Badge, R.M., Alisch, R.S., Moran, J.V. (2003). ATLAS: a system to selectively identify human-specific L1 insertions. Am J Hum Genet 72(4):823-38.

Baillie, J.K., Barnett, M.W., Upton, K.R., Gerhardt, D.J., Richmond, T.A., De Sapio, F., Brennan, P.M., Rizzu, P., Smith, S., Fell, M., Talbot, R.T., Gustincich, S., Freeman,

T.C., Mattick, J.S., Hume, D.A., Heutink, P., Carninci, P., Jeddeloh, J.A., Faulkner, G.J. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. Nature. 2011 479(7374):534-7.

Ballatore, C., Lee, V.M., Trojanowski, J.Q. (2007). Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. Nat Rev Neurosci 8(9):663-72.

Bar-Shira, A., Rashi-Elkeles, S., Zlochover, L., Moyal, L., Smorodinsky, N.I., Seger, R., Shiloh, Y. (2002). ATM-dependent activation of the gene encoding MAP kinase phosphatase 5 by radiomimetic DNA damage. Oncogene 21(5):849-55.

Barrachina, M., Ferrer, I. (2009). DNA methylation of Alzheimer disease and tauopathy-related genes in postmortem brain. J Neuropathol Exp Neurol 68(8):880-91.

Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. Cell 141(7):1159-70.

Beck, C.R., Garcia-Perez, J.L., Badge, R.M., Moran, J.V. (2011). LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet 12:187-215.

Belancio, V.P., Hedges, D.J., Deininger, P. (2008). Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 18(3):343-58.

Belancio, V.P., Whelton, M., Deininger, P. (2007). Requirements for polyadenylation at the 3' end of LINE-1 elements. Gene 390(1-2):98-107.

Belgnaoui, S.M., Gosden, R.G., Semmes, O.J., Haoudi, A. (2006). Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. Cancer Cell Int 6:13.

Bénit, L., Lallemand, J.B., Casella, J.F., Philippe, H., Heidmann, T. (1999). ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. J Virol 73(4):3301-8.

Beraldi, R., Pittoggi, C., Sciamanna, I., Mattei, E., Spadafora, C. (2006). Expression of LINE-1 retroposons is essential for murine preimplantation development. Mol Reprod Dev 73(3):279-87.

Bernard, V., Minnerop, M., Bürk, K., Kreuz, F., Gillessen-Kaesbach, G., Zühlke, C. (2009). Exon deletions and intragenic insertions are not rare in ataxia with oculomotor apraxia 2. BMC Med Genet 10:87.

Berry, C., Hannenhalli, S., Leipzig, J., Bushman, F.D. (2006). Selection of target sites for mobile DNA integration in the human genome. PLoS Comput Biol 2(11):e157.

Bertram, L., Parrado, A.R., Tanzi, R.E. (2013). TREM2 and neurodegenerative disease. N Engl J Med 369(16):1565.

Bettens, K., Sleegers, K., Van Broeckhoven, C. (2013). Genetic insights in Alzheimer's disease. Lancet Neurol 12(1):92-104.

Bierer, L.M., Hof, P.R., Purohit, D.P., Carlin, L., Schmeidler, J., Davis, K.L., Perl, D.P. (1995). Neocortical neurofibrillary tangles correlate with dementia severity in Alzheimer's disease. Arch Neurol 52(1):81-8.

Bodega, B., Orlando, V. (2014). Repetitive elements dynamics in cell identity programming, maintenance and disease. Curr Opin Cell Biol 31C:67-73.

Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V., Cullen, B.R. (2006). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci USA 103(23):8780-5.

Boissinot, S., Chevret, P., Furano, A.V. (2000). L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol Biol Evol 17(6):915-28.

Bourc'his, D., Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431(7004):96-9.

Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta Neuropathol 112(4):389-404.

Braak, H., Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 82(4):239-59.

Braak, H., Braak, E. (1997). Frequency of stages of Alzheimer-related lesions in different age categories. Neurobiol Aging 18(4):351-7.

Braak, H., Del Tredici, K. (2013). Amyloid-β may be released from non-junctional varicosities of axons generated from abnormal tau-containing brainstem nuclei in sporadic Alzheimer's disease: a hypothesis. Acta Neuropathol 126(2):303-6.

Braak, H., Thal, D.R., Ghebremedhin, E., Del Tredici, K. (2011). Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years. J Neuropathol Exp Neurol 70(11):960-9.

Branciforte, D., Martin, S.L. (1994). Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. Mol Cell Biol 14(4):2584-92.

Brouha, B., Meischl, C., Ostertag, E., de Boer, M., Zhang, Y., Neijens, H., Roos, D., Kazazian, H.H. Jr (2002). Evidence consistent with human L1 retrotransposition in maternal meiosis I. Am J Hum Genet 71(2):327-36.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., Kazazian, H.H. Jr (2003). Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci USA 100(9):5280-5.

Brunner, E., Brunner, D., Fu, W., Hafen, E., Basler, K. (1999). The dominant mutation Glazed is a gain-of-function allele of wingless that, similar to loss of APC, interferes with normal eye development. Dev Biol 206(2):178-88.

Brunnström, H.R., Englund, E.M. (2009). Cause of death in patients with dementia disorders. Eur J Neurol 16(4):488-92.

Bundo, M., Toyoshima, M., Okada, Y., Akamatsu, W., Ueda, J., Nemoto-Miyauchi, T., Sunaga, F., Toritsuka, M., Ikawa, D., Kakita, A., Kato, M., Kasai, K., Kishimoto, T., Nawa, H., Okano, H., Yoshikawa, T., Kato, T., Iwamoto, K. (2014). Increased L1 retrotransposition in the Neuronal Genome in Schizophrenia. Neuron 81(2):306-13.

Calero, M., Rostagno, A., Matsubara, E., Zlokovic, B., Frangione, B., Ghiso, J. (2000). Apolipoprotein J (clusterin) and Alzheimer's disease. Microsc Res Tech 15;50(4):305-15.

Callinan, P.A., Batzer, M.A. (2006). Retrotransposable elements and human disease. Genome Dyn 1:104-15.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell 12(4):503-14.

Cavallucci, V., D'Amelio, M., Cecconi, F. (2012). Aβ toxicity in Alzheimer's disease. Mol Neurobiol 45(2):366-78.

Celniker, S.E., Wheeler, D.A., Kronmiller, B., Carlson, J.W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S.P., Frise, E., Hodgson, A., George, R.A., Hoskins, R.A., Laverty, T., Muzny, D.M., Nelson, C.R., Pacleb, J.M., Park, S., Pfeiffer, B.D., Richards, S., Sodergren, E.J., Svirskas, R., Tabor, P.E., Wan, K., Stapleton, M., Sutton, G.G., Venter, C., Weinstock, G., Scherer, S.E., Myers, E.W., Gibbs, R.A., Rubin, G.M. (2002). Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence. Genome Biol 3(12):RESEARCH0079.

Chapuis, J., Hansmannel, F., Gistelinck, M., Mounier, A., Van Cauwenberghe, C., Kolen, K.V., Geller, F., Sottejeau, Y., Harold, D., Dourlen, P., Grenier-Boley, B., Kamatani, Y., Delepine, B., Demiautte, F., Zelenika, D., Zommer, N., Hamdane, M., Bellenguez, C., Dartigues, J.F., Hauw, J.J., Letronne, F., Ayral, A.M., Sleegers, K., Schellens, A., Broeck, L.V., Engelborghs, S., De Deyn, P.P., Vandenberghe, R., O'Donovan, M., Owen, M., Epelbaum, J., Mercken, M., Karran, E., Bantscheff, M., Drewes, G., Joberty, G., Campion, D., Octave, J.N., Berr, C., Lathrop, M., Callaerts, P., Mann, D., Williams, J., Buée, L., Dewachter, I., Van Broeckhoven, C., Amouyel, P., Moechars, D., Dermaut, B., Lambert, J.C.; GERAD consortium (2013). Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. Mol Psychiatry 18(11):1225-34.

Chen, K.L., Wang, S.S., Yang, Y.Y., Yuan, R.Y., Chen, R.M., Hu, C.J. (2009). The epigenetic effects of amyloid-beta(1-40) on global DNA and neprilysin genes in murine cerebral endothelial cells. Biochem Biophys Res Commun 378(1):57-61.

Chin, J. (2011). Selecting a mouse model of Alzheimer's disease. Methods Mol Biol 670:169-89.

Chishti, M.A., Yang, D.S., Janus, C., Phinney, A.L., Horne, P., Pearson, J., Strome, R., Zuker, N., Loukides, J., French, J., Turner, S., Lozza, G., Grilli, M., Kunicki, S., Morissette, C., Paquette, J., Gervais, F., Bergeron, C., Fraser, P.E., Carlson, G.A., George-Hyslop, P.S., Westaway, D. (2001). Early-onset amyloid deposition and cognitive deficits in transgenic mice expressing a double mutant form of amyloid precursor protein 695. J Biol Chem 276(24):21562-70.

Chouliaras, L., Mastroeni, D., Delvaux, E., Grover, A., Kenis, G., Hof, P.R., Steinbusch, H.W.M., Coleman, P.D., Rutten, B.P.F., van den Hove, D.L.A. (2013). Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. Neurobiol Aging. 34(9):2091-9.

Chouliaras, L., Rutten, B.P.F., Kenis, G., Peerbooms, O., Visser, P.J., Verhey, F., van Os, J., Steinbusch, H.W.M., van den Hove, D.L.A. (2010). Epigenetic regulation in the pathophysiology of Alzheimer's disease. Prog Neurobiol. 90(4):498-510.

Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., Greally, J.M., Voinnet, O., Heard, E. (2011). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell 141(6):956-69.

Cordaux, R., Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat Rev Genet 10(10):691-703.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Marchetto, M.C., Muotri, A.R., Mu, Y., Carson, C.T., Macia, A., Moran, J.V., Gage, F.H. (2011). Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. Proc Natl Acad Sci USA 108(51):20382-7.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V., Gage, F.H. (2009). L1 retrotransposition in human neural progenitor cells. Nature 460(7259):1127-31.

Crews, L., Masliah, E. (2010). Molecular mechanisms of neurodegeneration in Alzheimer's disease. Hum Mol Genet 19(R1):R12-20.

Crichton, J.H., Dunican, D.S., Maclennan, M., Meehan, R.R., Adams, I.R. (2014). Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. Cell Mol Life Sci 71(9):1581-605.

Crouch, P.J., Harding, S.M., White, A.R., Camakaris, J., Bush, A.I., Masters, C.L. (2008). Mechanisms of A beta mediated neurodegeneration in Alzheimer's disease. Int J Biochem Cell Biol 40(2):181-98.

Cruchaga, C., Karch, C.M., Jin, S.C., Benitez, B.A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., Jeng, A.T., Cooper, B., Skorupa, T., Carrell, D., Levitch, D., Hsu, S., Choi, J., Ryten, M.; UK Brain Expression Consortium, Hardy, J., Ryten, M., Trabzuni, D., Weale, M.E., Ramasamy, A., Smith, C., Sassi, C., Bras, J., Gibbs, J.R., Hernandez, D.G., Lupton, M.K., Powell, J., Forabosco, P., Ridge, P.G.,

Corcoran, C.D., Tschanz, J.T., Norton, M.C., Munger, R.G., Schmutz, C., Leary, M., Demirci, F.Y., Bamne, M.N., Wang, X., Lopez, O.L., Ganguli, M., Medway, C., Turton, J., Lord, J., Braae, A., Barber, I., Brown, K.; Alzheimer's Research UK Consortium, Passmore, P., Craig, D., Johnston, J., McGuinness, B., Todd, S., Heun, R., Kölsch, H., Kehoe, P.G., Hooper, N.M., Vardy, E.R., Mann, D.M., Pickering-Brown, S., Brown, K., Kalsheker, N., Lowe, J., Morgan, K., David Smith, A., Wilcock, G., Warden, D., Holmes, C., Pastor, P., Lorenzo-Betancor, O., Brkanac, Z., Scott, E., Topol, E., Morgan, K., Rogaeva, E., Singleton, A.B., Hardy, J., Kamboh, M.I., St George-Hyslop, P., Cairns, N., Morris, J.C., Kauwe, J.S., Goate, A.M. (2014). Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 505(7484):550-4.

Cruts, M., Van Broeckhoven, C. (1998). Presenilin mutations in Alzheimer's disease. Hum Mutat 11(3):183-90.

de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 7(12):e1002384.

de la Chaux, N., Wagner, A. (2009). Evolutionary dynamics of the LTR retrotransposons roo and rooA inferred from twelve complete Drosophila genomes. BMC Evol Biol 9:205.

DeMattos, R.B., Cirrito, J.R., Parsadanian, M., May, P.C., O'Dell, M.A., Taylor, J.W., Harmony, J.A., Aronow, B.J., Bales, K.R., Paul, S.M., Holtzman, D.M. (2004). ApoE and clusterin cooperatively suppress Abeta levels and deposition: evidence that ApoE regulates extracellular Abeta metabolism in vivo. Neuron 22;41(2):193-202.

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

Dewannieux, M., Heidmann, T. (2005). L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. J Mol Biol 349:241–247.

Ding, W., Lin, L., Chen, B., Dai, J. (2006). L1 elements, processed pseudogenes and retrogenes in mammalian genomes. IUBMB Life 58(12):677-85.

Erwin, J.A., Marchetto, M.C., Gage, F.H. (2014). Mobile DNA elements in the generation of diversity and complexity in the brain. Nat Rev Neurosci 15(8):497-506.

Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, Parker JJ, Atabay KD, Gilmore EC, Poduri A, Park PJ, Walsh CA (2012). Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell 151(3):483-96.

Ewing, A.D., Kazazian, H.H. Jr (2010). High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res 20(9):1262-70.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., Waki, K., Hornig, N., Arakawa, T., Takahashi, H., Kawai, J., Forrest, A.R., Suzuki, H., Hayashizaki, Y., Hume, D.A., Orlando, V., Grimmond, S.M., Carninci, P. (2009). The regulated retrotransposon transcriptome of mammalian cells. Nat Genet 41(5):563-71.

Fedoroff, N.V. (2012). McClintock's challenge in the 21st century. Proc Natl Acad Sci USA 109(50):20200-3.

Ferrer, I. (2012). Defining Alzheimer as a common age-related neurodegenerative process not inevitably leading to dementia. Prog Neurobiol 97(1):38-51.

Ferri, C.P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., Hall, K., Hasegawa, K., Hendrie, H., Huang, Y., Jorm, A., Mathers, C., Menezes, P.R., Rimmer, E., Scazufca, M.; Alzheimer's Disease International (2005). Global prevalence of dementia: a Delphi consensus study. Lancet 366(9503):2112-7.

Fitzsimons, C.P., van Bodegraven, E., Schouten, M., Lardenoije, R., Kompotis, K., Kenis, G., van den Hurk, M., Boks, M.P., Biojone, C., Joca, S., Steinbusch, H.W., Lunnon, K., Mastroeni, D.F., Mill, J., Lucassen, P.J., Coleman, P.D., van den Hove,

D.L., Rutten, B.P. (2014). Epigenetic regulation of adult neural stem cells: implications for Alzheimer's disease. Mol Neurodegener 9:25.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T.D., Wu, Y.Z., Plass, C., Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci USA 102(30):10604-9.

Frame, I.G., Cutfield, J.F., Poulter, R.T. (2001). New BEL-like LTR-retrotransposons in Fugu rubripes, Caenorhabditis elegans, and Drosophila melanogaster. Gene 263(1-2):219-30.

Friedrich, R.P., Tepper, K., Rönicke, R., Soom, M., Westermann, M., Reymann, K., Kaether, C., Fändrich, M. (2010). Mechanism of amyloid plaque formation suggests an intracellular basis of Abeta pathogenicity. Proc Natl Acad Sci 107(5):1942-7.

Gabriel, A., Dapprich, J., Kunkel, M., Gresham, D., Pratt, S.C., Dunham, M.J. (2006). Global mapping of transposon location. PLoS Genet 2(12):e212.

Galimberti, D., Scarpini, E. (2012). Progress in Alzheimer's disease. J Neurol 259(2):201-11

Gao, Y., Sun, Y., Frank, K.M., Dikkes, P., Fujiwara, Y., Seidl, K.J., Sekiguchi, J.M., Rathbun, G.A., Swat, W., Wang, J., Bronson, R.T., Malynn, B.A., Bryans, M., Zhu, C., Chaudhuri, J., Davidson, L., Ferrini, R., Stamato, T., Orkin, S.H., Greenberg, M.E., Alt, F.W. (1998). A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. Cell 95(7):891-902.

Gasior, S.L., Wakeman, T.P., Xu, B., Deininger, P.L. (2006). The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol 357(5):1383-93.

Georgiou, I., Noutsopoulos, D., Dimitriadou, E., Markopoulos, G., Apergi, A., Lazaros, L., Vaxevanoglou, T., Pantos, K., Syrrou, M., Tzavaras, T. (2009). Retrotransposon

RNA expression and evidence for retrotransposition events in human oocytes. Hum Mol Genet 18(7):1221-8.

Gerrish, A., Russo, G., Richards, A., Moskvina, V., Ivanov, D., Harold, D., Sims, R., Abraham, R., Hollingworth, P., Chapman, J., Hamshere, M., Pahwa, J.S., Dowzell, K., Williams, A., Jones, N., Thomas, C., Stretton, A., Morgan, A.R., Lovestone, S., Powell, J., Proitsi, P., Lupton, M.K., Brayne, C., Rubinsztein, D.C., Gill, M., Lawlor, B., Lynch, A., Morgan, K., Brown, K.S., Passmore, P.A., Craig, D., McGuinness, B., Todd, S., Johnston, J.A., Holmes, C., Mann, D., Smith, A.D., Love, S., Kehoe, P.G, Hardy, J., Mead, S., Fox, N., Rossor, M., Collinge, J., Maier, W., Jessen, F., Kölsch, H., Heun, R., Schürmann, B., van den Bussche, H., Heuser, I., Kornhuber, J., Wiltfang, J., Dichgans, M., Frölich, L., Hampel, H., Hüll, M., Rujescu, D., Goate, A.M., Kauwe, J.S., Cruchaga, C., Nowotny, P., Morris, J.C., Mayo, K., Livingston, G., Bass, N.J., Gurling, H., McQuillin, A., Gwilliam, R., Deloukas, P., Davies, G., Harris, S.E., Starr, J.M., Deary, I.J., Al-Chalabi, A., Shaw, C.E., Tsolaki, M., Singleton, A.B., Guerreiro, R., Mühleisen, T.W., Nöthen, M.M., Moebus, S., Jöckel, K.H., Klopp, N., Wichmann, H.E., Carrasquillo, M.M., Pankratz, V.S., Younkin, S.G., Jones, L., Holmans, P.A., O'Donovan, M.C., Owen, M.J., Williams, J. (2012). The role of variation at AβPP, PSEN1, PSEN2, and MAPT in late onset Alzheimer's disease. J Alzheimers Dis 28(2):377-87.

Gilbert, N., Lutz-Prigge, S., Moran, J.V. (2002). Genomic deletions created upon LINE-1 retrotransposition. Cell 110(3):315-25.

Gire, V., Roux, P., Wynford-Thomas, D., Brondello, J.M., Dulic, V. (2004). DNA damage checkpoint kinase Chk2 triggers replicative senescence. EMBO J 23(13):2554-63.

Goodier, J.L., Kazazian, H.H. Jr (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell 135(1):23-35.

Goodier, J.L., Ostertag, E.M., Du, K., Kazazian, H.H. Jr (2001). A novel active L1 retrotransposon subfamily in the mouse. Genome Res 11(10):1677-85.

Goodier, J.L., Ostertag, E.M., Kazazian, H.H. Jr (2000). Transduction of 3'-flanking sequences is common in L1 retrotransposition. Hum Mol Genet 9(4):653-7.

Goodier, J.L., Zhang, L., Vetter, M.R., Kazazian, H.H. Jr (2007). LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. Mol Cell Biol 27(18):6469-83.

Gräff, J., Kim, D., Dobbin, M.M., Tsai, L.H. (2011). Epigenetic regulation of gene expression in physiological and pathological brain processes. Physiol Rev 91(2):603-49.

Gräff, J., Mansuy, I.M. (2008). Epigenetic codes in cognition and behaviour. Behav Brain Res 192(1):70-87.

Gräff, J., Rei, D., Guan, J.S., Wang, W.Y., Seo, J., Hennig, K.M., Nieland, T.J., Fass, D.M., Kao, P.F., Kahn, M., Su, S.C., Samiei, A., Joseph, N., Haggarty, S.J., Delalle, I., Tsai, L.H. (2012). An epigenetic blockade of cognitive functions in the neurodegenerating brain. Nature 29;483(7388):222-6.

Griciuc, A., Serrano-Pozo, A., Parrado, A.R., Lesinski, A.N., Asselin, C.N., Mullin, K., Hooli, B., Choi, S.H., Hyman, B.T., Tanzi, R.E. (2013). Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. Neuron 78(4):631-43.

Han, K., Sen, S.K., Wang, J., Callinan, P.A., Lee, J., Cordaux, R., Liang, P., Batzer, M.A. (2005). Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. Nucleic Acids Res 33(13):4040-52.

Hancks, D.C., Kazazian, H.H. Jr (2012). Active human retrotransposons: variation and disease. Curr Opin Genet Dev 22:191–203.

Haoudi, A., Semmes, O.J., Mason, J.M., Cannon, R.E. (2004). Retrotransposition-Competent Human LINE-1 Induces Apoptosis in Cancer Cells With Intact p53. J Biomed Biotechnol 2004(4):185-194.

Heckmann, J.M., Low, W.C., de Villiers, C., Rutherfoord, S., Vorster, A., Rao, H., Morris, C.M., Ramesar, R.S., Kalaria, R.N. (2004). Novel presenilin 1 mutation with profound neurofibrillary pathology in an indigenous Southern African family with early-onset Alzheimer's disease. Brain 127(Pt 1):133-42.

Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., Kazazian, H.H. Jr (1994). A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. Nat Genet 7(2):143-8.

Holtzman, D.M., Morris, J.C., Goate, A.M. (2011). Alzheimer's disease: the challenge of the second century. Sci Transl Med 3(77):77sr1.

Hong-Qi, Y., Zhi-Kun, S., Sheng-Di, C. (2012). Current advances in the treatment of Alzheimer's disease: focused on considerations targeting Aβ and tau. Transl Neurodegener 1(1):21.

Huang, C.R.L., Burns, K.H., Boeke, J.D. (2012). Active transposition in genomes. Annu Rev Genet 46:651-675.

Huang, C.R., Schneider, A.M., Lu, Y., Niranjan, T., Shen, P., Robinson, M.A., Steranka, J.P., Valle, D., Civin, C.I., Wang, T., Wheelan, S.J., Ji, H., Boeke, J.D., Burns, K.H. (2010). Mobile interspersed repeats are major structural variants in the human genome. Cell 141(7):1171-82.

Hyman, B.T., Phelps, C.H., Beach, T.G., Bigio, E.H., Cairns, N.J., Carrillo, M.C., Dickson, D.W., Duyckaerts, C., Frosch, M.P., Masliah, E., Mirra, S.S., Nelson, P.T., Schneider, J.A., Thal, D.R., Thies, B., Trojanowski, J.Q., Vinters, H.V., Montine, T.J. (2012). National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. Alzheimers Dement 8(1):1-13.

Iqbal, K., Grundke-Iqbal, I. (2008). Alzheimer neurofibrillary degeneration: significance, etiopathogenesis, therapeutics and prevention. J Cell Mol Med 12(1):38-55.

Iqbal, K., Liu, F., Gong, C.X., Grundke-Iqbal, I. (2010). Tau in Alzheimer disease and related tauopathies. Curr Alzheimer Res 7(8):656-64.

Isik, A.T. (2010). Late onset Alzheimer's disease in older people. Clin Interv Aging 5:307-11.

Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M., Devine, S.E. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. Cell 141(7):1253-61.

Jiang, T., Yu, J.T., Tian, Y., Tan, L. (2013). Epidemiology and etiology of Alzheimer's disease: from genetic to non-genetic factors. Curr Alzheimer Res 10(8):852-67.

Kaer, K., Speek, M. (2013). Retroelements in human disease. Gene 518(2):231-41.

Kalia, M. (2003). Dysphagia and aspiration pneumonia in patients with Alzheimer's disease. Metabolism 52(10 Suppl 2):36-8.

Kamenetz, F., Tomita, T., Hsieh, H., Seabrook, G., Borchelt, D., Iwatsubo, T., Sisodia, S., Malinow, R. (2003). APP processing and synaptic function. Neuron 37(6):925-37.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., Ashburner, M., Celniker, S.E. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol 3(12):RESEARCH0084.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev 19(4):489-501.

Kano, H., Godoy, I., Courtney, C., Vetter, M.R., Gerton, G.L., Ostertag, E.M., Kazazian, H.H. Jr (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes Dev 23(11):1303-12.

Karch, C.M., Cruchaga, C., Goate, A.M. (2014). Alzheimer's disease genetics: from the bench to the clinic. Neuron 83(1):11-26.

Karch, C.M., Goate, A.M. (2014). Alzheimer's Disease Risk Genes and Mechanisms of Disease Pathogenesis. Biol Psychiatry 77(1):43-51.

Karch, C.M., Jeng, A.T., Nowotny, P., Cady, J., Cruchaga, C., Goate, A.M. (2012). Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. PLoS One 7(11):e50976.

Katoh, I., Kurata, S. (2013). Association of endogenous retroviruses and long terminal repeats with human disorders. Front Oncol 11;3:234.

Kazazian, H.H. Jr (2004). Mobile elements: drivers of genome evolution. Science 303(5664):1626-32.

Kazazian, H.H. Jr, Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., Antonarakis, S.E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332(6160):164-6.

Keane, T.M., Wong, K., Adams, D.J., Flint, J., Reymond, A., Yalcin, B. (2014). Identification of structural variation in mouse genomes. Front Genet 5:192.

Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., Haugen, E., Zerr, T., Yamada, N.A., Tsang, P., Newman, T.L., Tüzün, E., Cheng, Z., Ebling, H.M., Tusneem, N., David, R., Gillett, W., Phelps, K.A., Weaver, M., Saranga, D., Brand, A., Tao, W., Gustafson, E., McKernan, K., Chen, L., Malig, M., Smith, J.D., Korn, J.M., McCarroll, S.A., Altshuler, D.A., Peiffer, D.A., Dorschner, M., Stamatoyannopoulos, J., Schwartz, D., Nickerson, D.A., Mullikin, J.C., Wilson, R.K., Bruhn, L., Olson, M.V., Kaul, R., Smith, D.R., Eichler, E.E. (2008). Mapping and sequencing of structural variation from eight human genomes. Nature 453(7191):56-64.

Kim, D.H., Yeo, S.H., Park, J.M., Choi, J.Y., Lee, T.H., Park, S.Y., Ock, M.S., Eo, J., Kim, H.S., Cha, H.J. (2014). Genetic markers for diagnosis and pathogenesis of Alzheimer's disease. Gene 545(2):185-93.

Kim, J., Basak, J.M., Holtzman, D.M. (2009a). The role of apolipoprotein E in Alzheimer's disease. Neuron 63(3):287-303.

Kim, J., Castellano, J.M., Jiang, H., Basak, J.M., Parsadanian, M., Pham, V., Mason, S.M., Paul, S.M., Holtzman, D.M. (2009b). Overexpression of low-density lipoprotein receptor in the brain markedly inhibits amyloid deposition and increases extracellular A beta clearance. Neuron 64(5):632-44.

Kim, W.S., Guillemin, G.J., Glaros, E.N., Lim, C.K., Garner, B. (2006). Quantitation of ATP-binding cassette subfamily-A transporter gene expression in primary human brain cells. Neuroreport 17(9):891-6.

Kim, W.S., Li, H., Ruberu, K., Chan, S., Elliott, D.A., Low, J.K., Cheng, D., Karl, T., Garner, B. (2013). Deletion of Abca7 increases cerebral amyloid-β accumulation in the J20 mouse model of Alzheimer's disease. J Neurosci 33(10):4387-94.

Kim, W.S., Weickert, C.S., Garner, B. (2008). Role of ATP-binding cassette transporters in brain lipid transport and neurological disease. J Neurochem 104(5):1145-66.

Kimberland, M.L., Divoky, V., Prchal, J., Schwahn, U., Berger, W., Kazazian, H.H. Jr (1999). Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. Hum Mol Genet 8(8):1557-60.

Kitazawa, M., Medeiros, R., Laferla, F.M. (2012). Transgenic mouse models of Alzheimer disease: developing a better model as a tool for therapeutic interventions. Curr Pharm Des 18(8):1131-47.

Kondo-Iida, E., Kobayashi, K., Watanabe, M., Sasaki, J., Kumagai, T., Koide, H., Saito, K., Osawa, M., Nakamura, Y., Toda, T. (1999). Novel mutations and genotype-

phenotype relationships in 107 families with Fukuyama-type congenital muscular dystrophy (FCMD). Hum Mol Genet 8(12):2303-9.

Konkel, M.K., Batzer, M.A. (2010). A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. Semin Cancer Biol 20(4):211-21.

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., Taillon, B.E., Chen, Z., Tanzer, A., Saunders, A.C., Chi, J., Yang, F., Carter, N.P., Hurles, M.E., Weissman, S.M., Harkins, T.T., Gerstein, M.B., Egholm, M., Snyder, M. (2007). Paired-end mapping reveals extensive structural variation in the human genome. Science 318(5849):420-6.

Kosik, K.S. (2013). Diseases: Study neuron networks to tackle Alzheimer's. Nature 503(7474):31-2.

Koudijs, M.J., Klijn, C., van der Weyden, L., Kool, J., ten Hoeve, J., Sie, D., Prasetyanti, P.R., Schut, E., Kas, S., Whipp, T., Cuppen, E., Wessels, L., Adams, D.J., Jonkers, J. (2011). High-throughput semiquantitative analysis of insertional mutations in heterogeneous tumors. Genome Res 21(12):2181-9.

Koval, A.P., Veniaminova, N.A., Kramerov, D.A. (2011). Additional box B of RNA polymerase III promoter in SINE B1 can be functional. Gene 10;487(2):113-7.

Krych-Goldberg, M., Moulds, J.M., Atkinson, J.P. (2002). Human complement receptor type 1 (CR1) binds to a major malarial adhesin. Trends Mol Med 8(11):531-7.

Kuff, E.L., Lueders, K.K. (1988). The intracisternal A-particle gene family: structure and functional aspects. Adv Cancer Res 51:183-276.

Kulpa, D.A., Moran, J.V. (2006). Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nat Struct Mol Biol 13(7):655-60

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., Hata, K., Li, E., Matsuda, Y.,

Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H., Nakano, T. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev 22(7):908-17.

LaFerla, F.M., Oddo, S. (2005). Alzheimer's disease: Abeta, tau and synaptic dysfunction. Trends Mol Med 11(4):170-6.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R.,

McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J.; International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature 409(6822):860-921.

Lavie, L., Maldener, E., Brouha, B., Meese, E.U., Mayer, J. (2004). The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. Genome Res 14(11):2253-60.

Levin, H.L., Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 18;12(9):615-27.

Li, X., Scaringe, W.A., Hill, K.A., Roberts, S., Mengos, A., Careri, D., Pinto, M.T., Kasper, C.K., Sommer, S.S. (2001). Frequency of recent retrotransposition events in the human factor IX gene. Hum Mutat 17(6):511-9.

Liu, D., Niu, Z.X. (2009). The structure, genetic polymorphisms, expression and biological functions of complement receptor type 1 (CR1/CD35). Immunopharmacol Immunotoxicol 31(4):524-35.

Livak, K.J., Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25(4):402-8.

Lyon, M.F. (1998). X-chromosome inactivation: a repeat hypothesis. Cytogenet Cell Genet 80(1-4):133-7.

Macia, A., Muñoz-Lopez, M., Cortes, J.L., Hastings, R.K., Morell, S., Lucena-Aguilar, G., Marchal, J.A., Badge, R.M., Garcia-Perez, J.L. (2011). Epigenetic control of retrotransposon expression in human embryonic stem cells. Mol Cell Biol 31(2):300-16.

Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., Mager, L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet 2:e2.

Malik, H.S., Burke, W.D., Eickbush, T.H. (1999). The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 16(6):793-805.

Malik, M., Simpson, J.F., Parikh, I., Wilfred, B.R., Fardo, D.W., Nelson, P.T., Estus, S. (2013). CD33 Alzheimer's risk-altering polymorphism, CD33 expression, and exon 2 splicing. J Neurosci 33(33):13320-5.

Malki, S., van der Heijden, G.W., O'Donnell, K.A., Martin, S.L., Bortvin, A. (2014). A role for retrotransposon LINE-1 in fetal oocyte attrition in mice. Dev Cell 29(5):521-33.

Malone, C.D., Hannon, G.J. (2009). Molecular evolution of piRNA and transposon control pathways in Drosophila. Cold Spring Harb Symp Quant Biol 74:225-34.

Mandell, J.W., Banker, G.A. (1996). Microtubule-associated proteins, phosphorylation gradients, and the establishment of neuronal polarity. Perspect Dev Neurobiol 4(2-3):125-35.

Martin, S.L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., Kwok, S.C., Hodges, R.S., Williams, M.C. (2005). LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. J Mol Biol 348(3):549-61.

Martínez, A., Otal, R., Sieber, B.A., Ibáñez, C., Soriano, E. (2005). Disruption of ephrin-A/EphA binding alters synaptogenesis and neural connectivity in the hippocampus. Neuroscience 135(2):451-61.

Martinez-Canabal, A. (2014). Reconsidering hippocampal neurogenesis in Alzheimer's disease. Front Neurosci 8:147

Martínez-Garay, I., Ballesta, M.J., Oltra, S., Orellana, C., Palomeque, A., Moltó, M.D., Prieto, F., Martínez, F. (2003). Intronic L1 insertion and F268S, novel mutations in RPS6KA3 (RSK2) causing Coffin-Lowry syndrome. Clin Genet 64(6):491-6.

Maside, X., Bartolomé, C., Assimacopoulos, S., Charlesworth, B. (2001). Rates of movement and distribution of transposable elements in Drosophila melanogaster: in situ hybridization vs Southern blotting data. Genet Res 78(2):121-36.

Mastroeni, D., Grover, A., Delvaux, E., Whiteside, C., Coleman, P.D., Rogers, J. (2008). Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. Neurobiol Aging 31(12):2025-37.

Mätlik, K., Redik, K., Speek, M. (2006). L1 antisense promoter drives tissue-specific transcription of human genes. J Biomed Biotechnol 2006(1):71753.

Matthews, F.E., Arthur, A., Barnes, L.E., Bond, J., Jagger, C., Robinson, L., Brayne, C.; Medical Research Council Cognitive Function and Ageing Collaboration. (2013). A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. Lancet 382(9902):1405-12.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34(7):939-44.

Mears, M.L., Hutchison, C.A. 3rd (2001). The evolution of modern lineages of mouse L1 elements. J Mol Evol 52(1):51-62.

Meischl, C., Boer, M., Ahlin, A., Roos, D. (2000). A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. Eur J Hum Genet 8(9):697-703.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., Nakamura, Y. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res 52(3):643-5.

Mikkers, H., Allen, J., Knipscheer, P., Romeijn, L., Hart, A., Vink, E., Berns, A. (2002). High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. Nat Genet 32(1):153-9.

Miné, M., Chen, J.M., Brivet, M., Desguerre, I., Marchant, D., de Lonlay, P., Bernard, A., Férec, C., Abitbol, M., Ricquier, D., Marsac, C. (2007). A large genomic deletion in the PDHX gene caused by the retrotranspositional insertion of a full-length LINE-1 element. Hum Mutat 28(2):137-42.

Mohamed, N.V., Herrou, T., Plouffe, V., Piperno, N., Leclerc, N. (2013). Spreading of tau pathology in Alzheimer's disease by cell-to-cell transmission. Eur J Neurosci 37(12):1939-48.

Moran, J.V., DeBerardinis, R.J., Kazazian, H.H. Jr (1999). Exon shuffling by L1 retrotransposition. Science 283(5407):1530-4.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., Kazazian, H.H. Jr (1996). High frequency retrotransposition in cultured mammalian cells. Cell 87(5):917-27.

Morris, J.C., Roe, C.M., Xiong, C., Fagan, A.M., Goate, A.M., Holtzman, D.M., Mintun, M.A. (2010). APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. Ann Neurol 67(1):122-31.

Morrish, T.A., Garcia-Perez, J.L., Stamato, T.D., Taccioli, G.E., Sekiguchi, J., Moran, J.V. (2007). Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. Nature 446(7132):208-12.

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., Moran, J.V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet 31(2):159-65.

Morrison, L.D., Smith, D.D., Kish, S.J. (1996). Brain S-adenosylmethionine levels are severely decreased in Alzheimer's disease. J Neurochem 67(3):1328-31.

Mukherjee, S., Mukhopadhyay, A., Banerjee, D., Chandak, G.R., Ray, K. (2004). Molecular pathology of haemophilia B: identification of five novel mutations including a LINE 1 insertion in Indian patients. Haemophilia 10(3):259-63.

Muñoz-López, M., García-Pérez, J.L. (2010). DNA transposons: nature and applications in genomics. Curr Genomics 11(2):115-28.

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature 435(7044):903-10.

Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. Nature 468(7322):443-6. doi: 10.1038/nature09544.

Muotri, A.R., Zhao, C., Marchetto, M.C., Gage, F.H. (2009). Environmental influence on L1 retrotransposons in the adult hippocampus. Hippocampus 19(10):1002-7.

Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Ishikawa, Y., Minami, R., Nakamura, H., Matsuo, M. (1993). Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. J Clin Invest 91(5):1862-7.

O'Donnell, K.A., Burns, K.H. (2010). Mobilizing diversity: transposable element insertions in genetic variation and disease. Mob DNA 1(1):21.

Ohshima, K., Hattori, M., Yada, T., Gojobori, T., Sakaki, Y., Okada, N. (2003). Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. Genome Biol 4(11):R74.

Ohshima, K., Okada, N. (1994). Generality of the tRNA origin of short interspersed repetitive elements (SINEs). Characterization of three different tRNA-derived retroposons in the octopus. J Mol Biol 243(1):25-37.

Olazarán, J., Reisberg, B., Clare, L., Cruz, I., Peña-Casanova, J., Del Ser, T., Woods, B., Beck, C., Auer, S., Lai, C., Spector, A., Fazio, S., Bond, J., Kivipelto, M., Brodaty, H., Rojo, J.M., Collins, H., Teri, L., Mittelman, M., Orrell, M., Feldman, H.H., Muñiz, R. (2010). Nonpharmacological therapies in Alzheimer's disease: a systematic review of efficacy. Dement Geriatr Cogn Disord 30(2):161-78.

Ostertag, E.M., DeBerardinis, R.J., Goodier, J.L., Zhang, Y., Yang, N., Gerton, G.L., Kazazian, H.H. Jr (2002). A mouse model of human L1 retrotransposition. Nat Genet 32(4):655-60.

Ostertag, E.M., Kazazian, H.H. Jr (2001). Biology of mammalian L1 retrotransposons. Annu Rev Genet 35:501-38.

Pace, J.K., Feschotte, C. (2007). The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res 17(4):422-32.

Panegyres, P.K., Chen, H.Y. (2013). Differences between early and late onset Alzheimer's disease. Am J Neurodegener Dis 2(4):300-6.

Peleg, S., Sananbenesi, F., Zovoilis, A., Burkhardt, S., Bahari-Javan, S., Agis-Balboa, R.C., Cota, P., Wittnam, J.L., Gogol-Doering, A., Opitz, L., Salinas-Riester, G., Dettenhofer, M., Kang, H., Farinelli, L., Chen, W., Fischer, A. (2010). Altered histone

acetylation is associated with age-dependent memory impairment in mice. Science 328(5979):753-6.

Peña-Longobardo, L.M., Oliva-Moreno, J. (2014). Caregiver Burden in Alzheimer's Disease Patients in Spain. J Alzheimers Dis [Epub ahead of print].

Penzkofer, T., Dandekar, T., Zemojtel, T. (2005). L1Base: from functional annotation to prediction of active LINE-1 elements. Nucleic Acids Res 33(Database issue):D498-500.

Perepelitsa-Belancio, V., Deininger, P. (2003). RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet 35(4):363-6.

Perl, D.P. (2010). Neuropathology of Alzheimer's disease. Mt Sinai J Med 77(1):32-42.

Pickeral, O.K., Makałowski, W., Boguski, M.S., Boeke, J.D. (2000). Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res 10(4):411-5.

Piskareva, O., Schmatchenko, V. (2006). DNA polymerization by the reverse transcriptase of the human L1 retrotransposon on its own template in vitro. FEBS Lett 580(2):661-8.

Pittoggi, C., Sciamanna, I., Mattei, E., Beraldi, R., Lobascio, A.M., Mai, A., Quaglia, M.G., Lorenzini, R., Spadafora, C. (2003). Role of endogenous reverse transcriptase in murine early embryo development. Mol Reprod Dev 66(3):225-36.

Ponicsan, S.L., Kugel, J.F., Goodrich, J.A. (2010). Genomic gems: SINE RNAs regulate mRNA production. Curr Opin Genet Dev 20(2):149-55.

Pornthanakasem, W., Mutirangura, A. (2004). LINE-1 insertion dimorphisms identification by PCR. Biotechniques 37(5):750, 752.

Radde, R., Duma, C., Goedert, M., Jucker, M. (2008). The value of incomplete mouse models of Alzheimer's disease. Eur J Nucl Med Mol Imaging 35 Suppl 1:S70-4.

Rademakers, R., Cruts, M., Sleegers, K., Dermaut, B., Theuns, J., Aulchenko, Y., Weckx, S., De Pooter, T., Van den Broeck, M., Corsmit, E., De Rijk, P., Del-Favero, J., van Swieten, J., van Duijn, C.M., Van Broeckhoven, C. (2005). Linkage and association studies identify a novel locus for Alzheimer disease at 7q36 in a Dutch population-based sample. Am J Hum Genet 77(4):643-52.

Rademakers, R., Dermaut, B., Peeters, K., Cruts, M., Heutink, P., Goate, A., Van Broeckhoven, C. (2003). Tau (MAPT) mutation Arg406Trp presenting clinically with Alzheimer disease does not share a common founder in Western Europe. Hum Mutat 22(5):409-11.

Raux, G., Guyant-Maréchal, L., Martin, C., Bou, J., Penet, C., Brice, A., Hannequin, D., Frebourg, T., Campion, D. (2005). Molecular diagnosis of autosomal dominant early onset Alzheimer's disease: an update. J Med Genet 42(10):793-5.

Reilly, M.T., Faulkner, G.J., Dubnau, J., Ponomarev, I., Gage, F.H. (2013). The role of transposable elements in health and diseases of the central nervous system. J Neurosci 33(45):17577-86.

Reitz, C., Mayeux, R. (2014). Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. Biochem Pharmacol 88(4):640-51.

Ren, G., Vajjhala, P., Lee, J.S., Winsor, B., Munn, A.L. (2006). The BAR domain proteins: molding membranes in fission, fusion, and phagy. Microbiol Mol Biol Rev 70(1):37-120.

Reuter, M., Berninger, P., Chuma, S., Shah, H., Hosokawa, M., Funaya, C., Antony, C., Sachidanandam, R., Pillai, R.S. (2011). Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. Nature 480(7376):264-7.

Richardson, S.R., Morell, S., Faulkner, G.J. (2014). L1 retrotransposons and somatic mosaicism in the brain. Annu Rev Genet 48:1-27.

Ridge, P.G., Ebbert, M.T., Kauwe, J.S. (2013). Genetics of Alzheimer's disease. Biomed Res Int 2013:254954.

Rizzi, F., Caccamo, A.E., Belloni, L., Bettuzzi, S. (2009). Clusterin is a short half-life, poly-ubiquitinated protein, which controls the fate of prostate cancer cells. J Cell Physiol 219(2):314-23

Rogaeva, E., Meng, Y., Lee, J.H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C.T., Cheng, R., Hasegawa, H., Chen, F., Shibata, N., Lunetta, K.L., Pardossi-Piquard, R., Bohm, C., Wakutani, Y., Cupples, L.A., Cuenco, K.T., Green, R.C., Pinessi, L., Rainero, I., Sorbi, S., Bruni, A., Duara, R., Friedland, R.P., Inzelberg, R., Hampe, W., Bujo, H., Song, Y.Q., Andersen, O.M., Willnow, T.E., Graff-Radford, N., Petersen, R.C., Dickson, D., Der, S.D., Fraser, P.E., Schmitt-Ulms, G., Younkin, S., Mayeux, R., Farrer, L.A., St George-Hyslop, P. (2007). The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat Genet 39(2):168-77.

Rogers, J., Li, R., Mastroeni, D., Grover, A., Leonard, B., Ahern, G., Cao, P., Kolody, H., Vedders, L., Kolb, W.P., Sabbagh, M. (2006). Peripheral clearance of amyloid beta peptide by complement C3-dependent adherence to erythrocytes. Neurobiol Aging 27(12):1733-9.

Ryan, N.S., Rossor, M.N. (2010). Correlating familial Alzheimer's disease gene mutations with clinical phenotype. Biomark Med 4(1):99-112.

Sala Frigerio, C., Piscopo, P., Calabrese, E., Crestini, A., Malvezzi Campeggi, L., Civita di Fava, R., Fogliarino, S., Albani, D., Marcon, G., Cherchi, R., Piras, R., Forloni, G., Confaloni, A. (2005). PEN-2 gene mutation in a familial Alzheimer's disease case. J Neurol 252(9):1033-6.

Samuelov, L., Fuchs-Telem, D., Sarig, O., Sprecher, E. (2011). An exceptional mutational event leading to Chanarin-Dorfman syndrome in a large consanguineous family. Br J Dermatol 164(6):1390-2.

Scarpa, S., Cavallaro, R.A., D'Anselmi, F., Fuso, A. (2006). Gene silencing through methylation: an epigenetic intervention on Alzheimer disease. J Alzheimers Dis 9(4):407-14.

Scarpa, S., Fuso, A., D'Anselmi, F., Cavallaro, R.A. (2003). Presenilin 1 gene silencing by S-adenosylmethionine: a treatment for Alzheimer disease? FEBS Lett 541(1-3):145-8.

Schrijvers, E.M., Koudstaal, P.J., Hofman, A., Breteler, M.M. (2011). Plasma clusterin and the risk of Alzheimer disease. JAMA 305(13):1322-6.

Schwahn, U., Lenzner, S., Dong, J., Feil, S., Hinzmann, B., van Duijnhoven, G., Kirschner, R., Hemberger, M., Bergen, A.A., Rosenberg, T., Pinckers, A.J., Fundele, R., Rosenthal, A., Cremers, F.P., Ropers, H.H., Berger, W. (1998). Positional cloning of the gene for X-linked retinitis pigmentosa 2. Nat Genet 19(4):327-32.

Sen, S.K., Huang, C.T., Han, K., Batzer, M.A. (2007). Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. Nucleic Acids Res 35(11):3741-51.

Serrano-Pozo, A., Frosch, M.P., Masliah, E., Hyman, B.T. (2011). Neuropathological alterations in Alzheimer disease. Cold Spring Harb Perspect Med 1(1):a006189.

Shah, P., Lal, N., Leung, E., Traul, D.E., Gonzalo-Ruiz, A., Geula, C. (2010). Neuronal and axonal loss are selectively linked to fibrillar amyloid-{beta} within plaques of the aged primate cerebral cortex. Am J Pathol 177(1):325-33.

Sharif, J., Shinkai, Y., Koseki, H. (2013). Is there a role for endogenous retroviruses to mediate long-term adaptive phenotypic response upon environmental inputs? Philos Trans R Soc Lond B Biol Sci 368(1609):20110340.

Sheen, F.M., Sherry, S.T., Risch, G.M., Robichaux, M., Nasidze, I., Stoneking, M., Batzer, M.A., Swergold, G.D. (2000). Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. Genome Res 10(10):1496-508.

Shukla, R., Upton, K.R., Muñoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S., Sinha, S., Iannelli, F., Radaelli, E., Dos Santos, A., Rapoud, D., Guettier, C., Samuel, D., Natoli, G., Carninci, P., Ciccarelli, F.D., Garcia-Perez, J.L., Faivre, J., Faulkner, G.J. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell 153(1):101-11.

Singer, T., McConnell, M.J., Marchetto, M.C., Coufal, N.G., Gage, F.H. (2010). LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? Trends Neurosci. 33(8):345-54.

Siomi, M.C., Sato, K., Pezic, D., Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol 12(4):246-58.

Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Andrews, R., Bird, A.P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol Cell 37(4):457-68.

Sleegers, K., Brouwers, N., Gijselinck, I., Theuns, J., Goossens, D., Wauters, J., Del-Favero, J., Cruts, M., van Duijn, C.M., Van Broeckhoven, C. (2006). APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. Brain 129(Pt 11):2977-83.

Smallwood, S.A., Kelsey, G. (2011). De novo DNA methylation: a germ cell perspective. Trends Genet 28(1):33-42.

Solyom, S., Ewing, A.D., Hancks, D.C., Takeshima, Y., Awano, H., Matsuo, M., Kazazian, H.H. Jr (2012). Pathogenic orphan transduction created by a nonreference LINE-1 retrotransposon. Hum Mutat 33(2):369-71.

Stewart, C., Kural, D., Strömberg, M.P., Walker, J.A., Konkel, M.K., Stütz, A.M., Urban, A.E., Grubert, F., Lam, H.Y., Lee, W.P., Busby, M., Indap, A.R., Garrison, E., Huff, C., Xing, J., Snyder, M.P., Jorde, L.B., Batzer, M.A., Korbel, J.O., Marth, G.T.;

1000 Genomes Project (2011). A comprehensive map of mobile element insertion polymorphisms in humans. PLoS Genet 7(8):e1002236.

St Laurent, G., Hammell, N., McCaffrey, T.A. (2010). A LINE-1 component to human aging: do LINE elements exact a longevity cost for evolutionary advantage? Mech Ageing Dev 131(5):299-305.

Szak, S.T., Pickeral, O.K., Makalowski, W., Boguski, M.S., Landsman, D., Boeke, J.D. (2002). Molecular archeology of L1 insertions in the human genome. Genome Biol 3(10):research0052.

Tanzi, R.E. (2012). The genetics of Alzheimer disease. Cold Spring Harb Perspect Med 2(10). pii: a006296.

Thies, W., Bleiler, L.; Alzheimer's Association (2013). 2013 Alzheimer's disease facts and figures. Alzheimers Dement 9(2):208-45.

Thomas, C.A., Muotri, A.R. (2012). LINE-1: creators of neuronal diversity. Front Biosci 4:1663-8.

Thomas, C.A., Paquola, A.C., Muotri, A.R. (2012). LINE-1 retrotransposition in the nervous system. Annu Rev Cell Dev Biol 28:555-73.

Tohgi, H., Utsugisawa, K., Nagane, Y., Yoshimura, M., Genda, Y., Ukitsu, M. (1999). Reduction with age in methylcytosine in the promoter region -224 approximately -101 of the amyloid precursor protein gene in autopsy human cortex. Brain Res Mol Brain Res 70(2):288-92.

Trelogan, S.A., Martin, S.L. (1995). Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. Proc Natl Acad Sci USA 92(5):1520-4.

Tsankova, N., Renthal, W., Kumar, A., Nestler, E.J. (2007). Epigenetic regulation in psychiatric disorders. Nat Rev Neurosci 8(5):355-67.

Uren, A.G., Mikkers, H., Kool, J., van der Weyden, L., Lund, A.H., Wilson, C.H., Rance, R., Jonkers, J., van Lohuizen, M., Berns, A., Adams, D.J. (2009). A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. Nat Protoc 4(5):789-98.

van den Hurk, J.A., van de Pol, D.J., Wissinger, B., van Driel, M.A., Hoefsloot, L.H., de Wijs, I.J., van den Born, L.I., Heckenlively, J.R., Brunner, H.G., Zrenner, E., Ropers, H.H., Cremers, F.P. (2003). Novel types of mutation in the choroideremia (CHM) gene: a full-length L1 insertion and an intronic mutation activating a cryptic exon. Hum Genet 113(3):268-75.

Vasquez, J.B., Fardo, D.W., Estus, S. (2013). ABCA7 expression is associated with Alzheimer's disease polymorphism and disease status. Neurosci Lett 556:58-62.

Vassetzky, N.S., Ten, O.A., Kramerov, D.A. (2003). B1 and related SINEs in mammalian genomes. Gene 13;319:149-60.

Verghese, P.B., Castellano, J.M., Garai, K., Wang, Y., Jiang, H., Shah, A., Bu, G., Frieden, C., Holtzman, D.M. (2013). ApoE influences amyloid-β (Aβ) clearance despite minimal apoE/Aβ association in physiological conditions. Proc Natl Acad Sci USA 110(19):E1807-16.

Viollet, S., Monot, C., Cristofari, G. (2014). L1 retrotransposition: The snap-velcro model and its consequences. Mob Genet Elements 4(1):e28907.

Vitullo, P., Sciamanna, I., Baiocchi, M., Sinibaldi-Vallebona, P., Spadafora, C. (2012). LINE-1 retrotransposon copies are amplified during murine early embryo development. Mol Reprod Dev 79(2):118-27.

Wallace, N.A., Belancio, V.P., Deininger, P.L. (2008). L1 mobile element expression causes multiple types of toxicity. Gene 419(1-2):75-81.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S.E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M.R., Brown, D.G., Brown, S.D., Bult, C., Burton, J., Butler, J., Campbell, R.D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A.T., Church, D.M., Clamp, M., Clee, C., Collins, F.S., Cook, L.L., Copley, R.R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K.D., Deri, J., Dermitzakis, E.T., Dewey, C., Dickens, N.J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D.M., Eddy, S.R., Elnitski, L., Emes, R.D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G.A., Flicek, P., Foley, K., Frankel, W.N., Fulton, L.A., Fulton, R.S., Furey, T.S., Gage, D., Gibbs, R.A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T.A., Green, E.D., Gregory, S., Guigó, R., Guyer, M., Hardison, R.C., Haussler, D., Hayashizaki, Y., Hillier, L.W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D.B., Johnson, L.S., Jones, M., Jones, T.A., Joy, A., Kamal, M., Karlsson, E.K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W.J., Kirby, A., Kolbe, D.L., Korf, I., Kucherlapati, R.S., Kulbokas, E.J., Kulp, D., Landers, T., Leger, J.P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D.R., Mardis, E.R., Matthews, L., Mauceli, E., Mayer, J.H., McCarthy, M., McCombie, W.R., McLaren, S., McLay, K., McPherson, J.D., Meldrim, J., Meredith, B., Mesirov, J.P., Miller, W., Miner, T.L., Mongin, E., Montgomery, K.T., Morgan, M., Mott, R., Mullikin, J.C., Muzny, D.M., Nash, W.E., Nelson, J.O., Nhan, M.N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M.J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K.H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C.S., Poliakov, A., Ponce, T.C., Ponting, C.P., Potter, S., Quail, M., Reymond, A., Roe, B.A., Roskin, K.M., Rubin, E.M., Rust, A.G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M.S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J.B., Slater, G., Smit, A., Smith, D.R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J.P., Von Niederhausern, A.C., Wade, C.M., Wall, M., Weber, R.J., Weiss, R.B., Wendl, M.C., West, A.P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R.K., Winter, E., Worley, K.C., Wyman, D., Yang, S.,

Yang, S.P., Zdobnov, E.M., Zody, M.C., Lander, E.S. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature 420(6915):520-62.

West, R.L., Lee, J.M., Maroun, L.E. (1995). Hypomethylation of the amyloid precursor protein gene in the brain of an Alzheimer's disease patient. J Mol Neurosci 6(2):141-6.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H. (2007). A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8(12):973-82.

Wimmer, K., Callens, T., Wernstedt, A., Messiaen, L. (2011). The NF1 gene contains hotspots for L1 endonuclease-dependent de novo insertion. PLoS Genet 7(11):e1002371.

Witherspoon, D.J., Xing, J., Zhang, Y., Watkins, W.S., Batzer, M.A., Jorde, L.B. (2010). Mobile element scanning (ME-Scan) by targeted high-throughput sequencing. BMC Genomics. 11:410.

Xing, J., Zhang, Y., Han, K., Salem, A.H., Sen, S.K., Huff, C.D., Zhou, Q., Kirkness, E.F., Levy, S., Batzer, M.A., Jorde, L.B. (2009). Mobile elements create structural variation: analysis of a complete human genome. Genome Res 19(9):1516-26.

Yin, B., Largaespada, D.A. (2007). PCR-based procedures to isolate insertion sites of DNA elements. Biotechniques 43(1):79-84.

Yoder, J.A., Walsh, C.P., Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13(8):335-40.

Yoshida, K., Nakamura, A., Yazaki, M., Ikeda, S., Takeda, S. (1998). Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. Hum Mol Genet 7(7):1129-32.

Yu, F., Zingler, N., Schumann, G., Strätling, W.H. (2001). Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. Nucleic Acids Res 29(21):4493-501.

Zemojtel, T., Penzkofer, T., Schultz, J., Dandekar, T., Badge, R., Vingron, M. (2007). Exonization of active mouse L1s: a driver of transcriptome evolution? BMC Genomics 8:392.

Appendix A: Human Gene Ontology Tables

Table A.1: Gene ontology of NIS associated genes, found common between AD samples (frontal cortex and kidney). Background frequency corresponds to the total number of genes associated to the term in the human genome; sample frequency corresponds to the AIS associated genes found for the term. p<0.05

Biological function	Background frequency	Sample frequency	P-value
regulation of focal adhesion assembly (GO:0051893)	37	2	7.21E-04
regulation of cell-substrate junction assembly (GO:0090109)	37	2	7.21E-04
regulation of adherens junction organization (GO:1903391)	38	2	7.60E-04
regulation of cell junction assembly (GO:1901888)	50	2	1.31E-03
regulation of cell-matrix adhesion (GO:0001952)	78	2	3.18E-03
negative regulation of ripoptosome assembly involved in necroptotic process (GO:1902443)	1	1	3.30E-03
regulation of ripoptosome assembly involved in necroptotic process (GO:1902442)	1	1	3.30E-03
regulation of cellular component biogenesis (GO:0044087)	520	3	6.25E-03
fasciculation of motor neuron axon (GO:0097156)	2	1	6.60E-03
fasciculation of sensory neuron axon (GO:0097155)	3	1	9.90E-03
regulation of cell-substrate adhesion (GO:0010810)	149	2	1.15E-02
peptidyl-amino acid modification (GO:0018193)	645	3	1.17E-02
protein phosphorylation (GO:0006468)	667	3	1.29E-02
apical constriction (GO:0003383)	4	1	1.32E-02
membrane to membrane docking (GO:0022614)	5	1	1.65E-02
regulation of neuron apoptotic process (GO:0043523)	185	2	1.75E-02
regulation of cellular component organization (GO:0051128)	1709	4	1.84E-02
cell adhesion (GO:0007155)	823	3	2.35E-02
biological adhesion (GO:0022610)	826	3	2.37E-02
regulation of neuron death (GO:1901214)	222	2	2.51E-02
regulation of actin cytoskeleton organization (GO:0032956)	231	2	2.71E-02
nervous system development (GO:0007399)	1922	4	2.84E-02
myoblast migration (GO:0051451)	9	1	2.97E-02
bleb assembly (GO:0032060)	9	1	2.97E-02
positive regulation of T cell cytokine production (GO:0002726)	9	1	2.97E-02
negative regulation of necroptotic process (GO:0060546)	9	1	2.97E-02
neuron differentiation (GO:0030182)	914	3	3.17E-02
cerebellar granule cell differentiation (GO:0021707)	10	1	3.30E-02
cerebellar granular layer formation (GO:0021684)	10	1	3.30E-02
prepulse inhibition (GO:0060134)	11	1	3.63E-02
negative regulation of necrotic cell death (GO:0060547)	11	1	3.63E-02
regulation of actin filament-based process (GO:0032970)	276	2	3.84E-02
regulation of neuron projection development (GO:0010975)	278	2	3.89E-02
cerebellar granular layer morphogenesis (GO:0021683)	12	1	3.96E-02
I-kappaB phosphorylation (GO:0007252)	12	1	3.96E-02
regulation of necroptotic process (GO:0060544)	12	1	3.96E-02
regulation of multicellular organismal process (GO:0051239)	2121	4	4.09E-02
phosphorylation (GO:0016310)	1017	3	4.28E-02
cerebellar granular layer development (GO:0021681)	15	1	4.94E-02
regulation of establishment of cell polarity (GO:2000114)	15	1	4.94E-02
leukocyte tethering or rolling (GO:0050901)	15	1	4.94E-02
regulation of T cell cytokine production (GO:0002724)	15	1	4.94E-02

Table A.2: Gene ontology of NIS associated genes, found common between healthy controls (frontal cortex and kidney). Background frequency corresponds to the total number of genes associated to the term in the human genome; sample frequency corresponds to the AIS associated genes found for the term. p<0.05.

Biological function	Background frequency	Sample frequency	P-value
cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742)	185	8	1.87E-06
cell-cell adhesion (GO:0098609)	186	8	1.95E-06
homophilic cell adhesion via plasma membrane adhesion molecules (GO:0007156)	137	7	4.74E-06
cell adhesion (GO:0007155)	823	12	7.38E-05
biological adhesion (GO:0022610)	826	12	7.66E-05
neuron differentiation (GO:0030182)	914	12	2.17E-04
central nervous system neuron differentiation (GO:0021953)	170	6	3.54E-04
generation of neurons (GO:0048699)	1237	13	8.85E-04
neuron development (GO:0048666)	743	10	1.28E-03
nervous system development (GO:0007399)	1922	16	1.38E-03
brain development (GO:0007420)	603	9	1.56E-03
neurogenesis (GO:0022008)	1312	13	1.64E-03
regulation of locomotion (GO:0040012)	616	9	1.84E-03
central nervous system development (GO:0007417)	801	10	2.42E-03
head development (GO:0060322)	639	9	2.45E-03
regulation of negative chemotaxis (GO:0050923)	4	2	3.26E-03
neuron recognition (GO:0008038)	32	3	5.32E-03
regulation of localization (GO:0032879)	1936	15	6.07E-03
regulation of nervous system development (GO:0051960)	575	8	7.44E-03
regulation of synapse assembly (GO:0051963)	39	3	9.50E-03
anatomical structure morphogenesis (GO:0009653)	2020	15	9.75E-03
regulation of cellular component movement (GO:0051270)	630	8	1.38E-02
axon development (GO:0061564)	472	7	1.39E-02
system development (GO:0048731)	3692	21	1.54E-02
trigeminal nerve development (GO:0021559)	9	2	1.63E-02
negative regulation of transcription by competitive promoter binding (GO:0010944)	10	2	2.01E-02
axon guidance (GO:0007411)	362	6	2.25E-02
neuron projection guidance (GO:0097485)	362	6	2.25E-02
forebrain neuron differentiation (GO:0021879)	53	3	2.32E-02
single organismal cell-cell adhesion (GO:0016337)	235	5	2.33E-02
regulation of cellular component biogenesis (GO:0044087)	520	7	2.49E-02
regulation of synapse organization (GO:0050807)	61	3	3.49E-02
single organism cell adhesion (GO:0098602)	258	5	3.55E-02
regulation of multicellular organismal development (GO:2000026)	1331	11	3.59E-02
multicellular organismal development (GO:0007275)	4225	22	3.66E-02
protein localization to synapse (GO:0035418)	14	2	3.92E-02
regulation of cell motility (GO:2000145)	563	7	3.98E-02
regulation of synapse structure and activity (GO:0050803)	64	3	4.01E-02
forebrain generation of neurons (GO:0021872)	64	3	4.01E-02
regulation of membrane potential (GO:0042391)	266	5	4.07E-02
olfactory bulb interneuron differentiation (GO:0021889)	15	2	4.49E-02
positive regulation of multicellular organismal process (GO:0051240)	1151	10	4.50E-02

Appendix B: Drosophila Gene Ontology Tables

Table B.1: Gene ontology of NIS associated genes (defined by at least 1 MapFragment). Background frequency corresponds to the total number of genes associated to the biological function in Drosophila; sample frequency corresponds to the NIS associated genes found for the term. p<0.05.

Biological function	Background frequency	Sample frequency	P-value
anatomical structure morphogenesis (GO:0009653)	1634	503	3.08E-45
biological regulation (GO:0065007)	3307	829	6.00E-45
regulation of cellular process (GO:0050794)	2829	733	7.58E-43
regulation of biological process (GO:0050789)	3047	772	2.48E-42
organ development (GO:0048513)	1220	400	1.77E-40
system development (GO:0048731)	2172	597	4.54E-40
single-multicellular organism process (GO:0044707)	3217	792	3.68E-39
organ morphogenesis (GO:0009887)	792	292	7.90E-37
anatomical structure development (GO:0048856)	2870	712	8.23E-35
single-organism developmental process (GO:0044767)	3155	764	1.06E-34
multicellular organismal process (GO:0032501)	3641	849	4.26E-34
developmental process (GO:0032502)	3178	764	1.28E-33
multicellular organismal development (GO:0007275)	2696	675	1.36E-33
generation of neurons (GO:0048699)	845	296	1.82E-33
locomotion (GO:0040011)	538	220	5.59E-33
response to stimulus (GO:0050896)	2415	615	9.94E-32
cell differentiation (GO:0030154)	2142	560	4.95E-31
tissue development (GO:0009888)	756	268	9.95E-31
cellular developmental process (GO:0048869)	2202	570	1.36E-30
neuron differentiation (GO:0030182)	732	261	3.25E-30
single-organism process (GO:0044699)	6262	1267	1.88E-29
cell development (GO:0048468)	1345	394	2.42E-29
post-embryonic organ development (GO:0048569)	477	196	2.61E-29
post-embryonic organ morphogenesis (GO:0048563)	449	187	1.55E-28
imaginal disc morphogenesis (GO:0007560)	449	187	1.55E-28
epithelium development (GO:0060429)	545	210	6.25E-28
imaginal disc development (GO:0007444)	624	229	8.60E-28
metamorphosis (GO:0007552)	560	212	3.15E-27
nervous system development (GO:0007399)	1504	419	5.41E-27
post-embryonic development (GO:0009791)	663	236	6.91E-27
post-embryonic morphogenesis (GO:0009886)	546	207	1.35E-26
appendage morphogenesis (GO:0035107)	394	168	1.76E-26
appendage development (GO:0048736)	398	169	1.87E-26
imaginal disc-derived appendage morphogenesis (GO:0035114)	392	167	2.88E-26
imaginal disc-derived appendage development (GO:0048737)	396	168	3.06E-26
instar larval or pupal development (GO:0002165)	630	226	4.63E-26
instar larval or pupal morphogenesis (GO:0048707)	536	203	5.70E-26
post-embryonic appendage morphogenesis (GO:0035120)	385	164	9.56E-26
single-organism cellular process (GO:0044763)	4829	1015	2.27E-25
neuron development (GO:0048666)	630	223	6.45E-25
cellular component movement (GO:0006928)	575	209	1.40E-24
cell communication (GO:0007154)	1384	386	1.53E-24
imaginal disc-derived wing morphogenesis (GO:0007476)	339	149	1.68E-24
wing disc morphogenesis (GO:0007472)	345	150	3.23E-24
tissue morphogenesis (GO:0048729)	313	141	4.58E-24
single organism signaling (GO:0044700)	1329	373	4.77E-24
signaling (GO:0023052)	1329	373	4.77E-24
cellular component morphogenesis (GO:0032989)	771	251	2.79E-23
pattern specification process (GO:0007389)	508	189	4.15E-23
response to chemical (GO:0042221)	836	265	4.22E-23
cell morphogenesis involved in differentiation (GO:0000904)	464	178	4.76E-23
morphogenesis of an epithelium (GO:0002009)	289	132	6.38E-23
neurogenesis (GO:0022008)	1355	373	1.53E-22
wing disc development (GO:0035220)	459	175	2.48E-22
cell projection organization (GO:0030030)	631	216	3.08E-22
regulation of developmental process (GO:0050793)	635	216	6.66E-22
taxis (GO:0042330)	312	136	7.52E-22
cell morphogenesis involved in neuron differentiation (GO:0048667)	423	164	1.65E-21
embryo development (GO:0009790)	502	181	1.46E-20
axon development (GO:0061564)	291	127	2.61E-20
cell morphogenesis (GO:0000902)	657	216	3.99E-20
neuron projection development (GO:0031175)	523	184	8.38E-20
regionalization (GO:0003002)	462	169	1.16E-19
behavior (GO:0007610)	523	183	1.96E-19
cellular component organization (GO:0016043)	2257	535	2.14E-19
axonogenesis (GO:0007409)	284	123	2.43E-19
regulation of multicellular organismal process (GO:0051239)	584	196	5.48E-19
regulation of metabolic process (GO:0019222)	1572	401	1.48E-18

Table B.2: Gene ontology of NIS associated genes (defined by at least 10 MapFragments). Background frequency corresponds to the total number of genes associated to the biological function in Drosophila; sample frequency corresponds to the NIS associated genes found for the term. p<0.05.

Biological function	Background frequency	Sample frequency	P-value
response to stimulus (GO:0050896)	2415	45	1.31E-05
single-multicellular organism process (GO:0044707)	3217	51	2.11E-04
cell communication (GO:0007154)	1384	28	1.39E-03
regulation of biological process (GO:0050789)	3047	47	1.47E-03
system process (GO:0003008)	548	16	1.81E-03
single-organism process (GO:0044699)	6262	77	2.19E-03
regulation of cellular process (GO:0050794)	2829	44	2.62E-03
axonogenesis (GO:0007409)	284	11	3.19E-03
axon development (GO:0061564)	291	11	3.96E-03
regulation of response to DNA damage stimulus (GO:2001020)	24	4	4.81E-03
neuron differentiation (GO:0030182)	732	18	5.10E-03
single organism signaling (GO:0044700)	1329	26	5.21E-03
signaling (GO:0023052)	1329	26	5.21E-03
multicellular organismal process (GO:0032501)	3641	51	7.78E-03
cell projection organization (GO:0030030)	631	16	9.47E-03
neuron projection extension (GO:1990138)	29	4	9.92E-03
axon extension (GO:0048675)	29	4	9.92E-03
generation of neurons (GO:0048699)	845	19	1.02E-02
biological regulation (GO:0065007)	3307	47	1.25E-02
organ morphogenesis (GO:0009887)	792	18	1.36E-02
cell morphogenesis involved in differentiation (GO:0000904)	464	13	1.65E-02
tissue development (GO:0009888)	756	17	2.37E-02
cell morphogenesis involved in neuron differentiation (GO:0048667)	423	12	2.60E-02
signal transduction (GO:0007165)	916	19	2.81E-02
neuron development (GO:0048666)	630	15	3.04E-02
cellular response to stimulus (GO:0051716)	1318	24	3.05E-02
neurological system process (GO:0050877)	501	13	3.41E-02
synapse organization (GO:0050808)	114	6	3.94E-02
neuron projection morphogenesis (GO:0048812)	512	13	4.17E-02
sensory perception of chemical stimulus (GO:0007606)	268	9	4.52E-02

Table B.3: Gene ontology of AIS associated genes (defined by at least 1 MapFragment). Background frequency corresponds to the total number of genes associated to the biological function in Drosophila; sample frequency corresponds to the NIS associated genes found for the term. p<0.05.

Biological function	Background frequency	Sample frequency	P-value
regulation of response to stimulus (GO:0048583)	855	16	1.25E-03
neuron projection extension (GO:1990138)	29	4	1.37E-03
axon extension (GO:0048675)	29	4	1.37E-03
regulation of signaling (GO:0023051)	701	13	8.86E-03
regulation of cell communication (GO:0010646)	710	13	1.00E-02
developmental cell growth (GO:0048588)	50	4	1.10E-02
regulation of signal transduction (GO:0009966)	629	12	1.25E-02
anatomical structure morphogenesis (GO:0009653)	1634	21	1.38E-02
regulation of cellular process (GO:0050794)	2829	30	1.40E-02
synapse assembly (GO:0007416)	57	4	1.81E-02
imaginal disc-derived appendage morphogenesis (GO:0035114)	392	9	2.18E-02
regulation of biological process (GO:0050789)	3047	31	2.23E-02
appendage morphogenesis (GO:0035107)	394	9	2.26E-02
imaginal disc-derived appendage development (GO:0048737)	396	9	2.34E-02
appendage development (GO:0048736)	398	9	2.43E-02
single-multicellular organism process (GO:0044707)	3217	32	2.58E-02
developmental growth involved in morphogenesis (GO:0060560)	63	4	2.63E-02

Table B.4: Gene ontology of AIS associated genes (defined by at least 10 MapFragments). Background frequency corresponds to the total number of genes associated to the biological function in Drosophila; sample frequency corresponds to the NIS associated genes found for the term. p<0.05.

Biological function	Background frequency	Sample frequency	P-value
sperm individualization (GO:0007291)	47	3	3.61E-04
spermatid development (GO:0007286)	99	3	3.20E-03
cellularization (GO:0007349)	99	3	3.20E-03
positive regulation of JAK-STAT cascade (GO:0046427)	24	2	4.20E-03
spermatid differentiation (GO:0048515)	111	3	4.45E-03
histone H3-K4 dimethylation (GO:0044648)	1	1	9.68E-03
regulation of JAK-STAT cascade (GO:0046425)	44	2	1.39E-02
peptidyl-lysine dimethylation (GO:0018027)	2	1	1.94E-02
histone H3-K4 trimethylation (GO:0080182)	2	1	1.94E-02
spermatogenesis (GO:0007283)	204	3	2.51E-02
male gamete generation (GO:0048232)	205	3	2.54E-02
positive regulation of nuclear-transcribed mRNA poly(A) tail shortening (GO:0060213)	4	1	3.86E-02
regulation of nuclear-transcribed mRNA poly(A) tail shortening (GO:0060211)	4	1	3.86E-02
positive regulation of nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay (GO:1900153)	4	1	3.86E-02
regulation of nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay (GO:1900151)	4	1	3.86E-02
peptidyl-lysine trimethylation (GO:0018023)	4	1	3.86E-02
positive regulation of mRNA processing (GO:0050685)	4	1	3.86E-02
positive regulation of mRNA 3'-end processing (GO:0031442)	4	1	3.86E-02
positive regulation of mRNA metabolic process (GO:1903313)	5	1	4.83E-02
response to auditory stimulus (GO:0010996)	5	1	4.83E-02
peptidyl-lysine methylation (GO:0018022)	5	1	4.83E-02
positive regulation of mRNA catabolic process (GO:0061014)	5	1	4.83E-02
regulation of mRNA catabolic process (GO:0061013)	5	1	4.83E-02

Table B.5: Gene ontology of genes associated to the *roo*-LTR annotated in our database that we didn't detect with the SPAM technique. Background frequency corresponds to the total number of genes associated to the term in Drosophila; sample frequency corresponds to the AIS associated genes found for the term. p<0.05.

Biological function	Background frequency	Sample frequency	P-value
heart contraction (GO:0060047)	6	2	4.72E-03
blood circulation (GO:0008015)	6	2	4.72E-03
single-organism process (GO:0044699)	6262	35	5.75E-03
anatomical structure development (GO:0048856)	2870	21	1.01E-02
leg disc development (GO:0035218)	101	4	1.03E-02
single-multicellular organism process (GO:0044707)	3217	22	1.80E-02
regionalization (GO:0003002)	462	7	2.66E-02
cell morphogenesis involved in differentiation (GO:0000904)	464	7	2.73E-02
axon target recognition (GO:0007412)	15	2	2.89E-02
single-organism cellular process (GO:0044763)	4829	28	3.29E-02
heart process (GO:0003015)	17	2	3.70E-02
circulatory system process (GO:0003013)	17	2	3.70E-02
single-organism developmental process (GO:0044767)	3155	21	3.70E-02
multicellular organismal process (GO:0032501)	3641	23	4.01E-02
developmental process (GO:0032502)	3178	21	4.08E-02
cellular response to steroid hormone stimulus (GO:0071383)	18	2	4.14E-02
response to steroid hormone (GO:0048545)	18	2	4.14E-02
steroid hormone mediated signaling pathway (GO:0043401)	18	2	4.14E-02
pattern specification process (GO:0007389)	508	7	4.56E-02
regulation of RNA biosynthetic process (GO:2001141)	813	9	4.68E-02
regulation of transcription, DNA-templated (GO:0006355)	813	9	4.68E-02
regulation of nucleic acid-templated transcription (GO:1903506)	813	9	4.68E-02

This page intentionally left blank