More than half of the human genome is composed of repeated DNA, often referred as to *junk* DNA or *selfish* DNA. Indeed, human genes (DNA coding regions which represents <5% of the genome) can be considered islands in an ocean of repeated DNA. Some classes of repeated DNA can be mobilized within the genome (mobile DNA or transposons), generating new copies dispersed through the genome. Thus, the presence of mobile DNA within the human genome implies that our genome is not static and that new pieces of DNA are continuously being accumulated, dispersed and shuffled in our genome. Long INterspersed element 1 (LINE-1 or L1) is the only non-LTR autonomous mobile DNA in the human genome, with ~600000 copies accumulated in ALL human chromosomes. Although most of the LINE-1 copies are non-active elements or fossils (~99.8% of them), it was recently shown that an average human genome contains 80-100 active LINE-1 (RetroCompetetent L1s (RC-L1)).

Active LINE-1 elements in the human genome are 6Kb in length, lack Long Terminal Repeats (presents in retroviruses), contains a ~900bp long 5'UTR with internal promoter activity, followed by two Open Reading Frames (ORFs), and end in a poly-A tail. ORF1 encodes for a 40kDa protein with RNA binding and Nucleic Acid Chaperone activities; ORF2 encodes a potential 150kDa protein with demonstrated ENdonuclease (EN) and Reverse Transcriptase activities (RT); both proteins are required for LINE-1 mobilization known as retrotransposition (Figure 1).

Figure 1: Structure of a human RC-L1.

In a round of retrotransposition, an active L1 produce an intermediate RNA that is translated and subsequently integrated into genomic DNA by a process known as Target Primed Reverse Transcription or TPRT(Figure 2).

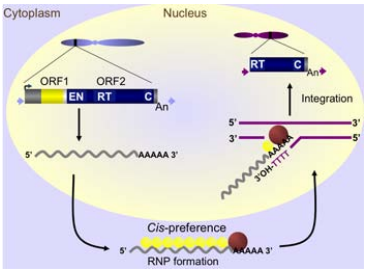

Figure 2: A working model of L1 retrotransposition. An RC-L1 is transcribed in the nucleus and then transported to the cytoplasm. ORF1p (yellow circle) and ORF2p (red circle) are translated and bind back to the L1 mRNA from which they were translated to form a cytoplasmic ribonucleoprotein particle (RNP). The RNP then translocates to the nucleus where retrotranspotion occurs by a process known as TPRT for Target Primed Reverse Transcription. The newly integrated L1 exhibits characteristics structures including 5´truncation, a poly(A) tail, and

Due to their ongoing mobilization in modern day humans, the activity of LINE-1 keeps impacting our genome. Indeed, recent estimates indicate that ~1/35 newborn individuals may contain a new LINE-1 mediated transposition event in their genome. Due to their mutagenic potential (as a LINE-1 insertion within a gene can disrupt its function), several human diseases have been reported to be caused by a LINE-1 retrotransposition event (cancer, haemophilia, muscular dystrophy, etc.). In most cases, both parents are non-carriers of the mutated gene but some of their newborn suffering a disease contains a *de novo* LINE-1 retrotransposition event, indicating that new LINE-1 copies are accumulated either in germ cells or soon after fertilization.

Other than as an insertional mutagen, the activity of LINE-1 (and mediated processes) has and had a great impact in the human genome and many human genes have evolved to use pieces of mobile DNA as part of their regulatory circuits (e.g., LINE-1

elements can alter the polyadenylation of the gene they insert in, can provide transcription factor binding sites, etc.). Other than the “dark side of LINE-1”, the presence of mobile DNA within a genome allows some degree of plasticity to the genome that can be beneficial over evolution. For instance, LINE-1 elements can directly create new genes by the mobilization of other mRNAs to new locus (termed exon shuffling). Finally, some types of retroelements (like *Alu*, with more than 1000000 copies dispersed per genome) rely on the activity of LINE-1 to mediate their mobilization (in *trans*). Overall, the activity of a single type of mobile DNA, LINE-1, has generated a THIRD of the human genome and its activity continues to impact our genome at different levels.

Due to the lack of a defined function, LINE-1 elements are considered as selfish DNA, which are present in the human genome solely because they are able to generate new heritable copies. As a result, the activity of LINE-1 is expected to be more prominent in cell types where new insertions could be dispersed in the population (e.g., germ cells and embryo). Indeed, we have recently shown that human embryonic stem cells (hESC, which represent a cell model of the Inner Cell Mass of a human embryo) express LINE-1 elements and can accommodate the activity of LINE-1 or retrotransposition. In addition, and using a mouse model of human L1 retrotransposition, it was surprisingly found that most *de novo* insertions occur in early embryo development. However, recent studies have described a somatic component of LINE-1 activity, which goes against the paradigm of selfish DNA. Why accumulate new LINE-1 copies in cell types that are not going to transmit them?.

Our hypothesis of work arises from our work in undifferentiated hESC and induced pluripotent stem cells, where LINE-1 are naturally overexpressed. Thus, we hypothesize that due to their stem status (*stemness*), and likely trough the panel of transcription factor present in stem cells, LINE-1 elements can be actively expresses in stem cells and as a consequence new LINE-1 copies could be accumulated. In the present research proposal, we have investigated the level of expression and activity of LINE-1 elements in several well characterized somatic stem cell populations: neuronal (NSCs), mesenchymal (MSCs), and haematopoietic (HSCs).

Since the beginning of the project, we have conducted L1 expression analyses in NSCs, MSCs and HSCs (in addition to fully differentiated cells as controls). In addition, we have conducted numerous LINE-1 mobilization assays in those same cell types. Altogether, the data obtained over the past 4 years has allowed us to demonstrate that LINE-1 elements are mostly expressed in NSCs, being the expression in MSCs and HSCs very low. Similarly, by the use of an engineered LINE-1 retrotransposition assay we have demonstrated that only NSCs can accommodate high levels of LINE-1 mobilization (a paper in preparation lead by a PhD student in my lab, Angela Macia). In NPCs, L1 insertions can occur into neuronally expressed genes; in addition, we determined the copy number of L1s in the human brain, and found that their copy number is higher than in other somatic tissues isolated from the same donor. In sum, these findings revealed that L1 elements are likely jumping in the human brain, and suggest that the genome of our brain is not static, although any biological significance remains to be determined. These findings were published as a letter in *Nature* in August 2009: “L1 retrotransposition in human neural progenitor cells” (accompanied by a *News&Views*). More recently, in late 2011 an independent group (directed by Dr.

Geoff Faulkner, The Roslin, UK) have obtained independent data in human brain samples, demonstrating ongoing LINE-1 mobilization in the human brain (also published in *Nature*). When compiled, the somatic activity of LINE-1 in humans indicate that is mostly a phenomenon restricted to NSCs and the human brain, opening a challenging question for the next years: what is the role of LINE-1 mobilization in the human brain? Is LINE-1 activity related to human brain disorders? Related to this question, in collaboration with groups in the US, we have recently demonstrated that LINE-1 mobilization is elevated in patients affected by the human disease Ataxia Telangiectasia, which has a number of brain specific symptoms (a study published in late 2011 in *Proc Natl Acad Sci U S A*).

Also related to this project, and by serendipity, while studying the activity of L1 in NSCs, we discovered that the expression of indicator cassettes delivered by LINE-1 into pluripotent cells is efficiently silenced during or immediately after integration. These data lead us to hypothesize that epigenetic processes may act as a host mechanism to counteract LINE-1 retrotransposition events in certain cell types. These data were published in a manuscript entitled: “Silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells” in *Nature* in 2010. Similarly, we have recently demonstrated that induced pluripotent stem cells do naturally express upon reprogramming (*Hum. Mol Genet, 2012*).

The completion of the project objectives has shed light in several unknown aspects of the LINE-1 and human biology in general as: how often is LINE-1 jumping? Where is LINE-1 jumping? What’s the impact of LINE-1 retrotransposition in somatic stem cells?. What are the repercussions of a mosaic genome to human health? How deep is the potential mutagenic effect of LINE-1 retrotransposition in somatic tissues? The answer to the above questions is just the beginning of a new field in human biology, as we now need to further understand the consequences of LINE-1 mobilization in the human brain. The data obtained in this proposal will be beneficial to the scientific community involved in the study of human biology and human health, in a worldwide level, and will firmly contribute to enhance EU scientific excellence.