



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# **Fanconi Anaemia and LINE-1 retrotransposition in the mammalian genome.**

Emma Clarke

PhD by Research

University of Edinburgh

2023





## Declaration

This thesis has been composed entirely by me. Apart from where stated, the experiments were carried out solely by me. The work in this thesis has not been submitted for any other degree or professional qualification.

Emma Clarke

March 2023



# Abstract

Transposable elements (TEs) are discrete, repetitive sequences of DNA that mobilise within genomes. For decades, TEs were dismissed as “junk DNA”, however, it is now clear that these elements have the potential to trigger genome instability, cause disease and shape the course of genome evolution.

L1 elements constitute the only autonomous elements which remain active in the human genome and comprises approximately 17% of human DNA. As a retrotransposon, L1 canonically mobilises through a “cut and paste” mechanism called target primed reverse transcription (TPRT). Due to the deleterious impacts of L1 activity, mammalian cells have evolved a range of mechanisms to suppress the mobilisation of these elements. The interactions between L1 elements and the host factors which regulate them are therefore an area of active research.

Several DNA repair genes have shown potential as regulators of L1 activity. Moreover, in cell lines deficient in non-homologous end-joining, L1 has shown the potential to retrotranspose without its ORF2p endonuclease, which is usually a requirement for canonical TPRT. This retrotransposition has been termed endonuclease independent (ENi) retrotransposition, and takes place at unrepaired double stranded breaks in the DNA. Interestingly, several DNA repair factors have also been identified as potential regulators of L1 retrotransposition (both positive and negative), including a number of proteins from the Fanconi Anaemia pathway. The relationship between these factors and L1 has yet to be fully characterised, and it remains to be seen whether L1 can exploit other DNA lesions in the way that it utilises DSBs in ENi retrotransposition.

This thesis aims to further investigate the relationship between L1 retrotransposition in the mammalian genome and DNA repair factors, particularly those comprising the Fanconi Anaemia pathway. Using cultured cell retrotransposition assays, I systematically tested a battery of mutant element in cells deficient in different proteins of the FANC pathway. In this way, I establish that ENi retrotransposition can be observed in a FANC background. I also demonstrate that FANC A deficient cells support retrotransposition of several L1 mutants which are immobile in parental cell lines. This includes elements with severe ORF1p mutations, mutations in the ORF2p endonuclease domain and mutations in the ORF2p PIP box. Despite testing a range of cell lines deficient in different DNA repair factors, including cells deficient in a range of FANC proteins, the retrotransposition of ORF1p, PIP and mutants appears to be unique to FANC A. My results are potentially indicative of a unique mechanism of retrotransposition in FANC A cells, a phenomena which has precedence in the ENi pathway of retrotransposition.

Mass spectrometry of immunoprecipitated T7-tagged ORF1p, both in FANC A and parental cells, demonstrated that a different selection of host factors interact with ORF1p in the two cell lines. Several of these have not been previously identified as L1 interactors, including YTHDF2, a protein which binds and destabilises m6A-containing RNA. Previous reports suggest that YTHDF2 regulates the stability of RNA:DNA hybrids *in vivo*, and associates with R loop containing loci. Through co-immunoprecipitation of YTHDF2 with ORF1p, I confirm that the protein interacts with L1 elements *in vitro*.

## Lay summary

DNA is the code for life which is shared by every organism. Small units of DNA are known as genes, and the combination of genes that each individual has can contribute to their physical characteristics as well as their predisposition to various diseases. The DNA of all living organisms also contains transposable elements, sections of genetic material that are capable of mobilising or “jumping” to different places in the code. These elements were once labelled as “junk DNA” and dismissed as inconsequential to genetic research, however the important role that they play in human health and evolution is now becoming apparent.

Some transposable elements move by cutting and pasting themselves into different sites in the DNA, whereas some move through more of a “copy and paste” mechanism. However, the only elements which are currently active in humans are the “copy and paste” elements. These elements produce proteins which work together to make an incision, or break, in the DNA, and then insert a copy of their own sequence into that break.

These elements behave as “selfish genes” and do not code any immediate benefits for their host organism. In fact, their activity is often disruptive. If they copy themselves into important genes, for example, they can prevent those genes from functioning correctly, resulting in instability within the DNA and increasing the risk of disease, even death. Because of this, our cells have evolved means of repressing the activity of transposable elements.

Fanconi Anaemia is a rare genetic disorder which results in the reduced production of blood cells in patients, as well as physical abnormalities and a predisposition to types of Anaemia. The disorder is triggered when any of the 21 proteins in the Fanconi Anaemia pathway stop working correctly. Recently, research suggests that some proteins in the Fanconi Anaemia pathway may also be involved in suppressing the activity of transposable elements in human DNA. In this thesis, I demonstrate that some of these genes reduce the capability of transposable elements to move or “jump” in living cells cultured in the lab.

By culturing and performing experiments in cells which do not have an important Fanconi Anaemia gene, I also showed that transposable elements missing key parts of their code can still move in these cells. This is interesting, because it means that removing parts of the Fanconi pathway could enable the elements to move in an unusual way. Seeing this difference helps us to better understand how the Fanconi Anaemia pathway could be working to repress the transposable elements in normal cells. This new, unusual way that transposable elements jump in cells that have defects in the Fanconi Anaemia pathway may be happening in Fanconi Anaemia patients and contributing to this disease.



Further research will be required to clarify this. Nonetheless, this research has taken the field one step closer to understanding how transposable elements can influence, and be influenced by, DNA repair genes.



## Acknowledgements

There are many people who made this work possible. I'd like to thank all the members of the Jose Garcia-Perez lab. Thank you to MJ, Kiko and Emilie who offered constant support, interesting discussion and were always willing to offer guidance. I couldn't ask for a kinder group. Thank you to Jose for the supervision of the project and creative ideas behind the experimentation, Ian Adams for offering a great deal of guidance and graciously taking the time to help me through the last stages of writing, and Nick Gilbert for continued positivity and encouragement.

Thank you to all of the incredible staff at the Institute, in particular Lizzie from FACs and Alex Von Kriegsheim for mass spectrometry.

Thank you to my friends and fellow PhD students and to Sam, who kept me sane by providing a space for laughter, vents and reassurance. Your camaraderie got me through this time.

Thank you to the members of the Andrew Jackson lab, who freely and generously shared resources and knowledge as we started out as a small lab.

Thank you to the MRC for funding my studies.

Thank you to the little community of friends I have made in Malta who continue to bring joy into my life and endured far more conversation about this thesis than they probably wanted to. Thank you to Ed for the amazing food, and the endless patience, positivity and understanding. You helped to make this possible for me. Thank you to my mum, dad and sister who have been a relentless source of love, the culmination of all of this is that we can have conversations without worrying about "the thing that can't be named".



## Table of Contents

List of Figures .....	6
List of Tables .....	9
Chapter 1.....	15
1.1. Transposable elements in the human genome .....	17
1.1.1. An introduction to transposable elements.....	17
1.1.2 L1 elements: active retrotransposition in the human genome .....	19
1.1.3 The impacts of L1 retrotransposition on the human genome .....	20
1.1.4 L1 elements and evolution of the human genome.....	23
1.2 The dynamics of L1 retrotransposition .....	24
1.2.1 The structure of an active L1 element .....	24
1.2.2 The mechanism of L1 retrotransposition.....	26
1.2.3 A cultured cell retrotransposition assay to measure retrotransposition .....	28
1.2.4 The formation of a ribonucleoprotein particle is essential for retrotransposition .....	30
1.2.5 The structure and function of ORF1p .....	31
1.2.6 ORF1p has been difficult to characterise experimentally.....	33
1.2.7 The enzymatic activity of ORF1p in TRPT.....	33
1.2.8 Expression of ORF1p is linked to cellular toxicity. ....	34
1.2.9 Identification of a PCNA interaction protein box in ORF1p .....	35
1.3 The endonuclease independent pathway of retrotransposition.....	36
1.4 L1 activity is regulated by host factors in the human genome.....	39
1.4.1 Transcriptional regulation of LINE- 1 .....	39
1.5 DNA repair in the human genome.....	41
1.5.1 DNA repair proteins regulate retrotransposition. ....	42
1.5.1.1. Nucleotide excision repair (NER) pathway .....	42
1.5.1.1. Non-homologous end joining pathway.....	43
1.5.1.2 BRCA1 identified as a L1 inhibitor.....	43
1.6 The Fanconi Anaemia pathway of DNA repair.....	44
1.6.1 The mechanism of ICL repair by the Fanconi Anaemia pathway.....	47
1.6.2 The Fanconi Anaemia pathway may also work to resolve stalled replication forks.....	48
1.6.3 R loops in Fanconi Anaemia .....	49
1.6.4 Additional roles of Fanconi Anaemia proteins.....	51
1.7 The Fanconi Anaemia pathway may regulate L1 retrotransposition.....	52
1.7.1 L1 retrotransposition at replication forks.....	53
1.8. Aims of this thesis .....	54

Chapter 2.....	56
2.1 General reagents.....	58
2.2 Cell culture methods:.....	58
2.2.1 Cell lines used in this thesis.....	59
2.2.2 Preservation of cell lines. ....	60
2.2.3 Transfection of cell lines for retrotransposition assay.....	60
2.2.3.1 Transfection of human cell lines for cultured cell retrotransposition assay. ....	60
2.2.3.2 Transfection of hamster cell lines. ....	60
2.2.3.4 Transfection of hamster cells for co-immunoprecipitation.....	61
2.2.4 Measurement of transfection efficiency .....	61
2.2.5 Cultured cell retrotransposition assay.....	62
2.2.6 Mitomycin C Sensitivity assay. ....	63
2.2.7 Generation of FANCC omplemented cells .....	63
2.3 Nucleic acid methods .....	64
2.3.1 Quantification of nucleic acids.....	64
2.3.2 Gel electrophoresis for visualisation of nucleic acids .....	64
2.3.3 Plasmids used in this thesis.....	64
2.4 Protein methods .....	67
2.4.1 Immunoprecipitation.....	67
2.4.2 Western blotting .....	68
2.4.2.1 Generation of lysate from whole cells. ....	68
2.4.2.2. Quantification of protein in whole cell lysate.....	69
2.4.2.3. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) for protein resolution .....	69
2.4.2.4 Immunoblotting .....	70
2.5 Microbial methods .....	71
2.5.1 Growth of bacteria .....	71
2.5.2 Transformation of E. coli.....	71
2.5.3 Generation of ultra-competent cells. ....	72
2.5.4 Testing of ultracompetent cells. ....	73
2.6 DNA methods .....	73
2.6.1 In house DNA sequencing .....	73
2.6.2 Primers for sequencing L1 plasmids .....	73
2.6.3 Small scale purification of plasmid DNA .....	76
2.6.4 Large scale purification of plasmid DNA .....	76
2.6.5. Restriction digests .....	76

2.6.6 Ligation of digested DNA.....	76
2.7 Mass Spectrometry .....	77
Chapter 3.....	79
3.1 Introduction .....	81
3.2 Cells deficient in Fanconi anaemia are hypersensitive to Mitomycin C .....	82
3.3 Optimisation of parameters for a retrotransposition assay in FANCA deficient cells. ....	83
3.4 Optimising the transfection of VH4 and V79 cells. ....	85
3.5 Optimising the seeding density .....	87
3.6 Optimising retrotransposition cell culture assay in FANCA complemented and NHEJ deficient cell lines.....	88
3.7 Main Conclusions .....	93
3.8 Discussion.....	93
Chapter 4.....	95
4.1 Introduction .....	97
4.2 Research Aims.....	98
4.3 Testing wild type retrotransposition in FANCA (VH4) and parental V79 cells. ....	99
4.4. Testing a battery of ORF1p mutants in FANCA mutant cells. ....	102
4.5. Rescuing unusual retrotransposition in VH4 cells with complementation using FANCA .....	106
4.6 Testing ORF2 only retrotransposition in FANCA (VH4) and parental (V79) cells.....	107
4.7 Testing EN mutants in FANCA deficient cells.....	109
4.8 Testing other ORF2 mutants in FANCA deficient cells. ....	111
4.9 Main conclusions .....	115
4.10 Discussion.....	115
Chapter 5.....	119
5.1 Introduction .....	121
5.2 Aims.....	122
5.4 Testing the retrotransposition of L1-ORF1p mutant elements in FANCC deficient cells.....	126
5.5 Testing the retrotransposition of L1-ORF1p mutant elements in NHEJ deficient cells. ....	128
5.6 Testing the retrotransposition of L1-ORF1p mutant elements in human FANCA deficient cells.....	130
5.7 Testing the retrotransposition of L1-ORF1p mutants in FANCC mutant cells. ....	134
5.8 Testing the retrotransposition of L1-ORF1p mutants in NHEJ cells.....	136
5.9 Testing the retrotransposition of L1-ORF1p mutants in FANCA deficient human cells.....	138
5.10 Testing for an interaction between FA proteins and L1 encoded proteins .....	141
5.11 Main conclusions .....	143
5.12 Discussion.....	143

Chapter 6.....	147
6.1 Introduction .....	149
6.2 L1-ORF1p proteomic studies in FA-deficient cells: identification of host interactors with L1-ORF1p.....	152
6.3 A comparative analysis of host factors previously identified in L1 proteomic studies. ....	159
6.4 L1-ORF1p proteomic studies in NHEJ-deficient cells: identification of host interactors with L1-ORF1p.....	162
6.5 The role of YTHDF2 in the regulation of L1 retrotransposition in FA deficient cells. ....	166
6.6 Testing for an interaction between YTHDF2 and ORF1p .....	166
6.7 Investigating LINE-1 retrotransposition in YTHDF2 KO cells .....	167
6.8 R-loops , FA deficient cells and deregulated L1 retrotransposition.....	171
6.9 Discussion.....	179
Chapter 6.....	183
7.1 Summary of the findings in this thesis.....	185
7.2 The role of ORF1p and the ribonucleoprotein particle in retrotransposition .....	187
7.3 The role of the PIP box, and the ORF1p/PCNA interaction.....	189
7.4 FANC proteins as regulators of L1 retrotransposition .....	190
7.5 YTHDF2 as a potential negative regulator of retrotransposition.....	191
7.6 A potential model: R-loops as a site for insertion. ....	192
7.7 Future work and outstanding questions.....	193
8.0 Appendices.....	196
8.1 Antibodies used in this thesis .....	196
References .....	196





## List of Figures

Figure 1.1. Transposable elements in the human genome.....	18
Figure 1.2. Mechanisms by which L1 can impact the human genome.....	21
Figure 1.3. The structure of an active L1 element.....	24
Figure 1.4. The process of target primed reverse transcription.....	26
Figure 1.5. A cultured cell assay to measure L1 retrotransposition.....	29
Figure 1.6. The mechanism of endonuclease independent retrotransposition.....	38
Figure 1.7. The Fanconi Anaemia pathway repairs interstrand crosslinks.....	47
Figure 1.8. The structure of an R loop.....	49
Figure 3.1. Sensitivity of FANCA deficient (VH4) cells to MMC in comparison with the parental (V79) cell line.....	83
Figure 3.2. Death Curve of FANCA deficient (VH4) and parental (V79) cells using G418.....	85
Figure 3.3. Optimisation of transfection efficiency in FANCA (VH4) and parental (V79) cell line.....	87
Figure 3.4. Optimisation of seeding density in FANCA deficient and parental cell line.....	88
Figure 3.5. Development of a cell line complemented with FANCA .....	90
Figure 3.6. Death curve to identify the optimal antibiotic concentration in FANCA complemented cells, NHEJ deficient (XR1) and parental (4364) cells.....	92
Figure 3.7. Optimisation of transfection efficiency in FANCA complemented, NHEJ deficient (XR-1) and parental (4364) cell lines.....	93
Figure 4.1. Testing wild type L1 retrotransposition in FANCA (VH4) and parental (V79) cells.....	102
Figure 4.2. Complementation with FANCA rescues elevated retrotransposition in VH4 cells.....	104
Figure 4.3. Scheme of previously characterised ORF1 mutants.....	106
Figure 4.4. Testing type I ORF1 mutations in FANCA deficient and parental cell lines.....	107
Figure 4.5. Testing type II ORF1p mutations in FANCA deficient (VH4) and parental (V79) cell line. 109	
Figure 4.6. Rescuing increased retrotransposition of ORF1 L1 elements with FANCA complementation. ....	111
Figure 4.7. Testing ORF2 only retrotransposition in FANCA (VH4) and parental (V79) cells.....	113
Figure 4.8. Testing Endonuclease mutants in VH4 and parental cells.....	115
Figure 4.9. Measuring retrotransposition of ORF2 mutants in FANCA (VH4) and parental (V79) cells.....	

Figure 4.10. Increased retrotransposition of ORF2 mutants is rescued by complementation with FANCA .....	117
Figure 5.1. Mitomycin C (MMC) sensitivity assay in FANCC (CLV5B) and parental (V79) cells.....	129
Figure 5.2. FANCC regulates WT L1 retrotransposition .....	130
Figure 5.3. Testing a series of L1-ORF1p mutants in FANCC deficient cells.....	132
Figure 5.4. Testing a series of L1-ORF1p mutants in NHEJ deficient cells. ....	134
Figure 5.5. Testing a series of L1-ORF1p mutants in FANCI deficient cells.....	137
Figure 5.6. Testing a series of L1-ORF1p mutants in FANCD2 deficient cells.....	138
Figure 5.7. Testing a series of L1-ORF1p mutants in FANCC deficient cells.....	139
Figure 5.8. Testing a series of L1-ORF1p mutants in FANCI deficient cells.....	144
Figure 5.9. Testing a series of L1-ORF1p mutants in FANCD2 deficient cells.....	145
Figure 5.10. Co-immunoprecipitation of FA proteins and T7 tagged L1-ORF1p.....	147
Figure 6.1. Immunoprecipitation of L1-ORF1p-T7 interactors.....	158
Figure 6.2. Volcano plot showing MS analysis of proteins following immunoprecipitation of L1-ORF1p-T7 in FANCA (VH4) cells.....	159
Figure 6.3. Volcano plot showing MS analysis of proteins following immunoprecipitation of L1-ORF1p in parental (V79) cells.....	160
Figure 6.4. GO term analysis of functional pathways enriched in L1-ORF1p interactors identified in FANCA cells.....	163
Figure 6.5. GO term analysis of functional pathways enriched in L1-ORF1p interactors identified in NHEJ mutant cells.....	169
Figure 6.6 Co-immunoprecipitation of YTHDF2 and L1-ORF1p.....	170
Figure 6.7. Optimisation of transfection efficiency in parental HAP 1 and YTHDF2 KO cells.....	172
Figure 6.8. Retrotransposition of L1-ORF1p mutants in HAP1 and YTHDF2 KO cells.....	173
Figure 6.9. Retrotransposition of L1-ORF1p mutants in HAP1 and YTHDF2 KO cells.....	174
Figure 6.10. Overexpression of YTHDF2 and L1 Retrotransposition in HeLa cells.....	177
Figure 6.11. Overexpression of YTHDF2 and L1 Retrotransposition in V79B cells.....	178
Figure 6.12. Overexpression of YTHDF2 and L1 Retrotransposition in VH4 cells.....	179
Figure 6.13. Overexpression of human nuclear RNaseH1 and L1 Retrotransposition in V79B cells... ..	180
Figure 6.14. Overexpression of human nuclear RNaseH1 and L1 Retrotransposition in VH4 cells.....	181
Figure 6.15. Model of L1 Retrotransposition in WT and FA deficient cells.....	182



## List of Tables

Table 2.1 Cell culture drug stock solutions.....	58
Table 2.2 Cell lines used in this thesis.....	59
Table 2.3 Relative surface area of different cell culture vessels used in conversion.....	62
Table 2.4 Plasmids used in this thesis.....	65
Table 2.5 Immunoprecipitation buffers used in this thesis.....	67
Table 2.6 Reagents used in western blotting.....	68
Table 2.7 Working concentrations of antibiotics.....	71
Table 2.8 Buffers for generation of ultracompetent cells .....	72
Table 2.9 Criteria used for generation of ultracompetent cells.....	73
Table 2.10 Primers for sequencing L1 plasmids.....	74
Table 5.1 FA proteins previously identified as L1 regulators.....	127
Table 6.1 List of significantly enriched ORF1p-T7 interactors in FANCA and parental cells .....	160
Table 6.2 Previously identified L1 interactors from the current study.....	163
Table 6.3 Overlapping L1 interactors listed by study.....	163
Table 6.4 List of significantly enriched ORF1p-T7 interactors in NHEJ deficient (XR-1) and parental (4634) cells.....	166



## Abbreviations

A	Alanine
BER	Base Excision Repair
Bp	Base Pair
BRCA1/2	BReast CAncer susceptibility gene
cDNA	Complementary DNA
C	Cysteine
CFU	Colony Forming Units
Co-IP	Co-immunoprecipitation
Conc	Concentration
Comp	Complemented
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DDR	DNA damage response
dH <sub>2</sub> O	Distilled water
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	DeoxyriboNucleic Acid
DSB	Double Strand Break
EN	Endonuclease
ENi	Endonuclease Independent
ENm	Endonuclease mutant
FA	Fanconi Anaemia
FACs	Fluorescence Activated Cell sorting
g	Grams
GFP	Green Fluorescent Protein
h	Hours
ICL	Interstrand Cross Link
IP	Immunoprecipitation
Kb	Kilobase
kDa	Kilodaltons

KO	Knockout
L1	Line-1
L	Lysine
M	Molar
Mass spec	Mass Spectrometry
MEFs	Mouse embryonic fibroblasts
mg	Milligram
Min	Minutes
MMC	Mitomycin C
mRNA	Messenger RNA
Neo	Neomycin
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End Joining
ng	Nanogram
nt	Nucleotide
ORF1	Open Reading Frame 1
ORF1p	Open Reading Frame 1 protein
ORF2	Open Reading Frame 2
ORF1p	Open Reading Frame 2 protein
ORF0	Open Reading Frame 0
PBS	Phosphate Buffer Saline
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PIP	PCNA Interacting Protein
PIPm	PCNA Interaction Protein mutant
RF	Replication Fork
RPM	Revolutions Per Minute
RNA	Ribonucleic Acid
RNAse	Ribonuclease
RNAi	RNA Interference
RNP	RiboNucleoprotein Particle



RT	Reverse Transcriptase
RTm	Reverse Transcriptase mutant
S	Serine
SD	Standard Deviation
SDS	Sodium dodecyl sulphate
sec	Second(s)
SEM	Standard Error of the mean
siRNA	Short/small interfering RNA
SNP	Single Nucleotide Polymorphism
SRF	Stalled Replication Fork
ssDNA	Single Stranded DNA
T	Threonine
TSD	Target Site Duplication
Ub	Ubiquitin
UTR	Untranslated Region
UV	Ultraviolet
V	volt
W	Tryptophan
WT-L1	Wild type Line-1
WT	Wild type
YTHDF2	YTH N6-Methyladenosine RNA Binding Protein F2



## Chapter 1

---

# Introduction



## 1.1. Transposable elements in the human genome

### 1.1.1. An introduction to transposable elements

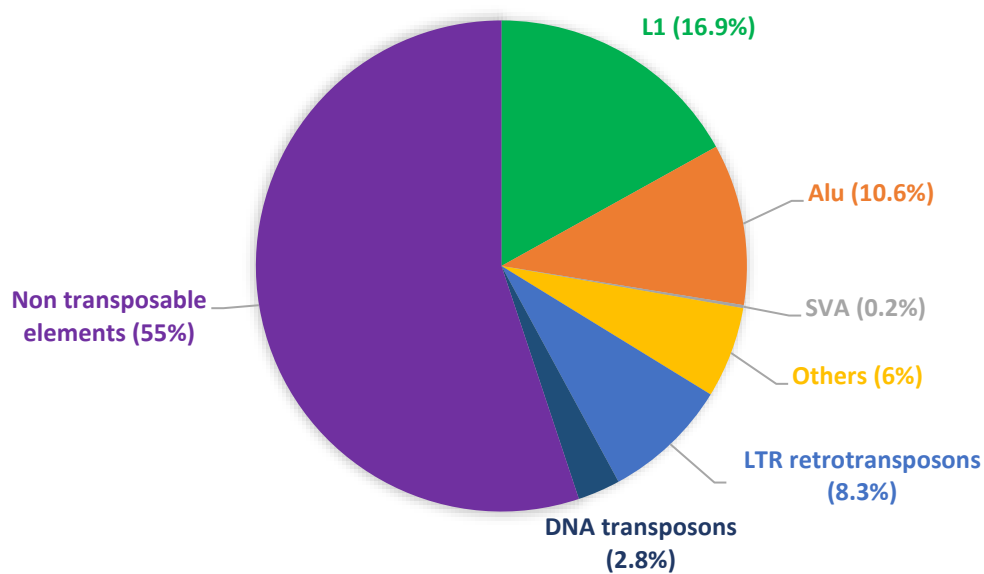
Transposable elements (TEs) are discrete repetitive DNA sequences which have the ability to mobilise within and invade genomes. TEs form a major component of both bacterial and eukaryotic genomes, though the specific proportion varies according to species. In eukaryotes, this proportion ranges from approximately 5% in yeast to approximately 90% in wheat genomes (Charles *et al.* 2008). TEs contribute to genome instability by a number of mechanisms, ranging from insertions to larger scale genomic rearrangements

Transposable elements are subdivided into two classes according to the mechanism of mobilisation. Retrotransposons, or Class I elements, move using a “copy and paste” mechanism, involving the reverse transcription of an intermediate retrotransposon mRNA which is integrated into the host genome at a new location (Gilbert *et al.* 2002). The use of an mRNA intermediate distinguishes retrotransposons from DNA transposons, the second class of TEs, which mobilise by a simpler “cut and paste” process. Though DNA transposons make up approximately 3% of the human genome, they are considered “fossils,” as research indicates that no human DNA transposon family has been active in the last 37 million years of primate evolution (Pace and Feshotte, 2007)

An estimated 40% of the human genome is comprised of retrotransposons (**Figure 1.1**), though this figure is likely an underestimation, as ancient elements will have been modified through evolution and may no longer be identifiable. Retrotransposons can be divided into long terminal repeat (LTR) containing elements, which make up approximately 8.3% of the genome, and non-LTR elements. Non-LTR elements can be further subdivided into LINEs (long interspersed elements) and SINEs (short interspersed elements) and make up approximately a third of the human genome together. (Cordaux and Batzer 2009; Lander *et al.* 2001).

Within these classes are distinct subfamilies, which show different levels of activity and can be further described as either autonomous (able to mobilise through use of their own reverse transcriptase), or non-autonomous (requiring the enzymatic machinery of another retrotransposon to mobilise). LINE-1 (L1) elements are currently the only autonomous, active elements in the human genome and provide the enzymatic machinery for the mobilisation of non-autonomous elements. Other prolific elements in the human genome include Alus and SVAs (both of which are categorised as SINEs), which are represented by approximately 1 million and 3,000 copies in the genome respectively (Ostertag *et al.* 2003; Lander *et al.* 2001).

Elements which have retained their activity through evolution are of a great deal of interest, as they continue to generate inter-individual genetic variability and contribute to genomic instability processes through a myriad of mechanisms (Beck *et al.* 2011) .



**Figure 1.1 Transposable elements in the human genome.**

Approximately 45% of the human genome is made up of mobile elements. Of these, 8.3% is made up of LTR retrotransposons and 33% is made up of non LTR transposons including L1, Alu and SVA elements (adapted from Cordaux and Batzer 2009).

### 1.1.2 L1 elements: active retrotransposition in the human genome

Any ongoing retrotransposition in the human genome is mediated by the enzymatic machinery of active L1 elements, a family of the LINE class of retrotransposons (Furano, 2000). Alu elements are believed to be more active, however as non-autonomous elements they still make use of the L1 enzymatic machinery to mobilise. There are several subfamilies of L1 elements, distinguished by sequence differences that have arisen through evolution. However, the only elements currently active in the human genome belong to the human specific subfamily L1HS-Ta (for transcriptionally active), the youngest subset of L1s. L1 elements make up approximately 17% of the genome (Lander *et al.* 2001), though it is estimated that over 99.8% of these have become modified through truncations, internal rearrangements and mutations throughout evolution, and are no longer capable of mobilisation (Cordaux and Batzer 2009). It is estimated that 80-100 retrotransposition-competent L1s continue to impact the genome (Beck *et al.* 2010; Brouha *et al.* 2003). Of these, a small proportion are highly active or “hot” L1s, which are thought to account for the majority of L1 retrotransposition (Brouha *et al.* 2003).

In recent years, genomic, molecular and computation approaches have been used to identify active L1 elements. Many human specific L1 elements appear to be polymorphic in human populations (Streva *et al.* 2015; Lutz *et al.* 2003), and many are not represented in the human genome reference sequence. Analysis of the in vitro activity of 82 L1 clones from the HGW showed that just six hot L1s accounted for 84% of retrotransposition activity. (Brouha *et al.* 2003) However, a more recent study identified 68 L1 elements in the genomes of six individuals of diverse geographical origin, of which 37 were shown to be “hot” for retrotransposition in a cultured cell assay (Beck *et al.* 2010; Beck *et al.* 2011). This indicates that “hot” L1s may be more abundant in human populations than initially thought, and may be an important source of interindividual genetic variation. Thus, understanding the dynamics of L1 mobilisation in the human genome is an area of ongoing research.

In order to ensure evolutionary success, L1 elements must mobilise in a cell type which will allow insertions to be passed on to the next generation, such as in the germline and during early embryogenesis (Trelogan and Martin 1995; Garcia-Perez *et al.* 2007). Active retrotransposition in the human germline is thought to provide a source of genetic variation and novel genetic material, influencing evolutionary processes from the DNA level (Cordaux and Batzer 2009). However, the genetic instability and insertional mutagenesis associated with retrotransposition events can be deleterious to the individual and the rate of de novo germline insertions is therefore likely under evolutionary constraint (Ivancevic *et al.* 2016). Numerous attempts have been made to establish the rate of L1 retrotransposition. Depending on the methodology, estimates range from around 1 in 20 to

1 in 200 births. Alu insertions are believed to take place more frequently, around 1 in 20 births (Cordaux and Batzer, 2009), and SVA insertions less frequently, at around 1 in 900 (Xing *et al.* 2009)

As a result of their deleterious impacts, retrotransposons are predominantly subject to epigenetic silencing in somatic tissues, and are reactivated during the genome-wide demethylation that occurs in early embryogenesis. Indeed, work in human pluripotent stem cells and using mouse models has shown that the majority of heritable insertions are generated within these contexts (Babushok *et al.* 2006; Garcia-Perez *et al.* 2007; van den Hurk *et al.* 2007; Kano *et al.* 2009; Levin and Moran 2011).

In recent years, however, a body of research has built up to suggest that retrotransposition may also take place in certain somatic tissues and is not limited to embryogenesis in humans. In particular, a number of studies suggest that L1 actively retrotransposes in the human brain, potentially contributing to somatic mosaicism in adult tissue and generating variation between neurons (Muotri *et al.* 2005; Zhao *et al.* 2019). Muotri *et al.* (2005) demonstrated that an engineered human L1 construct could mobilise in neuronal precursor cells generated from rat neural stem cells. The group showed that insertions from these events could influence the expression of neural genes, as well as altering cell fate. Furthermore, a study by Faulkner and Bilon (2018) revealed that a subset of L1 elements could be de-repressed in the soma despite being inactive in the germline.

### 1.1.3 The impacts of L1 retrotransposition on the human genome

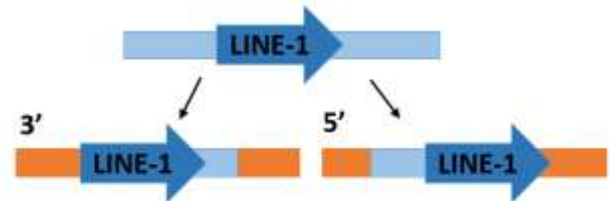
L1 elements, therefore, continue to impact the human genome in a myriad of ways, leading to disease and contributing to human evolution. Among other mechanisms, L1 elements can inactivate gene function through insertional mutagenesis, the triggering of aberrant splicing, target site deletions, recombination and transduction. **Figure 1.2** illustrates some additional examples of mechanisms by which L1 has been reported to impact the human genome.



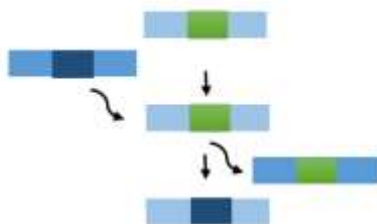
**a) Insertional Mutagenesis**



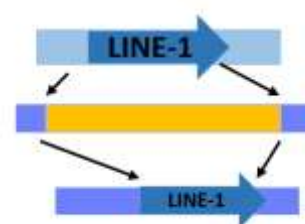
**b) Transduction**



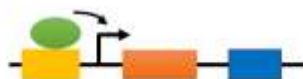
**c) Gene Conversion**



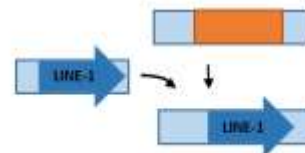
**d) Rearrangements**



**e) Modulation of gene expression**



**f) Insertion mediated deletions**



**Figure 1.2. Mechanisms by which L1 can impact the human genome.**

a) Insertional mutagenesis by L1 can lead to disease in some instances, as reported in the pathology of Haemophilia A (Kazazian *et al.* 1988) b) transduction by LINE-1 is believed to represent ~1% of the human genome (Pickeral *et al.* 2000) c) molecular biology and DNA sequence analysis has indicated that both LINE elements and Alus can act as substrates for gene conversion (Tremblay *et al.* 2000). d) genomic rearrangements, such as unequal homologous recombination, has been reported as a mechanism of L1 mediate genome instability (Burwinkle *et al.* 1998) e) L1 elements have been reported to act in cis modulating the expression of host genes (Wanichnopparat *et al.* 2013) f) Insertion of L1 elements at a genomic site has been reported to cause deletion of adjacent genetic material (Gilbert *et al.* 2002). Deletions range from 1bp to >130bp (Gilbert *et al.* 2002; Symer *et al.* 2002; Gilbert *et al.* 2005). (Adapted from Cordaux and Batzer 2009).

There are many reports linking L1 retrotransposition to disease through various mutagenic processes, with examples spanning haemophilia, cancer and diabetes (Kazazian *et al.* 1988; Miki *et al.* 1992; Martin-Nunez *et al.* 2014). Though the contribution of L1 to human pathology is still being established, Callinan and Batzer (2006) estimate that transposable elements cause ~0.27% of disease mutations and L1-mediated retrotransposition events are estimated to account for approximately 1 of every 1000 spontaneous, pathological mutations in the human genome (Chen *et al.* 2005; Kazazian *et al.* 1998). Reports of L1 mediated disease have been reviewed in Kaer and Speek (2013).

In 1988, Kazazian *et al.* published the first description of an endogenous L1 element causing a disease through insertional mutagenesis. The group observed L1 insertions in exon 14 of the factor VIII gene in two patients with haemophilia A (of 240 analysed). The insertions were 3.8 and 2.3 kilobases and generated target site duplications of 12 and 13 nucleotides respectively. Since this study, many instances of L1 causing disease through insertional mutagenesis have been reported (Van Den Hurk *et al.* 2003; Mukherjee *et al.* 2004; Qian *et al.* 2015; Brouha *et al.* 2002; Narita *et al.* 1993; Meischl *et al.* 2000; Lanikova *et al.* 2013; Morisada *et al.* 2010; Teugels *et al.* 2005; Kondo-lida *et al.* 1999).

As well as through insertional mutagenesis, L1 insertions can disrupt the genome via transductions (Pickeral *et al.* 2000). In some cases, transductions take place when RNA polymerase II reads through the weak poly A signal of the element to the DNA following the insertion. This is speculated to take place in 15-23% of L1 retrotransposition events. (Holmes *et al.* 1994; Goodier *et al.* 2000; Pickeral *et al.* 2000; Szak *et al.* 2002). In other cases, L1 activity can lead to the transduction of regulatory sequences. This process can affect gene expression or even result in the formation of new genes if different exons come together ectopically (Solyom *et al.* (2012a)). An example of a L1 mediated transduction was described by Solyom *et al.* (2012a), who identified the movement of a novel non-coding gene into exon 67 of the dystrophin gene.

L1 elements can also cause mutation through a mechanism called ‘gene breaking’. Gene breaking takes place when a L1 inserts into an intron in the antisense orientation, resulting in the splitting of the gene into two smaller transcripts (Wheelan *et al.* 2005). Wheelan *et al.* (2005) identified three genes which appeared to be split by L1 insertions, as well as 12 additional candidates. This phenomenon takes place because the presence of the element truncates the transcript from a gene’s promoter. These experiments were performed on episomal gene fusions; however, the authors propose that the same mechanism could lead to gene breaking in native chromosomal genes. In some

cases, L1 elements have also been reported to provide an alternative promoter once they insert into a new location (Dunn *et al.* 2003).

Alternatively, elements can create genetic rearrangements through ectopic recombination via non-allelic homologous recombination (NAHR; Beck *et al.* 2011). An example of this is reported by Wu *et al.* (2014), who observed one large deletion that had taken place in the F9 gene in patients with severe Haemophilia B. As it took place between two tandem L1 elements, the group proposed that the deletion had been mediated by NAHR.

#### 1.1.4 L1 elements and evolution of the human genome

As well as contributing to disease, the genomic changes mediated by L1 can also influence species formation and evolution by altering the structure of the genome (Cordaux and Batzer, 2009). L1 elements have remained persistently active in the genome for tens of millions of years, slowly accumulating in copy number and altering the genome through the previously described mechanisms and others (Beck *et al.* 2011; Cordaux and Batzer, 2009). In some instances, these changes can result in the creation of novel genetic material. For example, cell culture assays have found that L1 mediated 3' transductions can result in exon shuffling (Moran *et al.* 1999), and analysis of the human genome has suggested that 0.6-1% of human DNA could be derived from L1 mediated transduction (Goodier *et al.* 2000; Pickeral *et al.* 2000). Evidence that an SVA mediated transduction was responsible for generation of the acyl-malonyl condensing enzyme 1 (AMAC1) gene family has demonstrated the potential for retrotransposition activity to create genetic material (Xing *et al.* 2006).

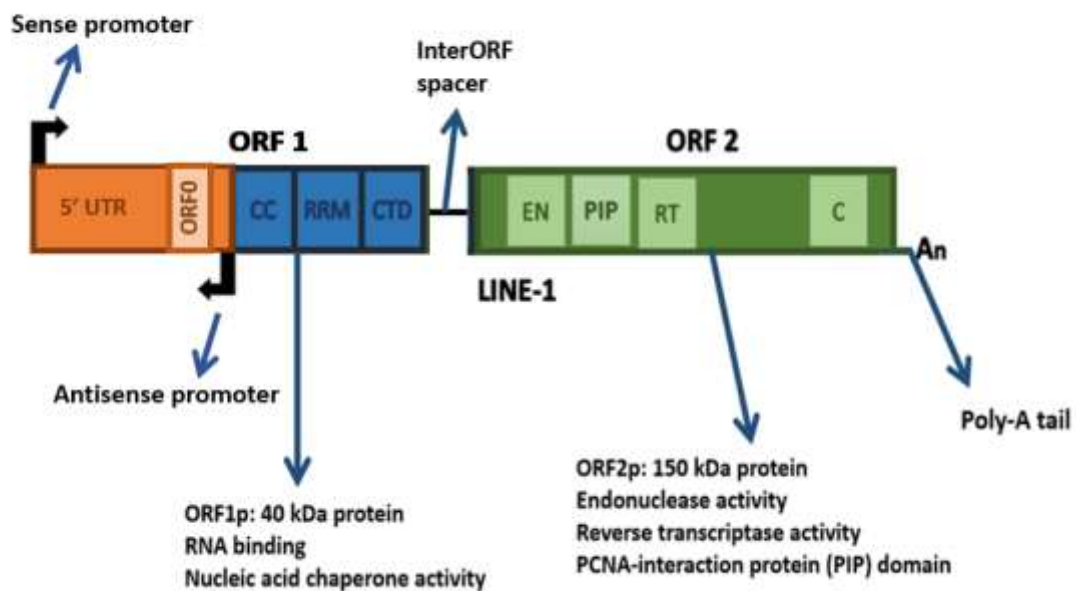
Transposable elements can also become recruited into genes in a process called "exonisation" (Gonçalves *et al.* 2017). It has been predicted that the proportion of coding sequences including TEs is around 0.1% (Gotea and Makalowski 2006).

It is also known that L1-mediated recombination can lead to genome deletions, leading in some cases to genetic disorders (Gilbert *et al.* 2002). However, in recent years, as sequencing technologies and comparative genomics have become more advanced, we have a better understanding of how L1 mediated deletions have altered human and other genomes on an evolutionary scale (Sen *et al.* 2006; Cordaux *et al.* 2008). Work to better understand the impact of L1 on the human genome on an evolutionary level is ongoing.

## 1.2 The dynamics of L1 retrotransposition

### 1.2.1 The structure of an active L1 element

A full-length active human L1 element is 6Kb in length (Beck *et al.* 2011; **Figure 1.3**). The 5' UTR region of human L1 is approximately 900bp long and contains an RNA polymerase II promoter, as well as a binding site for transcription factor YY1, and two binding sites for SOX-family transcription factors (Tchénio *et al.* 2000; Athanikar *et al.* 2004). Each of these binding sites is required for the activation of retrotransposition (Athanikar *et al.* 2004).



**Figure 1.3** The structure of an active L1 element

(Adapted from Zhang et al 2020)

The 5'UTR region of the L1 transcript also contains a short open reading frame, ORF0, which was identified in 2015 by Denli *et al.* ORF0 is located on the opposing strand to the subsequent open reading frames. ORF0 can form fusion proteins with proximal exons and appears to improve L1 mobility. Though its exact mechanistic role in TRPT is now known, its transcription by an antisense promoter may enhance L1 activity (Denli *et al.* 2015).

The 5' UTR region is followed by two open reading frames, ORF1 and ORF2 (both required for retrotransposition (Mathias *et al.* 1991; Moran *et al.* 1996)), separated by an inter-ORF spacer. This spacer contains two in frame stop codons as well as an in-frame AUG codon with the potential to generate short transcripts of 6 amino acids. Mutations within this inter-ORF spacer appear to have little impact on retrotransposition (Alisch *et al.* 2006).

The human ORF1 encodes a 500 amino acid, 40kDa protein: ORF1p. This protein contains a long coil coiled domain, which contains a leucine zipper (Holmes, 1992) upstream of an RNA recognition motif (RRM; Khazina and Weichenrieder 2009) and a C-terminal domain (Januszyk *et al.* 2007), both of which are highly conserved. The precise role of ORF1p in retrotransposition has been difficult to determine, partly because its evolutionary origin is unclear and also because its structure shares little homology with other proteins (Martin *et al.* 2006). However, from analysis of crystal and NMR solution structures it is evident that ORF1p is a highly efficient RNA packaging protein (Januszyk *et al.* 2011). It shows a high affinity for single stranded RNA and DNA, and in particular shows a cis-preference for its own transcript (Martin *et al.* 2006). Structural analysis has revealed that ORF1p forms dumbbell-like, non-spherical trimers via the coiled-coil domain (Martin *et al.* 2003), and that close coordination between the RRM and CTD is required for effective RNA binding. ORF1p is also believed to have nucleic acid chaperone activity, which may play a role in retrotransposition (Khazina *et al.* 2011).

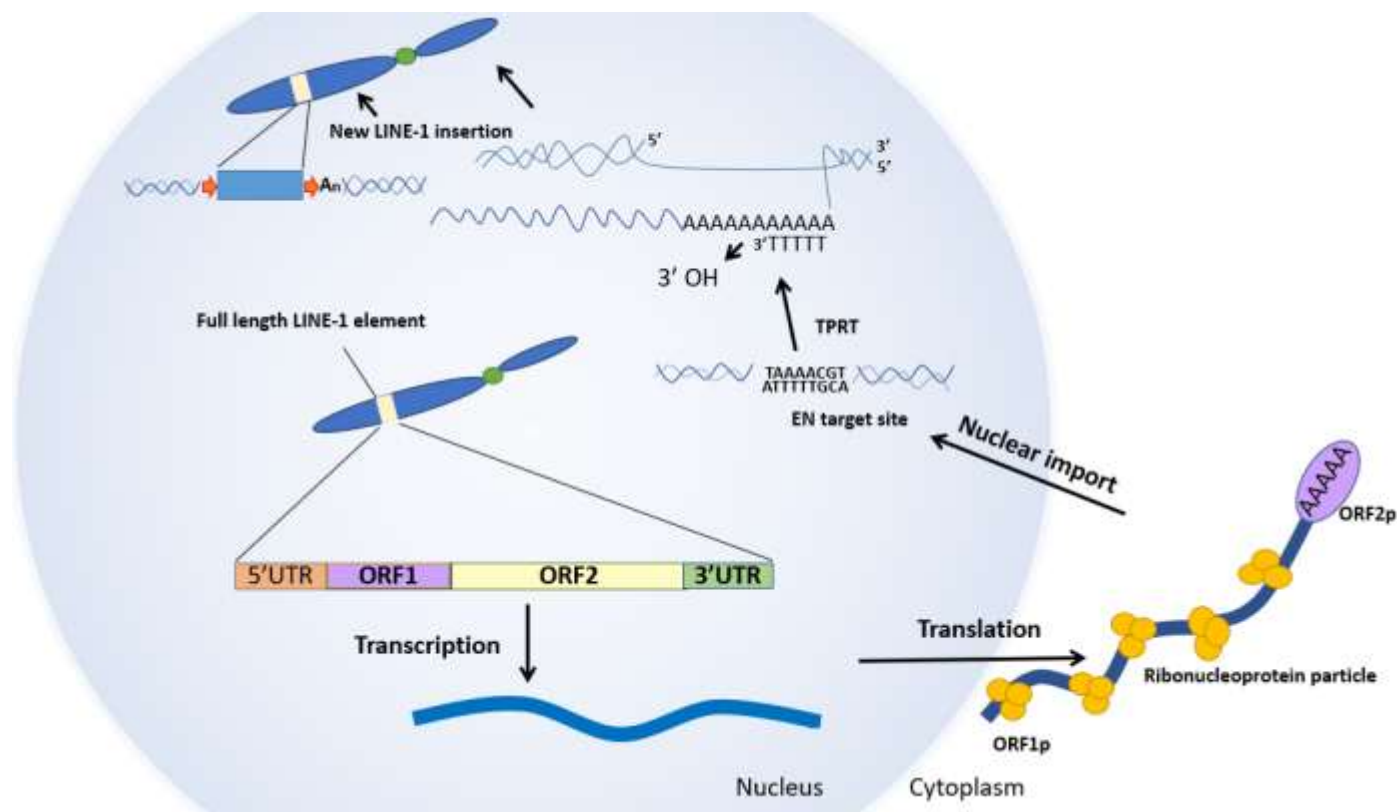
ORF2 encodes a 150kDa protein: ORF2p, which provides the endonuclease and reverse transcriptase activities required for canonical retrotransposition (Clements *et al.* 1998). Biochemical characterisation of the ORF2p protein has revealed that the reverse transcriptase is highly processive in comparison to similar enzymes (Piskareva *et al.* 2006) synthesising approximately 620 nucleotides per binding event using the L1 transcript as a template.

In addition, ORF2p contains a Z-domain and a cysteine rich domain, located close to the RT domain and the C terminus respectively. Both of these domains are of unknown function, though mutational analysis in cultured cell assays have shown that both regions are required for canonical retrotransposition (Christian *et al.* 2017).

At the terminus of an element there is a 3'UTR region, which contains a poly-A tail. Doucet *et al.* (2015) demonstrated that manipulation of the poly A tract in vitro resulted in inhibition of retrotransposition in *cis*. Furthermore, experimentation revealed that the poly A tract allows L1 ORF1p to bind and mobilise RNAs in *trans*.

### 1.2.2 The mechanism of L1 retrotransposition

L1 elements mobilise by a mechanism called target primed reverse transcription (TPRT). This is a “copy and paste mechanism” in contrast to the “cut and paste” mechanism used by DNA transposons. This mechanism is depicted in **Figure 1.4**.



**Figure 1.4. The process of target primed reverse transcription**

(Adapted from Zhang et al 2020)

The first step of TPRT is the transcription of the full-length element, initiated when RNA polymerase II binds to the promoter region in the 5'UTR of L1 s (Beck *et al.* 2011; Hancks and Kazazian 2012; Swergold 1990). The L1 mRNA transcript is exported to the cytoplasm, where ORF1 and ORF2 are translated into their respective protein products, ORF1p and ORF2p.

It is known that ORF1p is translated by a conventional cap-dependent mechanism (Dmitriev *et al.* 2007), however the difficulty in detecting ORF1p has made characterising ORF1p translation challenging. It is thought that ORF1p is expressed as a singularity, not as a fusion protein together with ORF2p, as detection with anti-ORF1p antisera indicates a 40kDa product (Leibold *et al.* 1990; McMillan and Singer 1993; Hohjoh and Singer 1996; Goodier *et al.* 2004; Kulpa and Moran 2005).

Using a cultured cell retrotransposition assay, Alisch *et al.* (2006) showed that ORF1p translation is initiated at the first in-frame AUG codon. Interestingly though, they observed that mutation of the AUG codon to any alternative amino acid does not impair retrotransposition of synthetic and mouse L1 transcripts in HeLa, CHO, or rat neural progenitor cells. This suggests that ORF1p translation can also begin from alternative codons. The same study demonstrated that sequences within ORF1p and the interORF spacer also appear to be dispensable both for retrotransposition and translation of ORF1p. However, mutation of the UAA ORF1p stop codon in a transcript without an interORF spacer abrogated retrotransposition, indicating that the presence of this codon is essential for retrotransposition and providing further evidence that ORF1p and ORF2p need to be encoded separately (Alisch *et al.* 2006). This series of experiments indicates that ORF1p is translated through an unconventional termination re-initiation mechanism that is not dependent on an AUG codon.

Once translated, both ORF1p and ORF2p show a cis-preference for their own transcripts, associating with their encoding mRNA to form a ribonucleoprotein particle (Esnault *et al.* 2000; Kulpa and Moran 2006; Wei *et al.* 2001). This cis preference is critical for ensuring the efficiency of retrotransposition, as it reduces the interaction of L1 machinery with unrelated cellular RNAs and RNAs from “fossilised L1s (Wei *et al.* 2001). Once formed, the ribonucleoprotein particle regains access to the nucleus. The mechanism of nuclear import is not yet known. However, using an adenoviral vector to deliver an L1 element into differentiated primary somatic cells and G1/S phase arrested cells, Kubo *et al.* (2006) demonstrated that the process is not dependent on cell-division.

The endonuclease of ORF1p can now locate and cut the genomic DNA at a specific consensus site, 3'-AA/TTT-5' (Cost and Boeke 1998; Cost *et al.* 2002; Morrish *et al.* 2002; Jurka *et al.* 1997). Often, this sequence is preceded by T-tracts of variable length, indicating that the sequence upstream of the EN consensus site has the potential to base pair with the polyA tail of the element. The endonuclease incision releases a 3' hydroxyl, which can be used as a primer for the reverse transcription of the new

L1 element by the ORF1p reverse transcriptase (Luan *et al.* 1993; Cost *et al.* 2002). The factors determining the genomic location of the new insertion are not fully known, however, in 2013, Monot *et al.* used a direct L1 extension assay to measure the efficiency of the ribonucleic particle to initiate reverse transcription using different primers. Notably, efficient priming was detected with just 4 matching nucleotides at the 3' end. Therefore, the specificity of the reverse transcriptase priming and cleavage endonuclease enzymes are likely to influence the location of new L1 insertions.

Once the L1 cDNA copy is inserted into the genomic DNA, an incision is made in the second DNA strand and a second copy of the L1 insertion is generated. However, the mechanism by which this process takes place is a topic of ongoing research.

Insertions generated by TPRT often contain ~7–20-bp target-site duplications (TSDs; Szak *et al.* 2002). TSDs are generated because the EN cleaves the second strand a variable distance away from the initial nicking site, these staggered gaps are later filled in to form repeat sequences flanking the element. (Szak *et al.* 2002). Approximately 35% of genomic L1 insertions generated by TPRT include the full-length element (Boissinot *et al.* 2000), and the rest are 5' truncated to some degree (Grimaldi *et al.* 1984). The cause for this frequent truncation is not known, but it has been speculated that it's a result of a host defence or DNA repair activity, potentially interrupting the reverse transcription process or partially degrading the transcript (Beck *et al.* 2011).

### 1.2.3 A cultured cell retrotransposition assay to measure retrotransposition

Mobile element activity can be measured by a cultured cell retrotransposition assay using a reporter cassette. This assay was developed by Moran *et al.* (1996) and can be used to quantify activity of many types of mobile elements. Development of the retrotransposition assay was a pivotal moment in L1 biology because it made the “real time” quantification of retrotransposition possible.

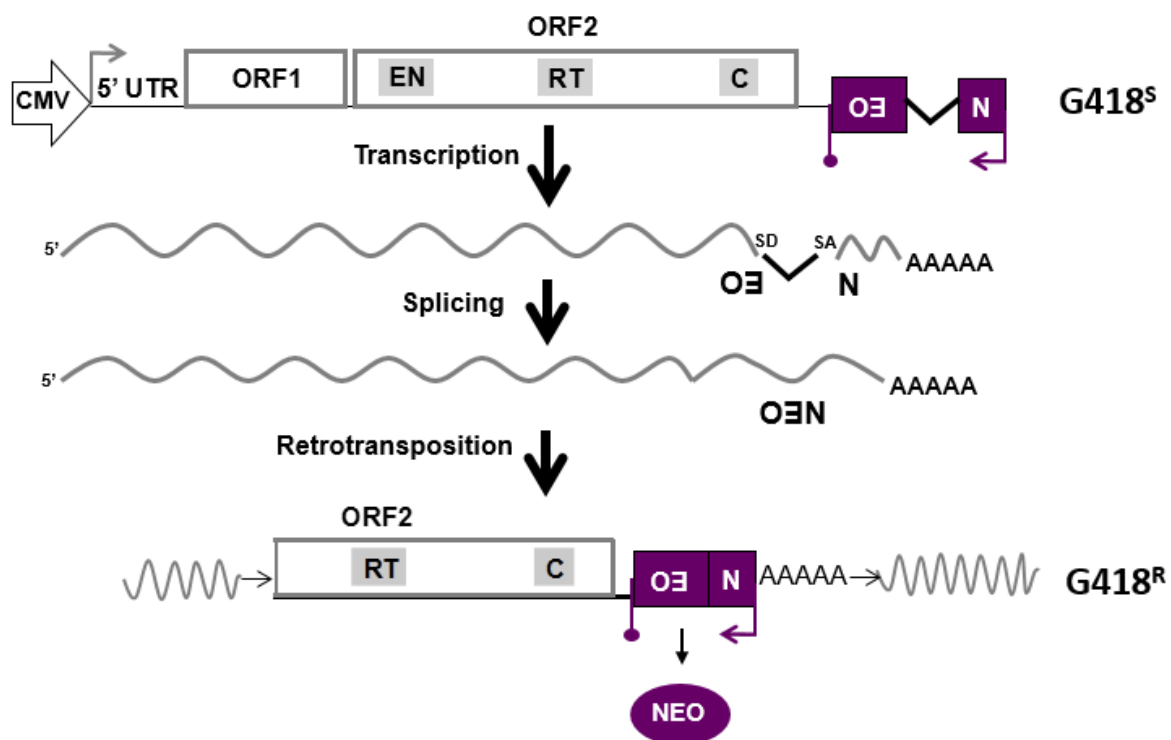
The assay was first used to characterise the activity of two elements, L1.2 and LRE3, which had been identified as the likely source of insertions in Factor VIII (Dombroski *et al.* 1991) and the dystrophin genes (Holmes *et al.* 1994). Expression in yeast systems had shown that these elements contained reverse transcriptase activity and contained two open reading frames (ORF1 and ORF2). Using the cultured cell retrotransposition assay, Moran *et al.* (1996) were able to show that both L1.2 and LRE3 were capable of high frequency autonomous retrotransposition.

To do this, Moran *et al.* (1996) cloned a reporter cassette designed by Freeman *et al.* (1994) into the 3'UTR of the L1.2 element, creating a transcript called *L1.2mneol* (**Figure 1.5**). The reporter gene,



*mneol*, consists of an antisense copy of the *neo* gene, which encodes resistance to the antibiotic G418 and is flanked by its own promoter and polyadenylation signal. Importantly, the antisense *neo* is interrupted by an intron IVS 2 of the  $\gamma$ -globin gene, in the sense orientation. This intron is flanked by a splice donor and splice acceptor sites (also in the sense orientation).

The design of the *L1.2mneol* transcript ensures that the gene product of *neo* will only be expressed when the full length of the cassette is expressed, spliced to remove the intron, and reverse transcribed into the genome (Moran *et al.* 1996). Once reintegrated into the genomic DNA, the *neo* reporter gene can be transcribed from its heterologous promoter and encode G418 resistance. In contrast, transcripts which originate from the *L1.2mneol* promoter can be spliced but will contain an antisense copy of *neo*, and transcripts generated from the *neo*'s own antisense promoter cannot be spliced due to the orientation of the donor and acceptor sites. This ensures that only cells which have harboured a full retrotransposition event are resistant to G418 and will survive a period of selection.



**Figure 1.5. A cultured cell assay to measure L1 retrotransposition**

(Adapted from Moran *et al.* 1996)

Moran *et al.* (1996) cloned *L1.2mneol* into an expression vector called pCEP4, which was chosen to ensure high levels of expression. pCEP4 replicates at a high copy number and contains a hygromycin

gene for the selection of successfully transformed cells. It also contains an origin for bacterial replication (Ori) and a selectable marker for prokaryotic cells. The construct was also designed with the cytomegalovirus (CMV) immediate-early promoter, which drives a high level of expression. In this way, an engineered L1 element in a pCEP4 vector can be transfected into any cell line. After waiting 48 hours after transfection, the cells were treated with a 12-day period of selection using the antibiotic hygromycin. After the period of selection, cells expressing the reporter gene survived and formed colonies, which were fixed and stained, whilst cells which have not will remain sensitive to the antibiotic and die. By modifying the expression vector, different selection methods can be used to adapt the retrotransposition assay to different cell lines.

#### 1.2.4 The formation of a ribonucleoprotein particle is essential for retrotransposition

The ribonucleoprotein particle, comprised of ORF1p, ORF1p and the encoding transcripts of each protein, is an important intermediate in target primed reverse transcription. In fact, there is a body of evidence suggesting that the formation of the particle is essential for retrotransposition (Kulpa and Moran 2005).

In 1990, working with pluripotent embryonic carcinoma cells, Deragon *et al.* identified that reverse transcriptase activity was associated with a macromolecular complex, which had a protein component of 37kDa and was detected in a 160,000 G pellet after differential centrifugation. Detection of L1 mRNA within the complex suggested that this particle was an intermediate of retrotransposition. This is thought to be the first recorded detection of the RNP.

Soon after, ORF1p and L1 mRNA were detected in other cell lines with high levels of endogenous L1, including F9 teratocarcinoma cells (Martin *et al.* 1991) and 2102EP embryonic carcinoma cells (Hohjoh and Singer, 1996). Together, these studies consolidated the theory of an RNP TPRT intermediate, however they did not distinguish between the RNP of a retrotransposition competent element or a retrotransposition defective element. To characterise a functional RNP, Kulpa and Moran (2005) engineered an epitope tagged ORF1p element and used cultured cell retrotransposition assays to observe the behaviour of mutant and wild type ORF1p.

Using differential centrifugation in extracts of HeLa cells, Kulpa and Moran (2005) confirmed that the epitope tagged ORF1p (etORF1p) localised to a cytoplasmic RNP and appeared to co-migrate with L1 RNA. They observed that catalytic mutations in the RT and EN domains of ORF1p did not appear to

affect localisation of etORF1p to the RNP. Furthermore, introducing a frameshift induced stop mutation into the amino terminus of ORF1p, which has been shown to prevent ORF2 synthesis, also did nothing to disrupt localisation of etORF1p to the RNP. ORF1p therefore appears to be dispensable for the formation of the RNP.

Kulpa and Moran also tested the impact of a series of ORF1p mutants on formation of the RNP with etORF1p. Three mutations in highly conserved amino acids in the carboxyl domain of ORF1p (REKG235-238, RR261-262 and YPAKLS282-287) have been previously shown to abrogate retrotransposition in a cultured cell assay. These mutations also appeared to obstruct the ability of etORF1p to localise to the RNP, as the tagged construct was no longer present in the 160 000 G sedimentation pellet. This data suggests that mutations in the carboxyl terminus of ORF1p disrupt interactions between ORF1p and L1 RNA, preventing the formation of the RNP.

Kulpa and Moran also tested a series of mutations altering a di-arginine motif of ORF1p. Mutation of both arginine residues to lysine resulted in the reduction of retrotransposition to less than 1% of the wild type level, as did a R262K mutation. Interestingly, each of these mutant constructs were detected in the 160 000 G pellet, suggesting that the RNP had successfully formed. The identification of the mutant constructs in the RNP pellet despite the abrogation of retrotransposition indicates that formation of the RNP, though necessary, is not sufficient for L1 retrotransposition.

These findings have advanced our understanding of the role of the ribonucleoprotein particle in TPRT and suggest that ORF1p may play an additional role downstream in the pathway. What this role is and the mechanistic impacts of the described ORF1p mutations remains to be elucidated.

### 1.2.5 The structure and function of ORF1p

Multiple studies have shown that mutations in ORF1p are detrimental to retrotransposition and can restrict retrotransposition with the same severity as mutations in the much more highly conserved ORF1p reverse transcriptase (Martin *et al.* 2005; Kulpa and Moran 2005; Martin *et al.* 2006).

The first coding sequence for ORF1p was identified from a mouse element called L1Md-A2 (Loeb *et al.* 1986). Since then, the ORF1p proteins of many different L1 elements (from various mammals and related elements in fish) have been analysed. Mammalian L1ORF1p consists of three domains, a coiled-coil domain, which allows the protein to form homo-trimers in a dumbbell shaped configuration, an RRM (RNA recognition motif) and a basic, conserved, C-terminal domain, which is thought to be involved in nucleic acid binding and chaperone activities (Martin *et al.* 2003; Martin *et*

*al.* 2006). In human L1 elements, the coiled-coil domain contains a leucine zipper (L1 elements from different organisms have poor sequence similarity in this region) and is preceded by a 51 residue-long N terminal region. There is also a 15-residue tail at the C terminus of L1ORF1p that can be truncated without impacting retrotransposition.

In the previously described study by Kulpa and Moran (2005), mutating two conserved motifs in the carboxyl domain abrogated efficient retrotransposition and prevented the association of ORF1p with the RNP. It is thought that these mutations disrupted the ability of the protein to bind back to its encoding transcript in order to form this critical retrotransposition intermediate. NMR structures of the CTD domain of ORF1p have shown that the protein appears to have three discrete arrangements, resting, lifting and twisting (Januszyk *et al.* 2007). Theoretically, the lifting motion would open up the RRM cleft, facilitating nucleic acid binding. In the twisting configuration, the nucleic acid would be locked into place. In the resting position, basic patches of the coiled coil domain would be exposed for further potential nucleic acid binding (Januszyk *et al.* 2007). In this way, the flexible structure of the protein is likely related to nucleic acid chaperone activity. It is not known precisely how mutating the AA and REKG motif in the carboxyl domain disrupts this structure and the binding of the template. However, structure based mutational analysis that combined RNA binding assays with cultured cell retrotransposition assays revealed that this structured flexibility is crucial for retrotransposition (Januszyk *et al.* 2007).

L1ORF1p also shows a strong resemblance to other coiled coil proteins, including the SNARE proteins found in eukaryotes (Sutton *et al.* 1998; Khazina and Weichenrieder 2018) the bacterial protein M1 (McNamara *et al.* 2008) and the influenza hemagglutinin protein (Chen *et al.* 1999). All of these proteins form homotypic trimers and have functions relating to membrane fusion. Furthermore, Horn *et al.* (2017) have shown that an interaction between L1ORF1p and the ESCRT membrane budding complex, which is generally involved in cellular trafficking and membrane fusion, is required for retrotransposition. It has therefore been speculated that L1ORF1p could have a function related to membrane fusion in an uncharacterised part of the L1 retrotransposition cycle (Khazina and Weichenrieder 2018).

X-ray crystallography of the ORF1p trimer has shown that the coiled coil domain mediates L1ORF1p trimerisation and enables the flexible attachment of the RRM and CTD domains (Khazina and Weichenrieder, 2009). Combining biophysical and cell-based techniques, Khazina and Weichenrieder (2009) determined the physical structure for the entire coiled coil domain of ORF1p. Coiled coil domains are made up of bundles of alpha helices, each helix constructed from seven amino acids. The centre of the helical bundle is usually made up of hydrophobic residues, leading to a hydrophobic core

which is important for the stability of the structure. Interestingly, the N-terminal heptads of the coiled coil domain appear to exhibit metastability, needing to switch between partially and fully structured states. The authors propose that this feature could explain the strong sensitivity of the protein to mutation within these domains, despite being less evolutionarily conserved than ORF1p.

#### 1.2.6 ORF1p has been difficult to characterise experimentally

In contrast to ORF1p, the expression of ORF1p has been poorly characterised in human cell lines and tissues. It is believed that ORF1p is expressed at a much lower level than ORF1p, with as little as one molecule of ORF1p generated per L1 RNA transcript, and an estimated ratio of 30 ORF1p molecules for every ORF1p molecule (Wei *et al.* 2001). In fact, the RNP retrotransposition intermediate is believed to contain many ORF1p molecules, just one ORF1p molecule and one L1 mRNA. A study investigating the expression and cellular localisation of ORF1p and ORF1p detected ORF1p in 97% of cells, whereas ORF1p was detected in approximately 10% of cells (Mita *et al.* 2018).

As such, the detection of ORF1p has been highly challenging in both an *in vitro* and an *in vivo* context (Dai *et al.* 2014), whilst ORF1p has been relatively easy to detect by western blotting (both with antibodies directed both to the endogenous protein and to proteins with epitope tags, including TAP, T7 and Flag tags, immunoprecipitation, mass-spectrometry and T7 RNA polymerase expression assays (Bratthauer *et al.* 1994; Rodic *et al.* 2014; An *et al.* 2011)).

Attempts to detect ORF1p by western blot have had mixed success. Kirilyuk *et al.* (2008) successfully detected ORF1p in rat chloroleukemia (RCL) cells after generating monoclonal antibodies directed to recombinant ORF1p and ORF1p. Detection of a 150 kDa band with weak signal intensity appeared consistent with the relatively low expression of ORF1p in comparison with ORF1, potentially due to the unconventional translation, re-initiation of ORF1p transcription (Alisch *et al.* 2006). However, many efforts have been fruitless, and the difficulty in isolating ORF1p has made identification of proteomic interactors and regulators challenging (Dai *et al.* 2014).

#### 1.2.7 The enzymatic activity of ORF1p in TRPT

ORF1p contains the enzymatic machinery required for target primed reverse transcription. It is believed to function as one contiguous protein (Alisch *et al.* 2006). An apurinic/apyrimidic (AP) endonuclease, first identified by Feng *et al.* (1996), is located at the N terminus of the protein. The group identified several conserved AP endonuclease motifs and demonstrated that the domain could convert supercoiled DNA into open circle, and preferentially cleaved sites similar to L1 target sites in vitro. When 2.6ng of L1 EN was incubated with 0.2ug of supercoiled DNA, 50% of it was converted into open circle DNA.

Adjacent to the EN domain is a Z domain (Kines *et al.* 2014). This, and the sequence at the C terminus of the EN domain, are thought to modulate the activity of the EN domain during retrotransposition, as the addition of these sequences to the EN domain reduced EN-induced DNA damage in vitro (Kines *et al.* 2014). Further experimentation will be needed to see how this sequence between the EN domain and the RT domain affects retrotransposition and whether cellular factors may be interacting with the element via this sequence.

The EN activity of ORF1p is physically coupled to its reverse transcriptase activity, which has also been shown to be required for retrotransposition (Moran *et al.* 1996). Unlike retroviral RTs, the ORF1p reverse transcriptase has been shown to be highly processive, producing a transcript up to 5 times longer than that of the Moloney murine leukaemia virus when tested in processive, in vitro conditions (Piskareva *et al.* 2006).

### 1.2.8 Expression of ORF1p is linked to cellular toxicity.

Expression of L1 has been linked to cellular toxicity through a number of mechanisms (Gasior *et al.* 2006; Wallace *et al.* 2008; Belgnaoui *et al.* 2006). This toxicity appears to be primarily associated with the expression of L1 proteins, with Gasior *et al.* (2006) estimating that L1 expression generates 10-fold more DSBs than insertions. Expression of the element results in an increase in  $\gamma$ -H2AX foci (a histone marker that indicates ionising radiation and DSBs), and a COMET assay in HeLa cells showed that expression of L1 resulted in a direct increase in DSBs (Gasior *et al.* 2006). In the same study, expression of L1 (but not empty vector, or EN- transcripts) also lead to cell cycle arrest at G2, reducing the number of cells which enter mitosis.

Gasior *et al.* (2006) observed that abrogation of the EN domain of ORF1p reduced the cellular toxicity, while mutation of the RT domain had no notable effect. This indicates that the ORF1p endonuclease is a driver of L1 induced toxicity, most likely through the generation of nicks in the genomic DNA, as

opposed to a deleterious effect from the generation of cDNA intermediates by the RT domain. More recent experiments, however, have suggested that the RT domain can also induce a substantial number of DSBs, though to a lesser degree than the EN domain (Wallace *et al.* 2008)

As expected, these effects can be fatal to cells. DNA damage induced by expression of the L1 element was also observed to induce apoptosis in cancer cells (Belgnaoui *et al.* 2006). Interestingly, the same study demonstrated that L1 induced DSBs are usually repaired 48 hours post transfection leading to speculation that DNA repair proteins are upregulated as part of the L1 lifecycle.

Interestingly, the EN domain of ORF1p has been observed to tolerate multiple mutations before its toxicity is mitigated. Kines *et al.* (2016) analysed a full length L1 in order to identify naturally occurring mutations from the EN domain. A range of mutations were observed, some of which were believed to be structurally important and others which were thought to act as sites for phosphorylation by different kinases. A subset of these mutations was tested for their ability to impact endonuclease activity, induce H2AX phosphorylation and drive Alu retrotransposition. Surprisingly, almost all of the mutated endonucleases were still capable of inducing DNA damage. Seemingly, many of these putative phosphorylation sites are not necessarily related to the capacity of the ORF1p EN to generate DNA damage, or to drive retrotransposition of Alu elements.

#### 1.2.9 Identification of a PCNA interaction protein box in ORF1p

Proliferating Cell Nuclear Antigen (PCNA), a sliding clamp that improves the processivity of DNA polymerase, has been identified as a high specificity interactor of ORF1p (Taylor *et al.* 2013; Mita *et al.* 2018). In 2013, Taylor *et al.* identified a canonical PCNA-interaction protein (PIP) box, located between the EN and RT domains of ORF1p, from residues 407-415. The group identified four amino acids I407, I411, Y414, and Y415, which were highly conserved across species. Pull-down experiments revealed that mutation of some of these residues to alanine disrupted the co-precipitation of PCNA with ORF1p.

Cultured cell retrotransposition assays revealed that mutations disrupting the ORF1p/PCNA interaction also led to reduction in retrotransposition. Mutation of a less conserved residue in the PIP box, Q408A, did not prevent the ORF1p/PCNA interaction, and did not lead to the subsequent reduction in retrotransposition. Reducing the expression of endogenous PCNA using short hairpin RNAs also led to a reduction in retrotransposition. Together, these data suggest that interaction of ORF1p with the PIP domain is required for retrotransposition

These studies validate a role for PCNA in the lifecycle of L1, likely during or following TPRT (Taylor *et al.* 2013). The precise nature of this role, however, is currently under investigation.

### 1.3 The endonuclease independent pathway of retrotransposition

As well as TPRT, a second pathway of L1 retrotransposition has been recently described. Using a retrotransposition assay in CHO-K1 cells, Morrish *et al.* (2002), noted that two constructs with mutated residues in the endonuclease site (Asp205Ala and His230Ala) were still capable of retrotransposition, at 19% and 5% of wild type levels respectively. In HeLa cells, these mutations had been observed to reduce retrotransposition levels to <5% of wild type levels, likely due to the abolishment of endonuclease activity. This mechanism of retrotransposition was termed “endonuclease-independent retrotransposition” (ENi) as it could take place without a functional EN domain.

Because the L1 lifecycle involves ORF1p and ORF2p showing a cis-preference for their own mRNA, Morrish *et al.* (2002) hypothesised that it is unlikely that the proteins are being complemented in trans to complete TPRT. Potentially, the elements were exploiting breaks that were already present in the DNA, instead of creating new DNA incisions using the EN domain.

This theory aligned with a previous study by Teng *et al.* (1996) in yeast *Saccharomyces cerevisiae*, which had shown that DSBs could be fixed by insertions in the break site by a functional reverse transcriptase (either of the yeast Ty1 element, the human L1, or *Crithidia* CRE1). Similarly, Voliva *et al.* (1984) had previously suggested that L1 could disperse through the genome through the “patching” of broken chromosomes. This mechanism would ensure that the L1 element was useful to its host and would therefore undergo positive selection.

To further understand the mechanism of ENi retrotransposition, Morrish *et al.* (2002) tested whether proteins from the non-homologous end joining (NHEJ) pathway of DNA repair were required to generate an ENi insertion. Using cultured cell retrotransposition assays, the group investigated L1 retrotransposition in XR-1 cells, a Chinese hamster ovary cell line mutated in XRCC4 and lacking ligase IV activity. Li *et al.* (1995) had previously shown that this cell line has impaired double stranded break repair and is also unable to perform normal V(D)J recombination.

The group observed that both EN mutant transcripts (Asp205Ala/L1.3 and His230Ala/L1.3) showed increased levels of retrotransposition in comparison with their activity in wild type cells (89.2% and 52.1% of wild type L1.3 levels, respectively). The EN mutants were also tested in hamster cells lacking



DNA-PKcs activity due to a mutation in the DNA-PKcs gene. Again, the EN mutants showed increased levels of retrotransposition in V3 cells. In the parental cell lines, ENi retrotransposition only took place at low levels. The group also observed that when a portion of chromosome 5 was present containing a functional copy of XRCC4, ENi retrotransposition was reduced. Interestingly, assays in *irs1SF* cells, which are unable to perform homologous recombination due to a mutation in XRCC3, did not show an increase in ENi retrotransposition, indicating that this pathway is specifically triggered by defects in the NHEJ pathway, rather than recombination pathways in general.

Structural analysis of insertions generated by wild type L1.3 in XR-1 cells indicated that they were derived from canonical TPRT. ENi insertions, however, were structurally distinct. Firstly, the insertion sites were not the canonical L1 consensus site. In addition, they lacked the target site duplications characteristic of L1 insertions and exhibited deletions at the target site ranging from 11bp and to 1.5kb. Though the insertions were truncated at the 5' end, which is conventional, they lacked a poly A tail and were 3' truncated. Two of the insertions analysed also contained additional nucleotides at the 5' end, and two contained additional nucleotides at the 3' end. Comparisons with EN insertions from V3 cell lines showed similar structural abnormalities, indicating that the mechanism of ENi retrotransposition is likely similar for both cell lines.

When analysing insertions from the parental, NHEJ competent, cell line, the group once again found a small number of insertions with 3' and 5' truncations and lacking TSDs. This indicates that the mechanism of ENi retrotransposition can take place in NHEJ competent cells, though the rate of retrotransposition may be lower.

As reports are primarily based on in vitro studies, the frequency at which ENi retrotransposition is happening in vivo is currently unknown (though one could speculate that genomic ENi insertions would be identifiable due to their structural hallmarks). Furthermore, as these experiments were performed in hamster cells, it is currently unclear whether the results could be extrapolated to humans, due to differences in DNA repair pathways between the two organisms. Priestly *et al.* (1998), have previously shown that NHEJ in rodent cells requires 50-fold less concentration of protein than in human cells. It could therefore be expected that the rate of ENi would be lower in human cells, as the L1 element has more competition from DNA repair proteins.

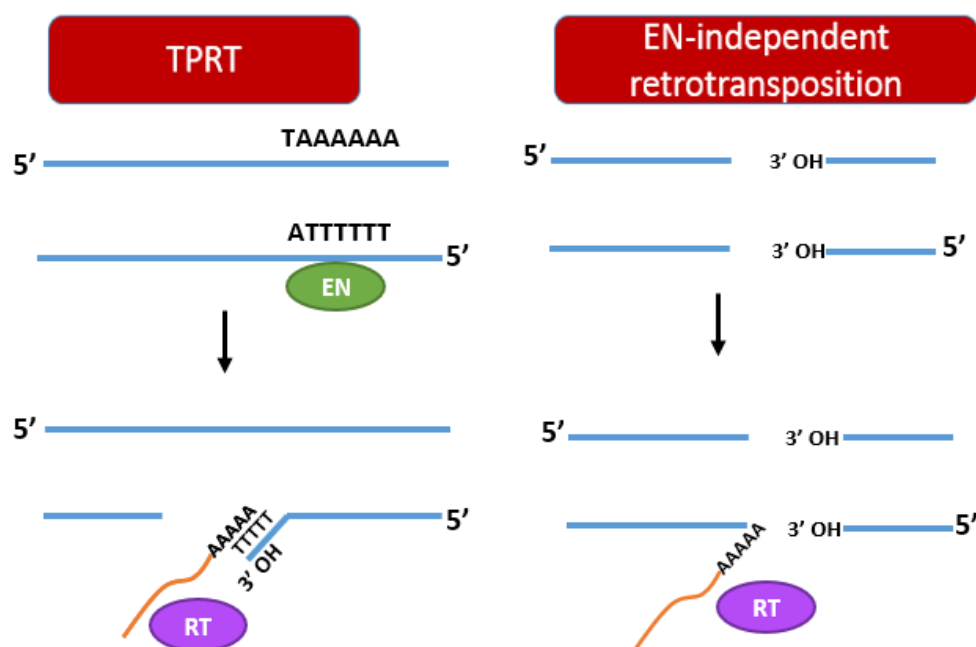
Similar to previous speculation about L1 elements "patching" up breaks in chromosomes, the authors propose that the EN deficient element is exploiting breaks in the DNA which have been left unresolved due to a lack of functional DNA repair. Breaks that arose during the G1 or early S phase of the cell cycle would leave an open 3' hydroxyl which could be exploited as a primer by the reverse transcriptase of

the L1 element. Priming could potentially take place from an internal site within the L1 RNA, without necessarily requiring micro-complementarity between the target DNA and the L1 cDNA.

In line with this data, a further study by Morrish *et al.* (2007) demonstrated that ENi retrotransposition takes place at mammalian telomeres. These studies were performed in V3 cells, which, as well as a deficiency in DNA-Pkcs, contain dysfunctional telomeres. The authors noted that approximately 30% of ENi retrotransposition events take place next to certain sequence adjacent to a telomere repeat (5'-TTAGGG-3'). The authors therefore propose that as well as double stranded DNA breaks, L1 elements can exploit dysfunctional telomeres as a substrate for ENi retrotransposition. Expression of a dominant negative TERF2 protein, which is reported to disrupt telomere capping, allowed the detection of ENi retrotransposition events, indicating that it is the deficiency in telomere function which is allowing the abnormal retrotransposition.

The authors also investigated whether ENi retrotransposition could take place at functional telomeres. Though they identified some faint bands by southern blotting, sequencing suggested that these were nonspecific products and suggest that wild type L1 events seem to initiate by conventional TPRT.

A model depicting ENi retrotransposition is shown in **Figure 1.6**.



**Figure 1.6. The mechanism of endonuclease independent retrotransposition**

## 1.4 L1 activity is regulated by host factors in the human genome.

Because of the various ways in which L1 elements can disrupt gene function, L1 retrotransposition is regulated by host factors at several stages of its lifecycle (Goodier and Kazazian 2008). Importantly, L1 retrotransposons are able to mobilise in the germline, which is how insertions are transmitted from generation to generation (Cordaux and Batzer 2009). This germline retrotransposition also means that L1 elements have the potential to influence the evolution of the mammalian genome (Cordaux and Batzer 2009). L1 elements have a range of expression during the germline cycle, and are subject to regulation by host factors both pre transcriptionally and post transcriptionally (Wang 2017).

### 1.4.1 Transcriptional regulation of LINE- 1

The most effective regulation of L1 is transcriptional regulation. Several mechanisms to transcriptionally repress L1 retrotransposition have been identified in germline cells. These include DNA methylation (Kaneda *et al.* 2004), histone modification (Day *et al.* 2010) and the fetal Piwi-interacting RNA (piRNA) pathway (Aravin *et al.* 2007).

During development, there are two stages where CpG methylation becomes removed from L1: in embryos (prior to implantation) and in migrating primordial germ cells (Messerschmidt *et al.* 2014). Otherwise, this methylation contributes to the repression of L1 elements. In humans there are three primary DNA methyltransferases: DNMT1, DNMT3A, and DNMT3B (in mice there is a fourth called DNMT3C) (Lyko *et al.* 2018). DNMT1 is responsible for maintenance of methylation, whereas DNMT3A and DNMT3B are responsible for de novo methylation in animal development (Lyko *et al.* 2018). Studies have shown that both DNMT3A and DNMT3B are required for de novo methylation of L1 elements (Kaneda *et al.* 2004; Kato *et al.* 2007). DNMT3L, which is a homologue of DNMT3A/B and is expressed in oocytes in females and spermatogonia in males, may also play an important role in the methylation of repetitive elements (Bourc'his *et al.* 2001; Hata *et al.* 2002; Bourc'his and Bestor 2004). In fact, deletion of DNMT3L results in the loss of de novo methylation at L1 elements (Bourc'his and Bestor 2004). In males, by birth, most full length L1 elements are re-methylated. It is thought that L1 elements remain hypermethylated throughout the onset of meiosis (Oakes *et al.* 2007).

The piRNA pathway is important for the transcriptional regulation of L1 elements during epigenetic reprogramming (Pezic *et al.* 2014). Though the pathway was originally identified in drosophila (Aravin *et al.* 2001; Vagin *et al.* 2006), subsequent research has shown that they are active in a number of species, including humans. The PIWI pathway is made up of two central components, the PIWI proteins and the PIWI interacting RNAs (piRNAs), which are small non-coding RNAs which are highly expressed in the germline (Aravin *et al.* 2006; Aravin *et al.* 2008). The PIWI proteins bind these piRNAs to form piRNA ribonucleoprotein (piRNP) or piRNA-induced silencing complexes (piRISCs)(Aravin *et al.* 2006). The pathway works through RNA interference, similar to the microRNA pathway, and targets and represses mRNA transcripts, including those of retrotransposons. (Aravin *et al.* 2008; Sijen and Plasterk 2003; Ernst *et al.* 2017)

The fetal piRNA pathway is a related pathway which functions primarily in fetus cells (Williams *et al.* 2015). It mediates transcriptional silencing primarily through methylation, though experimentation in mouse germ cells has shown that the pathway can also mediate repressive histone modifications (Williams *et al.* 2015). Deletion of Mili and Miwi 2, which are key components of the pathway, appears to reduce methylation of transposable elements but not of global methylation patterns, demonstrating a degree of specificity towards repetitive elements (Aravin *et al.* 2007; Nerkirk *et al.* 2017).

In mouse ES cells, histone methylation appears to play an important role in regulation of L1 elements. Day *et al.* (2010) observed through high throughput sequencing analysis that the histone marks H3K9me3 and H3K27me3 were associated with repression of retrotransposition. Similarly, experimentation in mouse ES cells has revealed that SETDB1 and SUV39H1/SUV39H2, which mediate the trimethylation of H3K9, are involved in the repression of L1 elements (Bulut-Karslioglu *et al.* 2014). A study by Garcia Perez *et al.* (2010) investigated the silencing of engineered L1 elements in human embryonic carcinoma cells. The group showed that treating the cells with histone deacetylases reversed L1 silencing, indicating that regulation was taking place at the chromatin level. Interestingly, promoting differentiation in this cell line inhibits the silencing of reporter genes, indicating that mechanisms of silencing are likely to be different in pluripotent and differentiated cells.

Two histone marks in particular are believed to mediate L1 repression, H2A/H4R3me2 and H3K9me2. Kim *et al.* (2014) showed that in primordial germ cells, the symmetric methylation of arginine 3 of histones H2A and H4 is important for repression of retrotransposition. These modifications are

mediated by protein arginine methyltransferase 5 (PRMT5). Kim *et al.* (2014) show that this marker is dominant during the global demethylation, a time when the genome is most vulnerable to retrotransposition activity.

Finally, L1 elements can also be silenced through the activity of zinc finger proteins, a diverse family of transcription factors that can bind target sequences through KRAB zinc finger binding domain. The most prominent subgroup is the KRAB-zinc finger proteins (KRAB-ZFPs), which are distinguished by the Kruppel-associated box (KRAB) domain (Urrutia *et al.* 2003). KRAB-ZFPs have been found to repress both endogenous and exogenous retroviruses in both human and mouse ESCs (Rowe *et al.* 2010). KRAB-ZFPs bind to specific retrotransposition sequences, and recruit KAP1 which facilitates the repressive H3K9me3 histone modification at these sites by SETDB1 (Friedman *et al.* 1996; Rowe *et al.* 2010). Because of the high specificity required for KRAB-ZFPs to bind L1, certain L1 subfamilies have been found to be more vulnerable to KAP-mediated repression than others. As the sequence of the element has evolved over time to create new sub types, modifications in the KRAB-ZFP binding sites have allowed some younger families to “escape” this regulation (Jacobs *et al.* 2014; Castro-Diaz *et al.* 2014).

rfo1.5 DNA repair in the human genome.

The stability and integrity of DNA are vitally important to life. DNA is subject to insult both from environmental and internal metabolic processes. UV radiation and tobacco smoke are known causes of DNA damage, as are reactive oxygen species and free radicals generated through cellular metabolism (Reviewed in Huang and Zhou 2021). Even DNA replication during cell division is an error prone process. Through these factors, errors can arise either in the form of DNA damage or mutation. Mutation is a change in the base sequence of the DNA, whereas DNA damage is physical abnormality in the DNA that can be detected by an enzyme. It is estimated that DNA suffers tens of thousands of lesions every day.

The eukaryotic cell cycle includes a number of checkpoints, at which point the cell uses specified internal cues to determine whether to proceed with division. A variety of DNA repair processes are constantly active within cells to maintain genomic integrity in the face of these disruptions. A reduction of DNA repair in mammalian genomes is associated with accelerated aging and disease (Reviewed in Huang and Zhou 2021). Indeed, deficiency in specific DNA repair pathways is a common causative factor in various types of cancer.

Due to the universal importance of genome integrity, many DNA repair processes are highly conserved in eukaryotic cells. The mechanism of DNA repair used depends on the type of lesion. Pyrimidine dimers, for example, caused by UV radiation, are repaired by a process called nucleotide excision repair, which involves almost 30 proteins (Kemp *et al.* 2012; Lloyd *et al.* 2005). Base excision repair (BER), on the other hand, targets damage caused by free radicals generated during metabolism. A further pathway, mismatch repair, identifies and resolves incorrect bases which have been misincorporated by the DNA polymerase (Stojic *et al.* 2004).

Double stranded breaks, which are highly deleterious and can result from ionising radiation, are resolved either using non-homologous end joining (NHEJ) or homologous recombination repair (HRR; Hefferin *et al.* 2005). HRR is used when there is a template strand available to the cellular machinery in order to identify the sequence, whereas NHEJ is used when no such template is available, and is a more error prone form of DNA repair (Hefferin *et al.* 2005).

Several DNA repair pathways are interconnected, and proteins may feature in different DNA damage response pathways or facilitate cross talk allowing the coordination of a specialised response (Huang and Zhou 2021). Extensive communication also exists between the DNA damage response and RNA processing pathways, DNA proliferation pathways and cell death. Disruption of this communication can lead to genomic instability and tumorigenesis. It is perhaps not surprising, then, that several DNA repair proteins have also been identified as potential regulators of L1 retrotransposition.

### 1.5.1 DNA repair proteins regulate retrotransposition.

The signalling cascades linked to DNA repair are complex, and as such, L1 elements can be impacted by a number of cellular events that take place in relation to DNA damage. This thesis is primarily concerned by the relationship between the Fanconi Anaemia DNA repair pathway and L1 retrotransposition, however specific proteins from a number of different DNA repair pathways have been directly linked to the regulation of retrotransposition. This section outlines the research which provides a precedence for the regulation of L1 by DNA repair proteins.

#### 1.5.1.1. Nucleotide excision repair (NER) pathway

The Nucleotide Excision Repair (NER) pathway is a key repair pathway which works to process 3' flaps and to remove the bulky adducts caused by exposure to UV light. Several proteins from this pathway

have been found to influence L1 retrotransposition. In cultured cell assays in CHO UV20s cells expression of ERCC1/XPF, a key component of NER, led to a reduction of retrotransposition (Gasior and Deininger 2008), indicating that the protein may serve as a negative regulator. Separately, Servant *et al.* (2017) demonstrated that XPD and XPA, two other core proteins in the NER pathway, also have the capacity to limit retrotransposition in cultured cell assays. Retrotransposition was elevated in cells mutant for the two proteins, and complementation with the wild type proteins appeared to reduce retrotransposition to wild type levels. A similar effect was seen with XPC, an NER protein involved in the binding of DNA lesions (Servant 2017).

#### 1.5.1.1. Non-homologous end joining pathway

Non-homologous end joining (NHEJ) is a mechanism of DNA repair which resolves double stranded breaks (DSBS) in DNA (Hefferin *et al.* 2005). In mammalian cells, NHEJ is the primary pathway for the repair of these lesions. Unlike, homologous recombination, NHEJ does not require a homologous sequence to direct the DNA repair. Recently, Suzuki *et al.* (2009) examined the retrotransposition frequencies of two types of LINE, the zebrafish ZfL2-2 and human L1 in a series of mutant DT40 cells. Mutant cell lines tested included Ku70, Artemis, LigIV, all of which are proteins involved in the NHEJ pathway. Retrotransposition levels of both the zebrafish and L1 element were reduced in all of the NHEJ mutant cells in comparison to the wild type control.

#### 1.5.1.2 BRCA1 identified as a L1 inhibitor

BRCA1 encodes the Breast cancer type 1 susceptibility protein, which is involved in a number of DNA repair processes including homologous recombination and the Fanconi Anaemia DNA repair pathway (Garcia-Higuera *et al.* 2001)

In 2020, Mita *et al.* undertook a global siRNA screen in HeLa-M2 cells to identify cellular factors regulating retrotransposition. The study, which made use of an siRNA, image based retrotransposition assay, identified 790 “inhibitors”, of which depletion increased retrotransposition levels. The group also identified 1,133 “supporters” of which depletion reduced retrotransposition. Gene ontology (GO) term analysis of the inhibitors revealed that the gene function was enriched for DNA repair proteins. A CRISPR screen by Liu *et al.* (2017) in human chronic myeloid leukaemia K562 and HeLa cells using an L1-G418R retrotransposition reporter identified several overlapping genes with the siRNA screen by Mita *et al.* that were involved in DNA repair processes.

After performing secondary validation of the DNA repair genes, the group observed a strong inhibitory effect of BRCA1. BRCA1 helps to protect replication forks from degradation by nucleases (Daza-Martin *et al.* 2019) and is often found at stalled replication forks. This finding lead Mita *et al.* 2020 to hypothesise about a “battleground” at replication forks between HR and L1 retrotransposons. In such instances, HR and FANCC components may be functionally competing with L1 retrotransposition. Activities of the HR proteins, such as formation of complexes for DNA end resection, could form physical barriers to TRPT (Mita *et al.* 2020). Furthermore, both pathways are active during the S/G2 phase when retrotransposition would be taking place (Mita *et al.* 2018). Because of the complex and as yet unclear nature of the relationship between DNA repair and L1 retrotransposition, such mechanistic insights into these interactions are intriguing.

Importantly, Mita *et al.* (2020) and several other studies also identified several proteins from the Fanconi Anaemia pathway as potential L1 regulators. The Fanconi Anaemia pathway and its role in the regulation of L1 retrotransposition is a key theme of this these, and the next section will go on to discuss the pathway in detail.

## 1.6 The Fanconi Anaemia pathway of DNA repair

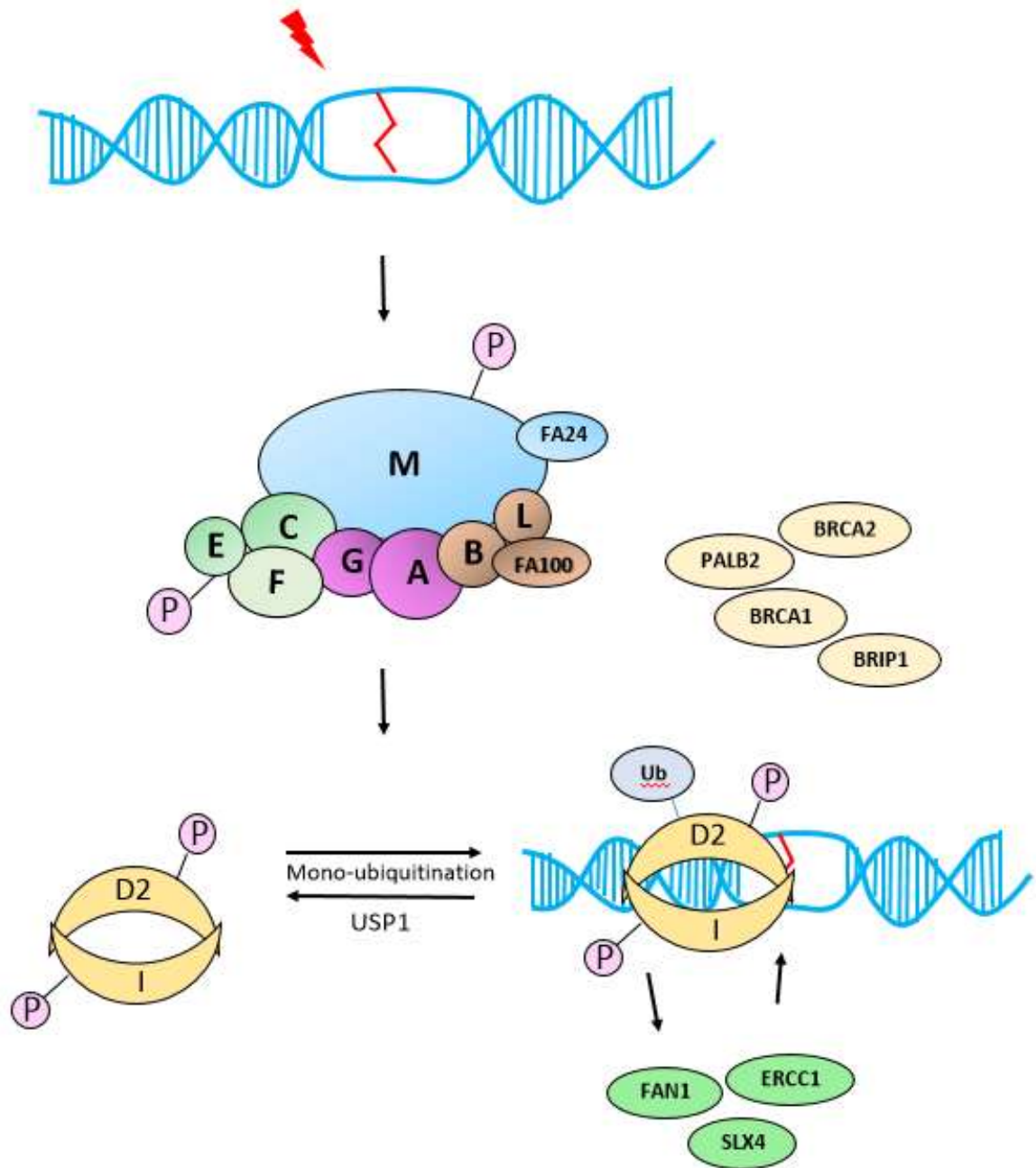
The Fanconi Anaemia pathway is a DNA repair pathway involved in the resolution of DNA interstrand cross links (ICLS). The pathway is made up of 22 genes, and mutation in any one of these (with the exception of the X chromosomal FANCB) is associated with the rare genetic disorder Fanconi Anaemia (Reviewed in Ceccaldi *et al.* 2016). Fanconi Anaemia is autosomal recessive, and affects approximately 1 in every 1000 people. Hallmarks include bone marrow failure, congenital abnormalities such as skeletal defects and a predisposition to cancer (Ceccaldi *et al.* 2016).

ICLs are highly toxic DNA lesions which can arise following exposure to environmental mutagens. They are created when an irreversible covalent linkage forms between bases on the two strands of DNA (Lopez-Martinez *et al.* 2016). This linkage prevents the separation of DNA strands which is required for replication and transcription and can lead to abnormalities such as R loops and stalled replication forks. The FANCC pathway is currently the only known mechanism for the resolution of ICLS and, indeed, cells deficient in the Fanconi Anaemia pathway are hypersensitive to DNA crosslinking agents such as mitomycin C (MMC) and diepoxybutane (DEB; Auerbach *et al.* 1998).

To repair cross links, the Fanconi Anaemia proteins coordinate the activities of three key DNA repair pathways, including the Nucleotide excision repair pathway (NER), homologous recombination (HR) and translesion synthesis (TLS; Ceccaldi *et al.* 2016). Depending on their function, the 22 FA genes



have been categorised into three different groups. The FA core complex, an E3 ubiquitin ligase, makes up the first group. This complex comprises FANCA, B, C, E, F, G, L, FAAP20 and FAAP100 (Kolinjivadi *et al.* 2020). A second group includes the heterodimeric complex of FANCD2 and FANCI (Kolinjivadi *et al.* 2020). Finally, group III includes a range of proteins including repair factors such as nucleases (FANCP (SLX4) and FANCO (ERCC4)), trans-lesion synthesis polymerases (FANCV (REV7) and polymerase  $\zeta$ ) and HR factors ((FANCD1 (BRCA2), FANCI (BRIP1), FANCL (PALB2), FANCO (RAD51C), FANCR (RAD51), FANCS (BRCA1) and FANCU (XRCC2); Kolinjivadi *et al.* 2020).



**Figure 1.7. The Fanconi Anaemia pathway repairs interstrand cross links**

(Adapter from Fang et al 2020)

### 1.6.1 The mechanism of ICL repair by the Fanconi Anaemia pathway

Crosslinks repaired by the Fanconi Anaemia pathway can only be repaired in the S phase of the cell cycle, coupled with replication (Knipscheer *et al.* 2009). In many cases, the presence of an ICL causes stalling of a replication fork, which leads to the detection of the lesion by an “anchor complex” made up of FANCM, FAAP24, MHF1 and MHF2 complex (Kolinjivadi *et al.* 2020). FANCM is thought to play a critical role in the detection of ICLS, and contains an N-terminal DEAH domain which is activated by branched oligos (Coulthard *et al.* 2013). Biochemical evidence suggests that UHRF1 is also involved in lesion sensing (Liang *et al.* 2015). Activation of this anchor complex recruits the core Fanconi Anaemia complex, which comprises FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL and FANCM to the lesion, where it monoubiquitinates the downstream FANCI and FANCD2 heterodimer (Walden and Deans 2014; D’Andrea and Grompe 2003; Meetei *et al.* 2004).

The Fanconi core complex is made up of three subcomplexes (Huang *et al.* 2014). The BL100 complex includes FANCB, FANCL and FAAP100, and is involved in directing the assembly of the whole core complex (Kolinjivadi *et al.* 2020). The CEF complex (including FANCC, FANCE, and FANCF), is responsible for the formation of a bridge facilitating the interaction between the core complex in its entirety and FANCD2 (Sweuc *et al.* 2017). Finally, there is the AG20 complex (comprising FANCA, FANCG and FAA20) which is thought to be required for localisation of the complex to the nucleus, but not for the ubiquitin ligase activity (Huang *et al.* 2014; Kolinjivadi *et al.* 2020).

The E3 ubiquitin ligase activity of the core complex is provided by FANCL (Meetei *et al.* 2003), which catalyses the conjugation of ubiquitin to lysine 561 of FANCD2 and lysine 523 of FANCI (Garcia-Higuera *et al.* 2001). Ubiquitin Conjugating Enzyme E2 T (UBE2T) is thought to provide the E2 enzyme activity required for FANCL-mediated ubiquitination (Machida *et al.* 2006). There is evidence that UBE2T undergoes autoubiquitination in vivo and that DNA damage in cells depleted of UBE2T leads to the accumulation of abnormal chromosomes characteristic of Fanconi Anaemia (Machida *et al.* 2006). Autoubiquitination of UBE2T is stimulated by the presence of FANCL, and could be one mechanism by which the Fanconi Anaemia pathway negatively regulates itself (Machida *et al.* 2006).

When a replication fork becomes stalled at an ICL, some of the replication machinery needs to be removed in order to repair the lesion. Breast cancer type 1 susceptibility protein (BRCA1)/FANCS works to remove the helicase CMG complex from the stalled fork, which allows the fork to move up to within just one nucleotide of the ICL. Once ubiquitylated, FANCD2 functions as a landing pad for

SLX/FANCP (Kim *et al.* 2012), which in turn acts as a scaffolding protein for DNA endonucleases MUS81, SLX1 and XPF/ERCC4/FANCD1 (Kolinjivadi *et al.* 2020). Nucleolytic cleavage by these endonucleases leaves a DNA adduct, as the crosslinked nucleotide remains bound to the complementary strand, and a double stranded DNA break. DNA translesion synthetase polymerases REV1, REV7/FANCV and REV3 bypass the DNA adduct, and repair of the double stranded DNA break is mediated by homologous recombination (Li *et al.* 2021), using the newly ligated DNA duplex as a template (Kolinjivadi *et al.* 2020). At some point, deubiquitylation of the FANCD2-FANCI complex by USP1-UAF1 takes place, though the exact timing of this is currently unknown. In this way, the FA pathway coordinates the action of multiple different DNA repair factors to resolve the ICL.

### 1.6.2 The Fanconi Anaemia pathway may also work to resolve stalled replication forks

As highlighted in the above section, replication forks running into ICLs or other unusual structures often become stalled. As well as indirectly removing stalled forks through the resolution of ICLs, there is also some evidence that the FA pathway works to resolve stalled forks directly. Indeed, Howlett *et al.* (2005) and Hussain *et al.* (2004) have shown that the FA pathway is activated upon exposure to hydroxyurea (HU) or APH, which are known to cause stalling of replication forks (and are not believed to cause ICLs). Treatment with mitomycin C, a chemotherapeutic agent which induces hypersensitivity in FANCC cells, is also thought to affect the architecture of the replication fork and the rate of replication (Zellweger *et al.* 2015).

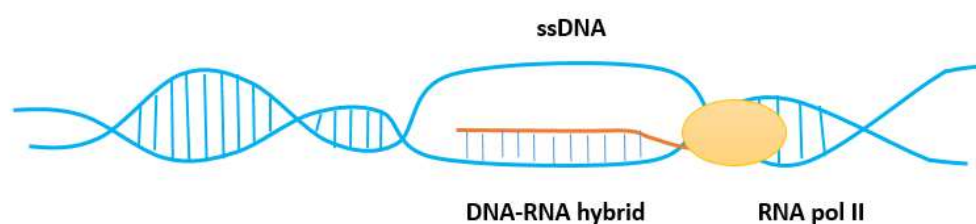
Zellweger *et al.* (2015) showed that exposure to genotoxic agents can lead to the formation of reversed forks (RVFs), a four-way junction formed from replicating nascent DNA. RVFs facilitate reorganisation after stalling in order to allow the restart of replication, however, their accumulation can also be deleterious. Work by Zellweger *et al.* (2015), has shown that fork reversal is dependent on RAD51 (FANCR) and isolation of Proteins enriched On Nascent DNA (iPOND) analysis performed in cells exposed to HU indicate that RAD51 is associated with stalled replication forks (Kolinjivadi *et al.* 2017). RAD51, together with BRCA1 and BRCA2, (key components of the homologous end joining and FANC pathways), were found to play a key role in fork protection through the stabilisation of nascent DNA (Schlachter *et al.* 2012). This finding was further supported by electron microscopy experiments in frog egg extracts, which captured replication forks at several intermediate stages (Hashimoto *et al.* 2010). This study found that RAD51 prevents the accumulation of ssDNA gaps at replication forks and protects DNA from degradation, enabling the resumption of DNA synthesis.

Several studies have suggested that FANCD2 also associates with BRCA1 and RAD51, and possibly has a function at stalled replication forks (Taniguchi *et al.* 2002; Kais *et al.* 2016; Michl *et al.* 2016). FANCD2 appears to maintain fork stability in tumours lacking BRCA1/BRCA2, and may also limit replication stress in cells lacking BRCA2. A particularly compelling study was performed by Tonzi *et al.* (2018), in FANCD2 deficient patient fibroblasts. The cells showed impaired fork restart that was rescued by complementation with FANCD2. Following exposure to HU for 5 h, the FANCD2 mutant cells had significantly shorter tracks of nascent DNA than healthy cells. The authors propose that this difference is caused by increased MRE11 mediated degradation of DNA in the FANCD2 deficient cells.

Finally, a specific role for FANCI in fork protection is also emerging. FANCI is a helicase from the iron sulphur family, and its loss has been implicated in hereditary breast and ovarian cancer (Cantor *et al.* 2004; Litman *et al.* 2005; Minion *et al.* 2015). Analysis of patient derived mutations has shown that helicase activity is required for ICL repair and iPOND analysis has identified FANCI as a replication fork associated protein (Peng *et al.* 2018). Fork degradation has also been observed in patient cells that were null for FANCI. It is thought that FANCI and the fork remodeler helicase-like transcription factor (HLTF) work together to balance degradation at the replication fork and to facilitate continuous DNA synthesis, especially during replication stress.

### 1.6.3 R loops in Fanconi Anaemia

As well as repairing DNA lesions such as ICLs and stalled replication forks, it is thought that the Fanconi Anaemia pathway protects the genome from a type of DNA lesion called an R loop. An R loop is a three stranded structure that is made up of a DNA: RNA hybrid bound by nucleic acid base pairing, and a displaced single strand of DNA (**Figure 1.7**). These structures can be involved in transcription and replication but can also generate genomic instability. R loops can be resolved through RNase H (reviewed in Cerritelli and Crouch 2009) or Senataxin (Boulé and Zakian 2007), or prevented by proteins such as Topo I, the THO complex or the SRSF splicing factor (Aguilera *et al.* 2012). These protective measures are important for the maintenance of genome stability, as the accumulation of R loops can act as a barrier to replication fork progression and cause replication stress (Gan *et al.* 2011; Wellinger *et al.* 2006).



**Figure 1.8. The structure of an R loop**

Because of the role of BRCA1 and BRCA2 in the FA pathway to resolve ICLs, Garcia Rubio *et al.* (2015) investigated whether R loops were a source of genomic instability in FA cells, and whether the FA pathway has a role in the regulation of R-loops. To do this, the group performed DRIP-PCR (DNA-RNA immunoprecipitation PCR) in human genes, *APOE*, *RPL13A*, *EGR1* and *BTBD1* in cells derived from FA patients. These genes are regions which have previously been observed to contain accumulations of R loops (Ginno *et al.* 2013). In all four genes, significantly higher frequencies of DNA: RNA hybrids were observed in cells lacking FANCA. Treatment with RNase H, which resolves R loops through the degradation of RNA, reduced the frequency of the R loops, confirming that the signal observed was resulting from DNA: RNA hybrids. A similar effect was reported in a well characterised PD20 human fibroblast cell line from a FANCD2<sup>-/-</sup> patient.

In 2019, Chang *et al.* performed a trigenic genome wide interaction screen in yeast to identify factors involved in the suppression of R loops. The screen was performed in cells lacking RNH1 and RNH201, and genes were identified which affected the fitness of the cells in this context. The pool of genes identified in the screen was enriched for DNA replication fork maintenance factors. The screen determined that depletion of MRE11, RAD50, or NBS1 resulted in R loop dependent DNA damage. The authors speculated that MRN suppressed R loops through the recruitment of FA factors. Depletion of FANCD2 by siRNA resulted in an increase in R loops, in line with previous work which has linked MRN to FA pathway activation (Roques *et al.* 2009) and the regulation of R loops by the FA pathway (Garcia Rubio *et al.* 2015).

Schwab *et al.* (2015) had previously suggested that the helicase FANCM has the ability to resolve R loops in vitro. Supporting this theory, Chang *et al.* (2019) observed through CHIP-cPCR that FANCM was binding to loci associated with R loops. The R loop is therefore another lesion linked to the FANCD2 phenotype.

#### 1.6.4 Additional roles of Fanconi Anaemia proteins

It has been thought since the 1980s that the Fanconi pathway may play a cytoprotective role in addition to its function in the repair of ICLS. Studies have implicated FA proteins in the protection of cells from ROS (Schindler and Hoehn, 1988), deprivation of growth factors (Haneline *et al.* 1998) and exposure to inflammatory cytokines (Cumming *et al.* 1996).

A number of studies have been performed on FANCC in particular. Reports as early as 1996 (Yamashita *et al.* 1996) have indicated that the protein may have cryoprotective roles independent of its function in resolving ICLS. FANCC has been shown to protect against pro-inflammatory cytokine-induced cell death through interactions with interferon signalling proteins such as signal transducer and activator of transcription 1 (STAT1), protein kinase R (PKR) and stress-inducible heat shock protein 70 (HSPA1A; Pang *et al.* 2000; Pang *et al.* 2002). FANCC has also been shown to protect against toxicity caused by withdrawal of growth factors in haematopoietic cells (Cumming *et al.* 1996).

Analysis of the link between genotype and phenotype in FANC patients has shown that individuals carrying a c.67delG mutant in FANCC have a milder disease than patients harbouring a FANCC which suppresses both roles in DDR and additional functions (Neveling *et al.* 2009; Sumpter *et al.* 2016). This suggests that FANCC has different roles, and alteration of them can influence the severity of the disease.

Several FANC proteins have also been associated with autophagy, an important pathway for cellular homeostasis (Orvedahl *et al.* 2010). Autophagy is important for the removal of faulty proteins, damaged organelles and pathogens. Its deregulation has been linked with multiple diseases. In a genome-wide screen Orvedahl *et al.* 2011) identified FANCC, FANCF and FANCL as potential regulators of autophagy. Further experimentation indicated that FANCC is required for a particular form of selective autophagy called virophagy, which cells use to degrade viruses or virus-derived antigens.

Other FANC proteins have also been implicated in cellular responses outside of DDR. FANCI, for example, is believed to be involved in the response to replication stress through regulating the firing of dormant origins in response to treatment with hydroxyurea (Chen *et al.* 2015). Though these origins of replication are normally silent, reports found that they can be utilised following the disruption of replication. Interestingly, FANCD2 appears to inhibit this activity. In an independent process, FANCD2 and BLM appear to work together to suppress firing from dormant origins (Chaudhury *et al.* 2013).

This work shows that there is a precedent for FANCD proteins to be involved in cellular processes outside of the FA pathway.

### 1.7 The Fanconi Anaemia pathway may regulate L1 retrotransposition.

Several studies have indicated that members of the Fanconi Anaemia pathway could play a role in the regulation of retrotransposition (Mita *et al.* 2020; Liu *et al.* 2018). A previously mentioned whole genome siRNA screen performed by Mita *et al.* (2020) identified several FA factors as inhibitors of retrotransposition. The strongest effect was seen in BRCA2, PALB2, and FANCB. siRNA depletion of these proteins was observed to increase retrotransposition approximately 3-fold. In contrast, siRNA depletion of FANCD1 and FANCD2 was found to reduce retrotransposition, indicating that these proteins are “supporters” of L1 mobility. Analysis of retrotransposition using a luciferase reporter corroborated that FANCD1 behaves differently to the other tested L1 proteins. A cluster of the FA proteins had previously been identified as potential L1 regulators in a genome wide screen performed by Liu *et al.* (2018).

The authors argue this data supports a model where L1 can exploit stalled replication forks as a site for retrotransposition. As retrotransposition is mainly believed to take place during S phase, the model is plausible (Mita *et al.* 2018). L1 retrotransposition assay in cells treated with varying concentrations of aphidicolin (APH) a DNA polymerase inhibitor that creates stalled replication forks, L1 retrotransposition increased with the concentration of APH, indicating that L1 may be recruited to stalled forks during S phase of the cell cycle.

The strongest effect seen in the siRNA screen was the depletion of BRCA1. Use of a reporter assay with BRCA1-null ovarian carcinoma cell line UWB1.289 and an isogenic cell line expressing of wild-type BRCA1 (UWB1.289+BRCA1), showed that expression of BRCA1 led to severely restricted L1 activity. The endonuclease independent activity of the element was also measured in both cell lines. Intriguingly, depletion of BRCA1 and FANCD2 led to an increase in the amount of endonuclease independent retrotransposition. This effect was confirmed using knockdowns of BRCA1 and FANCD2, and using a system with a different reporter gene. Interestingly, knockdown of BRCA2 led to a reduction in L1 endonuclease dependent retrotransposition, despite the fact that it increased endonuclease independent activity

Previous work by Mita *et al.* (2018) has suggested that L1 proteins localise to replication forks. This would support an interaction of L1 proteins with FA factors, which are known to act at stalled replication forks. Immunoblot analysis revealed that ORF1p is recruited to chromatin during S phase along with UPF1 and PCNA. Mass spectrometry analysis also confirmed that the ORF1p/PCNA



complex interacts with TOP1, RPA1 and PARP1, all of which are known to localise to stalled replication forks (Taylor *et al.* 2018). Mita *et al.* (2018) had speculated that L1 RNPs are recruited to replication forks during S phase, and that the stalled forks may help to mediate retrotransposition. The group also performed isolation of proteins on nascent DNA (iPOND) analysis, combined with immunoblotting in HeLa cells which expressed both recoded (ORFEUS) and non-recoded (L1rp) L1. ORF1p from both elements appeared to localise to replication forks. Due to its much lower expression profile, ORF1p was unsurprisingly not detected.

Further evidence that the Fanconi pathway is involved in the regulation of L1 comes from a cellular model of Fanconi Anaemia which is deficient in SLX4 (FANCP; Bregnard *et al.* 2016). SLX4 is believed to be involved in the regulation of pathogen stimulated cytokine production and directs the repair of DSBs by assembling several proteins into a complex named SLX4com. L1 DNA extracted from both RA3331 cells and RA3331 SLX4<sup>-</sup> cells and quantified by PCR or southern blot, revealed that RA3331 cells had higher levels of L1 DNA in comparison to RA3331 SLX4<sup>-</sup> cells, indicating that L1 DNA accumulates in the absence of SLX4. The group also tested for an interaction between ORF1p and SLX4 by immunoprecipitation with an N-terminally FLAG and HA tagged SLX4 protein. ORF1p was detected in the sample containing wild type SLX4, but not the mutant of the SAP domain. This indicates that ORF1p interacts with SLX4 and likely through the SAP domain. A cultured cell assay in HeLa cells using an SLX4 construct and a SAP mutant showed that the wild type protein restricts retrotransposition of L1. This result was also confirmed with a luciferase reporter system. Furthermore, this effect could be rescued by complementing the mutant cells with SLX4 protein.

### 1.7.1 L1 retrotransposition at replication forks.

Mita *et al.* have proposed that stalled replication forks could be exploited as sites for retrotransposition. The group found evidence suggesting that factors involved in fork protection are needed for BRCA1 mediated repression of LINE1. When Mita *et al.* (2020) knocked down BRCA2, which plays a key role in fork protection, they observed that cells with depleted BRCA1 had similar levels of retrotransposition to wild type cells. Furthermore, both RPA1 and RAD51 appear to restrict L1 retrotransposition. Depletion of BOD1L, which is important for the stabilisation of RAD51 at damaged replication forks, also abrogated BRCA1 induced L1 repression. (Higgs *et al.* 2015). This finding suggests that coating of the fork by RAD51 is important for the inhibition of L1 by BRCA1.

Mita *et al.* (2020) believe their findings could be explained by a model wherein DNA2 and CtIP play a BRCA1 dependent role in fork protection. In this model CtIP cuts the stalled fork and DNA2 partially degrades the DNA. Concordantly, the group found that retrotransposition levels in the absence of BRCA1, DNA2 and CtIP were similar to those in cells depleted with BRCA by siRNA. This would suggest that L1 can exploit replication forks which are not “coated” or protected by BRCA1 or DNA2.

## 1.8. Aims of this thesis

The relationship between DNA repair pathways and L1 retrotransposition is complex. DNA repair factors are increasingly being identified as modulators of L1 retrotransposition, and several studies have demonstrated how retrotransposition is mechanistically altered against different DNA repair backgrounds.

In this thesis, I aim to shed further light on the relationship between DNA repair pathways and retrotransposition of L1 elements. My experiments will focus on addressing three primary aims:

1. To determine which L1 domains are required for retrotransposition in a FANC background.
2. To determine whether FANC proteins demonstrate the capacity to regulate L1.
3. To identify new L1 interactors in a FANCC context.



## Chapter 2

---

# **Materials and methods**



## 2.1 General reagents.

Chemicals were purchased from Sigma Aldrich, BDH Laboratory Supplies (AnalaR, VWR), Fisher Chemicals, and Amersham Biosciences (GE Healthcare). Enzymes were obtained from New England Biolabs, Promega and Roche. Cell culture material was purchased from Gibco (Invitrogen) or Sigma-Aldrich unless otherwise stated.

## 2.2 Cell culture methods:

Cell lines used in this thesis and the cell culture conditions are listed in Table 2.1. For passaging, cells were detached using a mixture of Trypsin and Versene (ratio 1:1) and a 3-minute incubation at 37 degrees. Trypsin mix was inactivated using an equal amount of cell culture media and cells were pelleted at 1200 rpm, washed with PBS and then resuspended in the appropriate cell culture media. Cells were passaged at a ratio of 1:4 – 1:18, depending on the growth of the individual cell line.

**Table 2.1 Cell culture drug stock solutions.**

<b>Drug</b>	<b>Solvent</b>	<b>Stock concentration</b>	<b>Working concentration</b>	<b>Manufacturer</b>
Neomycin	H <sub>2</sub> O	250 mg/ml	500 µg/ml	Sigma-Aldrich
Blasticidin	H <sub>2</sub> O	10 mg/ml	2 µg/ml	Sigma-Aldrich
Mitomycin C	H <sub>2</sub> O	1.2 nM	5–360 nM	Sigma-Aldrich

## 2.2.1 Cell lines used in this thesis

**Table 2.2 Cell lines used in this thesis**

Cell line	Cell culture media	Media supplements	Cell culture conditions	Source
VH4	DMEM, Low Glucose, Pyruvate (Thermofisher scientific catalogue number 11885084)	10% Fetal Bovine serum, 1% non-Essential amino acids 100 U/ml Penicillin, 100 µg/ml Streptomycin	37°C, 5% CO <sub>2</sub> , normoxic	Jose-Garcia Perez Lab Genyo, Granada, Spain. ANKN and PD2DF are human fibroblasts derived from patients with Fanconi Anaemia, these were gifted to the Jose-Garcia Perez lab from Dr. Niall Howlett (University of Rhode island)
V79				
VH4 + FANCA				
XR-1				
4364				
CLV5B				
CLV5B+FANCC				
PD20F	DMEM high glucose (Thermofisher scientific catalogue no. 11965084)	10% Fetal Bovine serum, 1% non-Essential amino acids 100 U/ml Penicillin, 100 µg/ml Streptomycin	37°C, 5% CO <sub>2</sub> , Normoxic	
ANKN				
HAP1	DMEM/F12 (Thermofisher scientific catalogue no. <b>11320033</b> )	20% Fetal Bovine serum 100 U/ml Penicillin, 100 µg/ml streptomycin	37°C, 5% CO <sub>2</sub> , Normoxic	Prof. Alexander Ruzov (Nottingham University, UK)
YTHDF2				Prof. Alexander Ruzov (Nottingham University, UK)

### 2.2.2 Preservation of cell lines.

For storage, cell lines were grown to a confluency of 70% and then pelleted at 1200 rpm and washed with PBS. Cells were frozen in quantities of  $1-4 \times 10^6$  in 2ml cryostat tubes, in 500ul of FBS containing 10% DMSO. After transfer to freezing mix, cells were stored overnight in polystyrene boxes at 70 degrees, before being moved to liquid nitrogen.

### 2.2.3 Transfection of cell lines for retrotransposition assay

#### 2.2.3.1 Transfection of human cell lines for cultured cell retrotransposition assay.

Human cell lines were plated in six-well tissue culture plates 24 hours prior to transfection at seeding density of  $4 \times 10^4$ . Transfection was performed using FuGene 6 (Promega corporation). Transfection mix was made by first mixing 97 ul Opti-MEM Medium (Thermo Fisher Scientific, 31985062) per well to be transfected, with 1  $\mu$ G plasmid DNA and leaving to incubate for 5 minutes at RT. Next, 3ul of FuGene 6 was added to make a total volume of 100ul transfection mix per well, and mix was left to incubate at RT for a further 20 minutes. Transfection mix was added to wells drip wise while rotating tissue culture plate. Cells were incubated in transfection mix for 24 hours before being washed and transferred to normal tissue culture media.

#### 2.2.3.2 Transfection of hamster cell lines.

Transfection was largely performed as described above, using FuGene 6 as the transfection reagent. Hamster cells were seeded at a density of  $4 \times 10^4$  8 hours before transfecting. The 8h time window between plating and transfection was determined empirically by through experimentation, in order to maximise transfection efficiency.



#### 2.2.3.4 Transfection of hamster cells for co-immunoprecipitation

Cells were plated at a density of  $8.5 \times 10^5$  per 10cm dish, in cell culture media made up without antibiotics, to be 80% confluent. 8 hours after plating, transfection was performed using Lipofectamine 2000 transfection reagent (Invitrogen, Life Technologies. **Catalogue no.** 11668019). Conditions for transfection were optimised empirically.

For each 10cm plate, 15ug of DNA was diluted into 1.5ml of Opti-MEM reduced serum media, mixed, and allowed to incubate at RT for 5 minutes. In a separate falcon tube, 1.5ml of Opti-MEM media was mixed with 33ul of Lipofectamine and allowed to incubate for 5 minutes. Next, the diluted DNA was combined with the diluted lipofectamine, and allowed to incubate at RT for 20 minutes. For each 10cm plate, 3ml of transfection mix was added while the plate was gently rotated to mix. Cells were left in transfection mixture for 24 hours before it was removed and replaced with cell culture media.

#### 2.2.4 Measurement of transfection efficiency

To quantify transfection efficiency, a co-transfection was performed alongside the cultured cell retrotransposition assay. Each DNA plasmid to be transfected was mixed 0.5ug to 0.5ug with a plasmid which comprised of the CEP-4 vector, containing the EGFP reporter gene. Transfection was performed according to the protocol described above. 24 hours following transfection, the Opti-MEM mix was replaced with standard cell media. 48 hours following transfection, cells were harvested by washing twice with PBS and trypsinisation for three minutes at RT. Trypsin was neutralised using an equal volume of PBS supplemented with 20% FBS and cells were pelleted using centrifugation at 1200 rpm. Pellet was resuspended in 200 ul of PBS supplemented with 20% FBS and kept on ice.

The percentage of cells expressing GFP following the transfection was measured using the BD Accuri flow cytometer.

When scaling transfections up or down to different tissue culture vessels, the following relative surface areas were used. Experiments were predominantly performed in 6 well plates or 10cm dishes, however other vessels were used in addition.

**Table 2.3 Relative surface area of different cell culture vessels used in conversion**

Culture Vessel	24-well	12-well	6-well	60 mm	100 mm	150 mm	T25	T75
Surface area (cm <sup>2</sup> )	2	4	9	28.2	78.5	176.7	25	75
Ratio to 24-well plate	1	2	4.5	14.1	39.25	88.35	12.5	37.5

### 2.2.5 Cultured cell retrotransposition assay.

LINE-1 retrotransposition assays were carried out in triplicate with at least three biological replicates. The protocol was carried out as previously described (Moran et al 1996). Cells were plated and transfected according to the protocol described above and according to the specific cell line. Cells were left to incubate in transfection mix for 24 hours and then mix was replaced with normal culture media.

After 48 hours, cell culture media was replaced with media supplemented with 400ug/ul G418. Cells were kept under selection with G418 for 12 days, using a new media solution every other day. On the 13<sup>th</sup> day of selection, media was aspirated and cells were washed with PBS. Colonies were incubated in a PBS solution with 2% formaldehyde for 30 minutes at RT to fix the cells. Fixed colonies were stained by treatment with 0.1% Crystal Violet (W/V) for 10 minutes at room temperature.

For all experiments a co-transfection with pCEP-EGFP was performed in parallel to generate a transfection efficiency for each plasmid.

### 2.2.6 Mitomycin C Sensitivity assay.

MMC was obtained from Sigma Aldrich (Cat. No. M4287-2MG), from *Streptomyces caespitosus*. Dilutions of 1.2mM were made in distilled water and stored at -80 degrees in an opaque container to protect from light. MMC was diluted to 50 nM, 75 nM, 100 nM and 200 nM in distilled water. For the duration of the experiment, solutions containing MMC were wrapped in foil and all experimentation was carried out in limited light to minimise photosensitivity, as MMC is unstable, fresh dilutions were prepared for each experiment.

Cells were seeded at densities of  $1 \times 10^4$  in human FANCC cell lines and  $4 \times 10^4$  in hamster FANCC cell lines, the day before treatment with MMC. Treatment with MMC was maintained for 72 hours, with each condition in triplicate. Following treatment, cells were maintained overnight in standard cell culture media (no longer than 18 hours).

Cells were incubated for 10 mins at RT with mild agitation in 1ml fixing solution per well (Distilled water with 10% Acetic Acid and 10% Methanol). Fixed colonies were washed twice with distilled water then treated with 2% Crystal violet solution. Cells were left to dry overnight, then, to visualise, 500ul of dissolving solution was added to each well (methanol with 1% w/v SDS). Plates were sealed with parafilm then left to incubate for 2 hours with mild shaking at RT.

From each well, 100ul of solution was transferred to a 96 well plate. The OD595 of each sample was measured using a spectrophotometer and a Bradford protein ladder for reference.

### 2.2.7 Generation of FANCC complemented cells

FANCA complemented cells were generated using retroviruses. Briefly,  $4 \times 10^4$  FANCA deficient cells were plated per 10cm plate, and 18h later cells were washed twice with PBS1x and infected with each corresponding virus in the presence of polybrene. Antibiotic selection was used to select cells expressing the complementing FANC gene (or a control), and was started 48h after infection. After selection and generation of clonal cell lines, we used MMC to test the efficiency of the complementation. MMC treatments were conducted as described (Howlett et al., 2009). Briefly,  $5 \times 10^3$  cells were plated per well in a 12-well dish in triplicate. We plated parental cells (if available), FANC-mutant cells, and complemented clonal lines (or controls). 18h after plating, we treated cells with 0, 50, 75, 100 and 200 nM MMC during 72h. We used MMC from *Streptomyces caespitosus* (Sigma) and a fresh stock solution was prepared for each assay. 72h after treatment, cells were washed, fixed with acetic acid-methanol, and stained with 0.1% (w/v) crystal violet during 10 minutes. After staining, plates were washed with distilled water and dried overnight. Next day, crystal violet

was dissolved from wells using 250 µl of 1% (w/g) sodium dodecyl sulphate SDS in Methanol during 2h at room temperature. 100 µl from each well were transferred to a well of a 96-well plate and OD595 values were determined using a spectrophotometer. The value obtained with parental cells (or a wild type control) treated with 0nM MMC was assigned 100%, and the rest of values are referred to this control

## 2.3 Nucleic acid methods

### 2.3.1 Quantification of nucleic acids.

To quantify the concentration of nucleic acids in all experiments, a Nanodrop 1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific) was used. For each sample, 1ul was placed on the Nanodrop and the optical density at 260nm was measured. To determine the purity of the samples, light absorbance at three ratios 230 nm, 260 nm and 280 nm, was measured. These values are indicated in two ratios The 260/280 ratio reflects the amount of contaminating protein relative to nucleic acid, and should be 1.8-2.2. The 260/230 ratio shows other contaminants such as salts or lipids, and should be >1.7.

### 2.3.2 Gel electrophoresis for visualisation of nucleic acids

Nucleic acids were analysed using agarose gel electrophoresis. Gels were made by dissolving 1.5% agarose (Hi-Pure Low EEO agarose, Biogene) in 0.5% TBE by heating for 2 minutes in a microwave oven. After cooling solution under cold water, Ethidium Bromide was added to a concentration of 0.5 µg/ml. Prior to loading, DNA samples were mixed with 6X Gel loading dye (New England Biolabs B7024S). Gels were run at 5 volts/cm voltage. To visualise the resolution of the nucleic acids by size, a UV transilluminator was used (BioDoc-It System, UVP). Molecular weight was estimated using a 1Kb DNA ladder (Invitrogen) for reference.

### 2.3.3 Plasmids used in this thesis

#### **Table 2.4 Plasmids used in this thesis**

<u>Name of plasmid</u>	<u>Description</u>	<u>Figure reference</u>
<b>pAD2TEI</b>	pCEP6 vector with CMV and Neo cassette, contains ORF1p-T7 and ORF1p-TAP	<b>Figure 4.3</b>
<b>pAD2TEI-102</b>	pCEP6 vector with CMV promoter and Neo cassette, contains mutant ORF1p-T7 (REKG to AAAA) and ORF1p-TAP	<b>Figure 4.3</b>
<b>pAD2TEI-105</b>	pCEP6 vector with CMV promoter and Neo cassette, contains mutant ORF1p-T7 (ARR to AAA) and ORF1p-TAP	<b>Figure 4.3</b>
<b>pAD2TEI-106</b>	pCEP6 vector with CMV promoter and Neo cassette, contains mutant ORF1p-T7 (ARR to AKK) and ORF1p-TAP	<b>Figure 4.3</b>
<b>pAD2TEI-107</b>	pCEP6 vector with CMV promoter and Neo cassette, contains mutant ORF1p-T7 (ARR to AKR) and ORF1p-TAP	<b>Figure 4.3</b>
<b>pAD2TEI-108</b>	pCEP6 vector with CMV promoter and Neo cassette, contains mutant ORF1p-T7 (ARR to ARK) and ORF1p-TAP	<b>Figure 4.3</b>
<b>pAD2TEI-135</b>	pCEP6 vector with CMV promoter and Neo cassette, contains ORF1p-T7 and RT mutant ORF1p-TAP (D702A)	<b>Figure 4.1</b>
<b>pAD2TEI-136</b>	pCEP6 vector with CMV promoter and Neo cassette, contains ORF1p-T7 and EN mutant ORF1p-TAP (H230A)	<b>Figure 4.1</b>
<b>pAD2TEI-PIP6</b>	pCEP6 vector with CMV promoter and Neo cassette, contains ORF1p-T7 and mutant ORF1p-TAP (PIP6)	<b>Figure 4.1</b>
<b>pAD2TEI-162</b>	pCEP6 vector with CMV promoter and Neo cassette, contains ORF1p-T7 and mutant ORF1p-TAP (C-dom mutant)	<b>Figure 4.1</b>
<b>pSA500</b>	pCEP6 vector with CMV promoter and Neo cassette, contains ORF1p-TAP (C-dom mutant)	<b>Figure 4.7</b>
<b>pSA500-D205Aà ENm</b>	pCEP6 vector with CMV promoter and Neo cassette, contains ORF1p-TAP (EN mutant-D205A)	<b>Figure 4.7</b>

<b>pSA500-H230Aà ENm</b>	pCEP6 vector with CMV promoter and Neo cassette, contains EN mutant (H230A) ORF1p-TAP	<b>Figure 4.7</b>
<b>pSA500-YY414/15AA</b>	pCEP6 vector with CMV promoter and Neo cassette, contains PIP mutant ORF1p-TAP	<b>Figure 4.7</b>
<b>pSA500-D702A</b>	pCEP6 vector with CMV promoter and Neo cassette, contains RT mutant ORF1p-TAP	<b>Figure 4.7</b>
<b>pSA500-H230A/D702A</b>	pCEP6 vector with CMV promoter and Neo cassette, contains RT and EN mutant ORF1p-TAP	<b>Figure 4.7</b>
<b>JJ101/L1.3</b>	pCEP6 vector with CMV promoter and BLAST cassette, contains WT L1	<b>Figure 4.1</b>
<b>JJ101/L1.3-D205A</b>	pCEP6 vector with CMV promoter and BLAST cassette, contains ORF1p and EN mutant ORF1p	<b>Figure 4.1</b>
<b>JJ101/L1.3-YY414/15FF</b>	pCEP6 vector with CMV promoter and BLAST cassette, contains ORF1p and PIP mutant ORF1p	<b>Figure 4.1</b>
<b>JJ101/L1.3 RTm(D702A)-</b>	pCEP6 vector with CMV promoter and BLAST cassette, contains ORF1p and RT mutant ORF1p	<b>Figure 4.1</b>

## 2.4 Protein methods

### 2.4.1 Immunoprecipitation

The buffers used for the immunoprecipitation are shown in Table 2.3. During the immunoprecipitation protocol, all samples, tubes and buffers kept ice cold. Complete protease inhibitors (1X) are added on the day.

**Table 2.5 Immunoprecipitation buffers used in this thesis**

Buffer	Ingredients
Lysis Buffer	20 mM HEPES [pH 7.5], 150 mM NaCl, 0.5% NP40, 1mM EDTA, H <sub>2</sub> O Complete protease inhibitors 1X
5X Binding buffer	20 mM HEPES [pH 7.5], 150 mM NaCl, 1 mM EDTA, H <sub>2</sub> O, Complete protease inhibitors to 1X
Wash buffer	20 mM HEPES [pH 7.5], 150 mM NaCl, 0.1% NP40, 1mM EDTA, H <sub>2</sub> O Complete protease inhibitors 1X

Cells were harvested by spinning at 1200 RPM and washing twice with ice cold PBS. Pellet was resuspended in 60ul of Lysis buffer and transferred into a 2ml pre-cooled Eppendorf tube. Lysate was then maintained at constant agitation for 20 minutes at 4 degrees. Centrifuge in a micro centrifuge at 4°C, 13,000rpm for 10 min, lysate was transferred into a new microcentrifuge tube, leaving a sticky chromatin pellet behind. Supernatant was diluted at a ratio of 1:4 in 5X binding buffer. Next, 6ug/IP (6ul) of immunoprecipitating antibody (T7, abcam9138) or isotype control antibody (IgG) was added to the lysate and incubated at 4°C with gentle rotation for 2h.

Immunoprecipitation was performed using 10ul of protein G beads per sample. Beads were washed prior to use with 1ml Wash Buffer and separated from buffer using a magnetic rack. Beads were mixed with IP samples and mix was incubated O/N rotating at 4 degrees.

Lysate was next separated from beads using a magnetic rack. Supernatant was discarded or saved for western blotting. Beads were washed 2X with wash buffer. Finally, beads were washed once more with TBS and transferred to a fresh Eppendorf tube.

## 2.4.2 Western blotting

**Table 2.6 Reagents used in western blotting**

Buffer	Chemicals used for preparation
10X TBE	0.89 M Tris base, 0.89 M boric acid, 20 mM EDTA
10X TBS	0.5 M Tris base, 1.5 M NaCl (pH adjusted to 7.5 with HCl)
20X SSC	3 M NaCl, 0.3 M sodium citrate (pH adjusted to 7 with HCl)
1X urea lysis buffer	8 M urea, 50 mM Tris (pH 7.5), 150 mM $\beta$ -mercaptoethanol, cOmplete mini EDTA-free protease inhibitor cocktail (Roche, catalogue number 04693159001)
10X Tris-glycine SDS-PAGE running buffer	250 mM Tris base, 1.92 M glycine, 1% (w/v) SDS
1X immunoblotting transfer buffer	25 mM Tris base, 192 mM glycine, 0.1% (w/v) SDS, 20% (v/v) methanol
4X SDS protein sample loading buffer	0.5 M Tris-HCl (pH 6.8), 50% (v/v) glycerol, 2% (w/v) SDS, 0.1% (w/v) bromophenol blue
1X TE	10 mM Tris-HCl (pH 8), 1 mM EDTA

### 2.4.2.1 Generation of lysate from whole cells.

Mammalian cells were harvested using trypsin and pelleted by centrifugation at 1200rpm. Cells were lysed using 30-60ul of lysis buffer (depending on size of the pellet). Cells were subjected to sonication by Diagenode Bioruptor 300 Sonicator device at 4 degrees (sonication was performed in two rounds of 30 seconds with a 30 second gap in between). Sample was then centrifuged at 17,000 x g for 30 min at 4°C. Supernatant was removed for storage at -80 degrees Celsius.



#### 2.4.2.2. Quantification of protein in whole cell lysate

Before downstream experimentation, Quick Start Bradford Protein Assay (Bio-rad) was used to quantify the protein concentration in a sample. Protein samples at defined concentrations (BSA concentration range of 0.2, 0.4, 0.6, 0.8 and 1.0 mg/ml), provided by the manufacturer, were used to generate a standard ladder for reference. In the same plate as the protein ladder, 10ul of lysate was mixed with 190ul Bradford Dye Reagent. Using a spectrophotometer, absorbance at 595 nm ( $A_{595}$ ) was measured. Protein concentrations of lysate were determined using the absorbance readings of the standard proteins for reference.

#### 2.4.2.3. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) for protein resolution

IN preparation for western blotting, Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on whole cell lysates to resolve proteins according to molecular weight. Gels in this thesis were either home-made, or bought pre-cast using the 4–12% Bis-Tris NuPAGE® Novex (Invitrogen) gel system.

To make homemade gels, a resolving gel and a stacking gel were made separately. For A 10% resolving gel, 375 mM Tris (pH 8.8), 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate, 0.04 µl/ml TEMED and 10% acrylamide). The stacking gel was made using 125 mM Tris (pH 6.8), 0.1% SDS, 0.1% ammonium persulphate, 4.2% acrylamide/bisacrylamide and 1 µl/ml TEMED.

For pre-cast NuPAGE gels, the XCell SureLock™ Mini-Cell gel running chamber was used (Thermo fisher scientific, Catalogue number: EI0001) and gels were run using NuPAGE MOPS SDS running buffer (Thermo Fisher Scientific, NP0001) for Bis-Tris NuPAGE® gels. For homemade gels, the Bio-rad Mini-PROTEAN® Tetra Vertical Electrophoresis system (catalogue number 1658000FC) was used.

Prior to electrophoresis, samples were denatured in a heat block at 95 degrees using 1X SDS protein sample loading buffer. For size reference, samples were loaded onto gel alongside Precision Plus Protein All Blue Prestained Protein Standards (Bio-Rad). Gel was run at 145 V for 1 hour and 30 minutes or until proteins were appropriately resolved.

#### 2.4.2.4 Immunoblotting

Following SDS-PAGE, gels were processed in a Mini-Trans-Blot Cell system (Bio-Rad) to allow transfer of protein to a nitrocellulose membrane (Amersham). Transfer was performed at 4 degrees in an immunoblotting (**Table 2.6**) transfer buffer, at 100 V for 1 hour.

Following the protein transfer, the nitrocellulose membrane was incubated with a blocking solution containing 5% milk solution (prepared using Marvel milk powder, Premier Foods) in TBS/T (1X TBS with 0.1% TWEEN 20), for 1 hour at room temperature with gentle agitation. After blocking, the nitrocellulose membrane was incubated in primary antibody, diluted to the appropriate concentration in blocking buffer. Treatment with primary antibodies was performed overnight, at 4 degrees in a 50 ml Falcon tube with constant, gentle rotation.

The next morning, the nitrocellulose membrane was washed 3X with TBS/T. Each wash was performed for 15 minutes at room temperature, with constant agitation. Next the membrane was incubated for 1 hour with secondary antibody, diluted to the appropriate concentration in blocking buffer. Incubation with secondary antibody was performed at room temperature in a 50 ml falcon tube. Secondary antibodies were horseradish peroxidase (HRP)-labelled, and selected for compatibility with primary antibodies.

The HRP signal from bound secondary antibody was detected using Amersham ECL Prime Western Blotting Reagent (GE Healthcare LifeSciences). The nitrocellulose membrane was incubated for 5 minutes at room temperature in an Imaging solution was made up using a 1:1 ratio of solutions A and B (contained in the kit). Approximately 2 ml of ECL solution was used for 20cm of nitrocellulose membrane, gently applied to ensure even distribution across the membrane. Excess liquid was removed through blotting with Wattman paper. The membrane was then imaged using the ImageQuant LAS 4000 camera system, which measures the chemiluminescence signal. Optimal exposure was chosen by testing a range of exposures with different time intervals. Alternatively, the membrane was placed between two sheets of acetate paper and exposed to photographic film (Kodak Biomax XAR Film), developed using Konika SRX-101A Developer.

## 2.5 Microbial methods

### 2.5.1 Growth of bacteria

*E. coli* strains were grown at 37°C in/on Luria-Bertani (LB) medium (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl, 1 g/l glucose). To maintain selection for plasmid DNA in transformed cells, the relevant antibiotic (s; **Table 2.4**) were added to LB at the required concentration.

**Table 2.7 Working concentrations of antibiotics**

Antibiotic	Stock concentration	Working concentration	Solvent
Ampicillin	50 mg/ml	50 µg/ml	dH <sub>2</sub> O
Kanamycin	10 mg/ml	10 µg/ml	dH <sub>2</sub> O

### 2.5.2 Transformation of *E. coli*

*E. coli* were transformed by adding 1µg of plasmid or 2.5µg of ligation mixture to 25µl of ultra, chemically competent bacterial cells. The mix was incubated for 30 minutes on ice before a 45 second 42-degree heat shock. Bacteria was incubated on ice for another 2 minutes, before a further incubation of 60 minutes at 37 degrees. 100µl of transformed bacteria is then spread onto an LB -agar plate treated with the appropriate antibiotic.

### 2.5.3 Generation of ultra-competent cells.

Ultracompetent cells for recovery assay were generated using XL1-blue super competent cells (Agilent catalogue number: 200249).

Buffers used for the procedure are detailed in **Table 2.8**. SOB was autoclaved prior to used and TB was filtered using 0.2mm filter. Centrifuge bottles for bacteria were autoclaved prior to protocol, as were 1.5 ml Eppendorf tubes.

**Table 2.8 Buffers for generation of ultracompetent cells**

<b>1L SOB + glycine</b>	For 1 L, weight and mix: 20g Tryptone, 5g Yeast Extract, 0.5g NaCl, 625ul 4M KCl, 10g Glycine. PH was adjusted to 7 using NaOH.  At room temp, add 10ml of 1M MgCl
<b>100 ml TB</b>	10mM PIPES, CaCl <sub>2</sub> ·2H <sub>2</sub> O 15 mM, KCl 250mM, pH to 6.7 using 150-200ul 10M KOH, add 1.09 g MnCl <sub>2</sub> ·4H <sub>2</sub> O

Frozen XL-Blue cells were spread onto an LB agar plate and incubated overnight at 37 degrees.

250 ml of sterilised SOB + Glycine was inoculated with 10-12 colonies from the agar plate. Cells were incubated at 20 degrees with shaking of 140-160 rpm. Measurement was performed with a spectrometer to determine when the cells had reached OD<sub>600</sub> of 0.94. Doubling rate of the cells took approximately 3 hours.

Flask was removed from incubator and put on ice for 10 minutes. Cells were then pelleted in a sterilised flask at 2500 x g for 10 min at 4°C. Supernatant was removed and cells were gently resuspended in 80ml of ice-cold TB. Cells were left again on ice for 10 min before pelleting at 2500 x g for 10 min at 4°C. Supernatant was removed and cells were resuspended in 10ml ice cold TB. 10ml of TB+14%DMSO (Sigma Aldrich catalogue number: 200244). Cells were incubated on ice for a further 10 mins. A dry ice/ethanol bath was prepared in a Styrofoam box, and sterilised 1.5 ml Eppendorf tubes were pre-cooled in the bath. Cells were aliquoted into Eppendorfs at 500ul per tube, once frozen, cells were quickly transferred to -80 degrees.

#### 2.5.4 Testing of ultracompetent cells.

Efficiency of ultracompetent cells was tested through transformation with pUC18 DNA (Agilent) using the following conditions. 0.1, 1, or 10 pg of pUC18 control DNA was added to one aliquot of cells and tubes were swirled gently. Tubes were incubated on ice for 30 minutes. Tubes were heat pulsed in a 42°C water bath for 45 seconds, then incubated on ice for 2 minutes. 0.9 ml of preheated SOB medium was added and tubes were incubated at 37°C for 1 hour with shaking at 225–250 rpm. Cells were centrifuged at 500 rpm for 5 minutes and 250ul of supernatant was removed. Cells were resuspended in remaining 250 µl of the transformation mixture and plated on LB–ampicillin agar plates. Plates were incubated overnight at 37 degrees.

Guidelines to measure efficiency of ultracompetent cells are shown in Table 2.6 in colony forming units (cfu) per µg

**Table 2.9 Criteria used for generation of ultracompetent cells.**

0.1 pg pBSKS	100 colonies $10^9$ cfu/µg.
1 pg pBSKS	100 colonies $10^8$ cfu/µg.
10 pg pBSKS	:100 colonies $10^7$ cfu/µg.

## 2.6 DNA methods

### 2.6.1 In house DNA sequencing

DNA sequencing Dye terminator sequencing reactions (ABI) were performed and processed by the Institute of Genetics and Molecular Medicine (IGMM) sequencing service on a 3130/3730 genetic analyser (Applied Biosystems). DNA sequencing data was analysed using Sequencher 5.4.6 (Gene Codes Corp.) and SnapGene (Dotmatics)

### 2.6.2 Primers for sequencing L1 plasmids

**Table 2.10 Primers for sequencing L1 plasmids**

Primer	Sequence
-5'UTRs.5'	<u>Cagctcaaggaggcctgcctgc</u>
-5'UTRs5'	<u>Gcaggcaggcctccttgagctg</u>
JA8s5'	<u>Ccaaaagtagataaaacc</u>
-JA8as5'	<u>Ggttttatctacttttgg</u>
-ORF1As5'	<u>Gcctctcctcctccaaaggaacg</u>
-ORF1Aas5'	<u>Cgttcctttggaggaggagaggc</u>
-ORF1Bs5'	<u>Atgggactatgtgaaaagacc</u>
-ORF1Bas5'	<u>Ggtcttttcacatagtcccat</u>
-1598new-s5'	<u>Gcaggccaacgttcagattcagg</u>
-1598new-as5'	<u>Cctgaatctgaacgttggcctgc</u>
-ORF1Cs5'	<u>Ggaagcgctaaacatggaaag</u>
-ORF1Cas5'	<u>Ctttccatgttttagcgcttcc</u>
-IGs5'	<u>Ggaagaaactgcatcaactaatg</u>
-IGas5'	<u>Cattagttgatgcagtttcttcc</u>
-Bgl2s5'	<u>gcaggggttgcaatcctagtctctg</u>
-Bgl2as5'	<u>cagagactaggattgcaaccctgc</u>
-ORF2As5'	<u>Atcaaaagagacaaagaaggc</u>
-ORF2Aas5'	<u>Gccttcttgtctcttttgat</u>
-ORF2Bs5'	<u>Cagaactctccaccccaatc</u>
-ORF2Bas5'	<u>Gatttggggtggagagtctg</u>
-jORF2Cs5'	<u>Gacaccctaacatcacaaattaa</u>

-jORF2Cas5'	<u>Ttaattgtgatgttagggtgtc</u>
-jORF2Ds5'	<u>Ccagatggattcacagccgaattc</u>
-jORF2Das5'	<u>Gaattcggctgtgaatccatctgg</u>
-jORF2Es5'	<u>Caataaattaggattgatgggac</u>
-jORF2Eas5'	<u>Gtcccatcaatacctaatttattg</u>
-jORF2Fs5'	<u>ggaagaatcaatatcgtgaaaatg</u>
-jORF2Fas5'	<u>Catttcacgatattgattcttcc</u>
-jORF2Gs5'	<u>Cacctatacaaaaatcaattc</u>
-jORF2Gas5'	<u>Gaattgattttgtataaggtg</u>
-jORF2Hs.	<u>Gatatcatctcacaccagttag</u>
-jORF2Has.	<u>Ctaactggtgtgagatgatatc</u>
-jORF2Is5'	<u>Gactggattaagaaaatgtg</u>
-jORF2Ias5'	<u>Cacattttcttaatccagtc</u>
-jORF2Js5'	<u>Cattagtgggtgcagcgacc</u>
-jORF2Jas5'	<u>Ggtgcgctgcaccactaatg</u>

### 2.6.3 Small scale purification of plasmid DNA

Plasmid DNA was prepared using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's instructions. DNA was extracted from 5 ml of stationary phase *E. coli* culture 99 and eluted in 50 µl of elution buffer (10mM Tris-HCl pH 8.5) or DNase/RNase-free distilled water (Gibco).

### 2.6.4 Large scale purification of plasmid DNA

For larger scale preparation of plasmid DNA, the ZymoPURE II Plasmid Maxiprep Kit (Zymo Research) was used. DNA was extracted from 100-150 ml of stationary phase *E. coli* following the manufacturer's instructions and eluted in 400 µl of the provided elution buffer

### 2.6.5. Restriction digests

Plasmid DNA was digested with the appropriate restriction endonuclease in the buffer supplied by the manufacturer (NEB or Roche). To ensure complete restriction digestion before subsequent cloning steps, the digest was performed in a 100 µl of buffer with 2–5 µg of DNA and 20 U of the appropriate enzyme(s) overnight at the appropriate temperature, generally 37°C. For double digests, the optimal buffer conditions were selected for both enzymes using the manufacturer's guidelines.

### 2.6.5. Purification of restriction digested DNA

DNA fragments produced by restriction digestion were resolved by agarose gel electrophoresis. The desired DNA fragment was excised from the gel using a scalpel and purified using the QIAquick Gel Extraction kit (Qiagen) according to the manufacturer's instructions. DNA was eluted in 30 µl elution buffer and stored indefinitely at –20°C.

### 2.6.6 Ligation of digested DNA

Ligation of DNA molecules Ligation reactions contained: 100–200 ng of vector DNA, 2–3X this molar amount of insert DNA, 1 U T4 DNA Ligase (Roche) and 1X Ligation Buffer (Roche). Following incubation for 4–5 h at room temperature, 1 µl ligation mixture was used to transform *E. coli* (Section 2.2.2.2)



## 2.7 Mass Spectrometry

Immunoprecipitation samples were prepared for mass spectrometry as described in Methods section 2.1

The sample were incubated for 3 h at 37°C and stopped by freezing on dry ice for subsequent liquid chromatography (LC)-tandem MS (MS/MS; performed by Alex Von Kriegsheim, IGMM Mass Spectrometry facility; Section 2.5.11.2).

To prepare samples for MS analysis, an 8 M urea buffer with 100 mM Tris (pH 8.5) was added to samples to a final concentration of 2 M urea, followed by addition of a 1 M DTT to a final concentration of 10 mM. Samples were evaporated in a SpeedVac Concentrator™ (Thermo Scientific), resuspended in 12 µl 0.1% TFA buffer and analysed by MS. The tryptic peptides were analysed on a mass spectrometer connected to an Ultimate Ultra3000 chromatography system (Thermo Scientific) incorporating an autosampler. For each sample, 5 µl of the tryptic peptides were loaded on a homemade column (100 mm length, 75 µm inside diameter) packed with 1.9 µm ReprosilAQ C18 (Dr. Maisch HPLC) and separated by an increasing acetonitrile gradient (3%–32%) at a flow rate of 250 nl/min

The mass spectrometer was operated in positive ion mode with a capillary temperature of 220°C, with 2,000 V applied to the column. Data were acquired with the mass spectrometer operating in automatic data-dependent switching mode, selecting the 12 most intense ions prior to tandem MS (MS/MS) analysis. Mass spectra were analysed and label-free quantitation performed using the MaxQuant Software package (Tyanova et al., 2016).



## Chapter 3

---

# **Optimising a cultured cell retrotransposition assay in Chinese hamster cell lines**



### 3.1 Introduction

This thesis aims to investigate L1 retrotransposition in the context of Fanconi Anaemia. To do this, I will test the frequency at which a series of mutant L1 elements move in cell lines that exhibit a FANC phenotype. In this way, I will be able to probe the mechanism of retrotransposition in a FANCC context by testing how mutating different domains of the element impact the efficiency of retrotransposition in these cells. The cultured cell retrotransposition assay will therefore be a fundamental technique in this thesis.

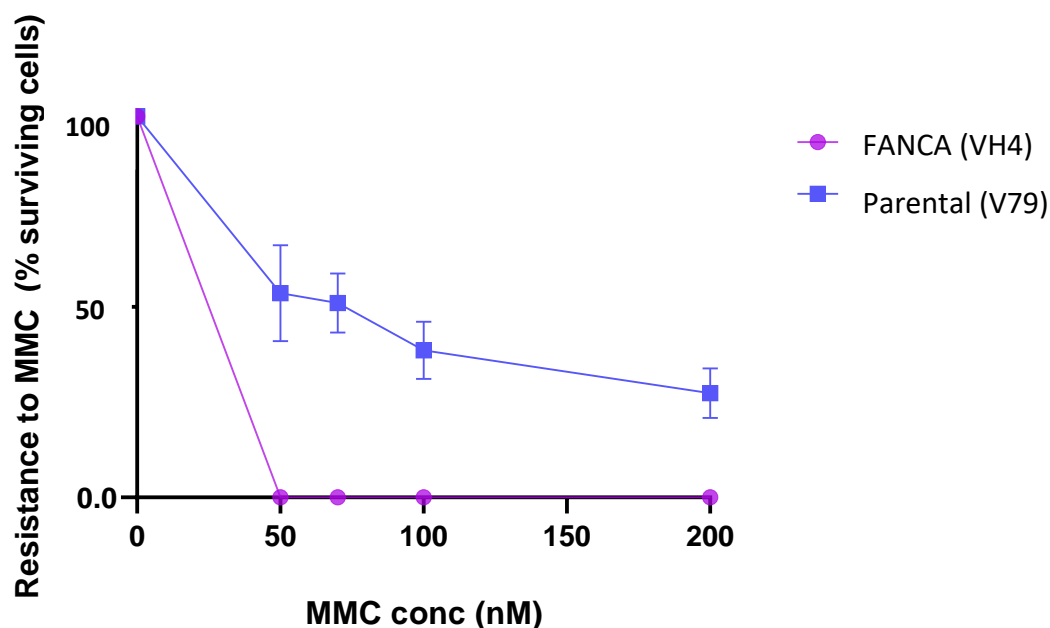
The retrotransposition assay is a highly sensitive tool that can be used to quantify efficiency of mobility in different contexts (Moran *et al.* 1996). It is a versatile tool, and has been vital in determining domains of L1 which are vital for retrotransposition. Different factors, such as transfection efficiency or seeding density, can alter the sensitivity of the assay and should be considered (Moran *et al.* 1996; Kopera *et al.* 2016). In this chapter, I aim to optimise the assay in several key cell lines, in order to ensure that the foundational experiments I perform are robust. Furthermore, as I plan to investigate whether the FANC phenotype alters the capacity of mutant L1 elements to move, I want the assay to be sufficiently sensitive to pick up rare retrotransposition events.

My foundational experiments were performed in two cell lines, VH4 and V79. V79 is a Chinese hamster ovary (CHO) cell line (Jones *et al.* 1994). A mutation of that FANCA gene in this line generated the VH4 cells (Overkamp *et al.* 1993; Telleman *et al.* 1996). As FANCA is mutated in 65% of FANC patients (Morgan *et al.*, 1999)(Solomon *et al.* 2015), I chose to investigate L1 retrotransposition first in these cells. Furthermore, having the parental line for comparison provides a good control, and we know that CHO cells support high levels of retrotransposition. In addition, previous work on endonuclease independent retrotransposition was performed in hamster cells. making them an excellent context in which to perform these experiments.

### 3.2 Cells deficient in Fanconi anaemia are hypersensitive to Mitomycin C

Due to the role of the Fanconi Anaemia pathway in repairing ICLS, cells deficient in FANC proteins exhibit a well-documented hypersensitivity to Mitomycin C (MMC; Overkamp *et al.* 1993). Treatment of FANCC cells with MMC results in an increase in apoptosis, chromosome breakage and cell cycle arrest at G2 (Zdzienicka *et al.* 1990).

To investigate L1 retrotransposition in FANCC cells, I used VH4 Chinese hamster cells which are deficient in FANCA (Overkamp *et al.* 1993; Telleman *et al.* 1996). The cell line has been reported to exhibit a Fanconi Anaemia phenotype, due to the loss of function of the protein (Zdzienicka *et al.* 1990). These cells are derived from the Chinese hamster cell line V79, which has been extensively used to investigate DNA damage and repair (Bradley *et al.* 1981; Chaung *et al.* 1997; Boorstein *et al.* 1992). The cell line was key for foundational experiments in this research because FANCA is mutated in 65% of FANC patients and also because hamster cells support a high level of L1 retrotransposition (Morrish *et al.* 2002).



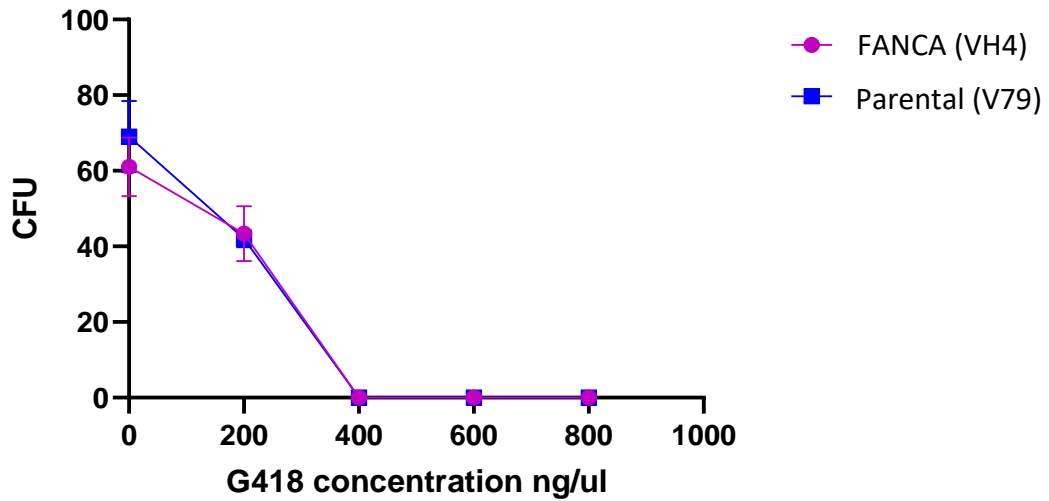
**Figure 3.1: Sensitivity of FANCA deficient (VH4) cells to MMC in comparison with the parental (V79) cell line.** A cell line deficient in FANCA and a parental line were tested for their relative sensitivity to crosslinking agent MMC. After a 24-hour treatment with MMC, surviving cells were fixed, stained and quantified using a spectrophotometer. Error bars indicate standard error of the mean (SEM) and this experiment was performed three times.

As expected, FANCA cells demonstrated a much higher sensitivity to MMC than the parental cell line with a functional Fanconi Anaemia pathway (**Figure 3.1**). This confirms that the cell line used is reflecting the expected Fanconi phenotype.

### 3.3 Optimisation of parameters for a retrotransposition assay in FANCA deficient cells.

To test L1 retrotransposition in FANCA deficient cells, a cultured cell retrotransposition assay was used (as described by Moran *et al.* 1996). The retrotransposition assay is a sensitive tool, which is highly useful for quantifying retrotransposition in different cell lines and testing the mobility of different elements. In order to optimise the sensitivity of the assay, however, several variables need to be controlled for. As I planned to carry out several critical experiments in a FANCA deficient cell line (VH4) and a parental line (V79). I performed a series of experiments to optimise the retrotransposition assay in these cell lines.

Firstly, I performed a death curve to optimise the concentration of antibiotic to be used in the assay (**Figure 3.2**). Critically, the antibiotic should be of a high enough concentration that cells which don't express the resistance gene are killed within the 12-day selection period. If, however, the concentration is too high, then even cells expressing the reporter gene may be adversely affected.



**Figure 3.2: Death Curve of FANCA deficient (VH4) and parental (V79) cells using G418.** A death curve was performed in FANCA deficient cells and the parental line to establish the optimal concentration of antibiotics to use in the cultured cell retrotransposition assay. Cells were treated with different concentrations of G418 for a selection period of 12 days, with media changes on alternating days. Error bars represent the SEM and the experiment was repeated three times. The lowest concentration that resulted in complete cell death was selected for future experiments, in this case 400ng/ul for both cell lines.

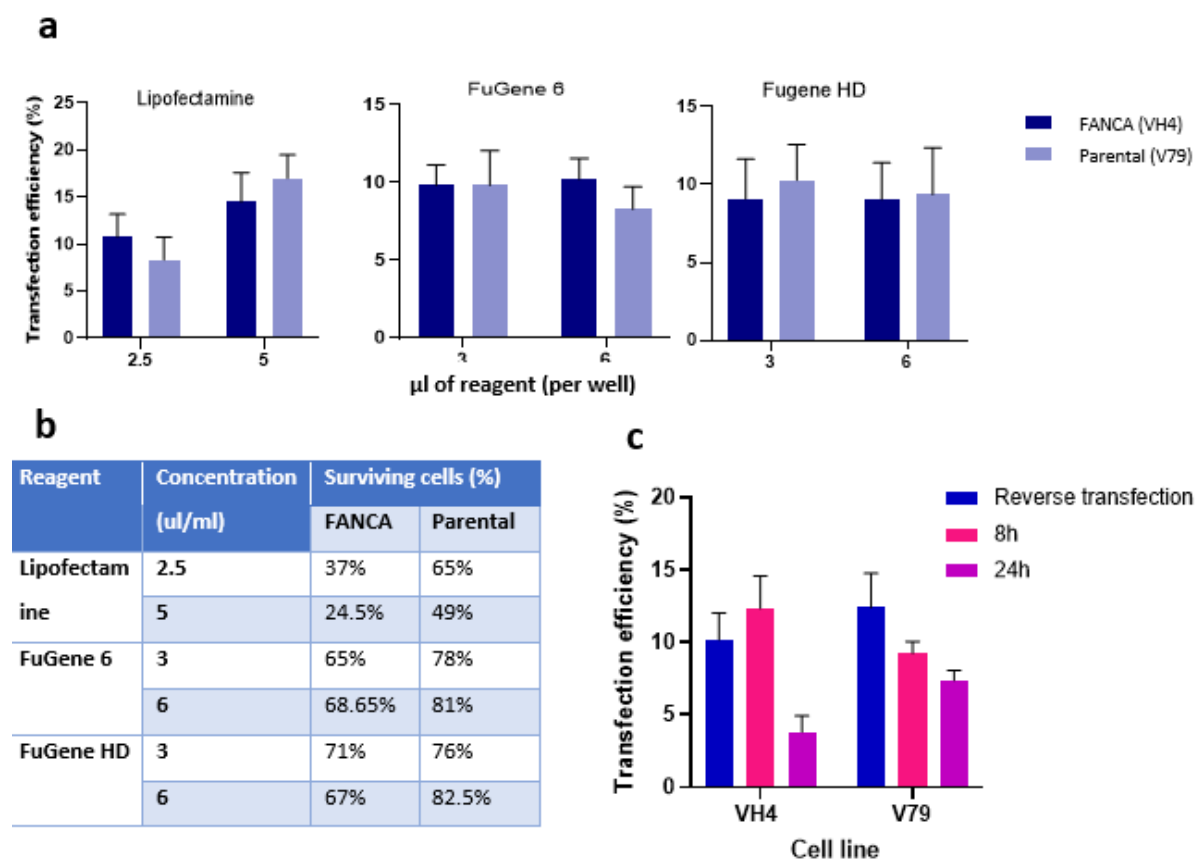
Based on the death curve shown in **Figure 3.2** I decided that 400ng/ul was an optimal choice of G418 concentration to use for the retrotransposition assay. This was the minimal concentration that killed all cells after a 12-day period of selection.



### 3.4 Optimising the transfection of VH4 and V79 cells.

In initial experiments with the VH4 and V79 cells, transfection efficiency was lower than expected (1-5%). As a high transfection efficiency improves the sensitivity of the assay to detect low levels of retrotransposition, I tested a range of conditions to maximise the number of cells transfected for future experiments. Three widely used reagents were tested, each at two concentrations (**Figure 3.3**). When measuring transfection efficiency, cell count was also taken into account to determine the degree of toxicity following transfection. Though transfection was more efficient using Lipofectamine 2000, this reagent caused notably more toxicity than FuGene6, as can be seen by the lower cell count at FACS, despite the same seeding density. Transfection with either FuGene6 or FuGene HD generated a similar efficiency.

Further experimentation showed that an 8h time delay between plating and transfection provided better transfection efficiency than a 24 h time delay. Similarly, a reverse transfection protocol, where cells were immediately transfected after plating, improved transfection efficiency.



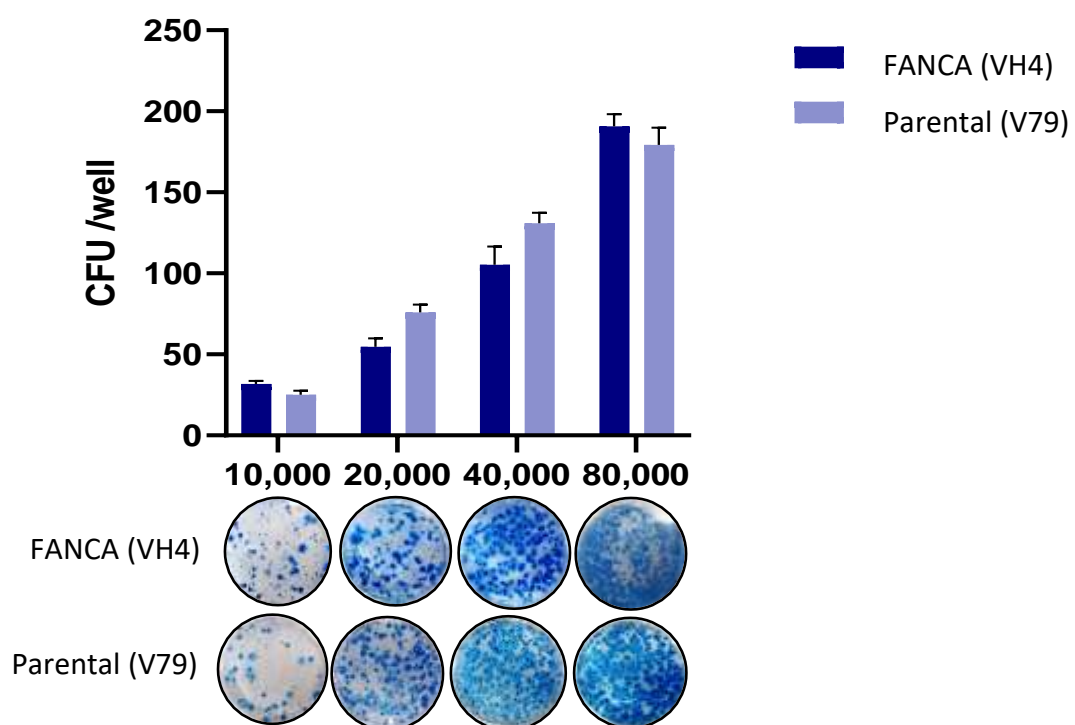
**Figure 3.3: Optimisation of transfection efficiency in FANCA (VH4) and parental (V79) cell line.**

- Transfection efficiency following the use of three different reagents, at two concentrations, in FANCA mutant and parental cell line
- Cell count measured by FACS, following the use of three different reagents, at two concentrations in FANCA mutant and parental cell line.
- Transfection efficiency in FANCA mutant and parental cell line with different time delays between plating and transfection.

### 3.5 Optimising the seeding density

Thirdly, I performed an assay to determine the optimal seeding density in the FANCA deficient and parental cell line. A density should be chosen that is not too high, or surviving colonies will be very densely packed on the plate and less distinct from each other, making them difficult to count accurately. If the density is too low, the cells may be too sparse and the assay will be less sensitive to lower levels of retrotransposition (Moran *et al.* 1996).

To do this, I plated both FANCA deficient and parental cell line at four different seeding densities and transfected with a wild type LINE-1 element. The results are shown in **Figure 3.4**.



**Figure 3.4: Optimisation of seeding density in FANCA deficient and parental cell line**

The number of colony forming units (CFU) following a retrotransposition assay (introduced in **Figure 1.5**) performed with four different seeding densities (x-axis). Either FANCA deficient or parental cells were transfected with a wild type L1 element. G418-resistant colonies were visualised by staining with crystal violet. CFU: colony forming units.

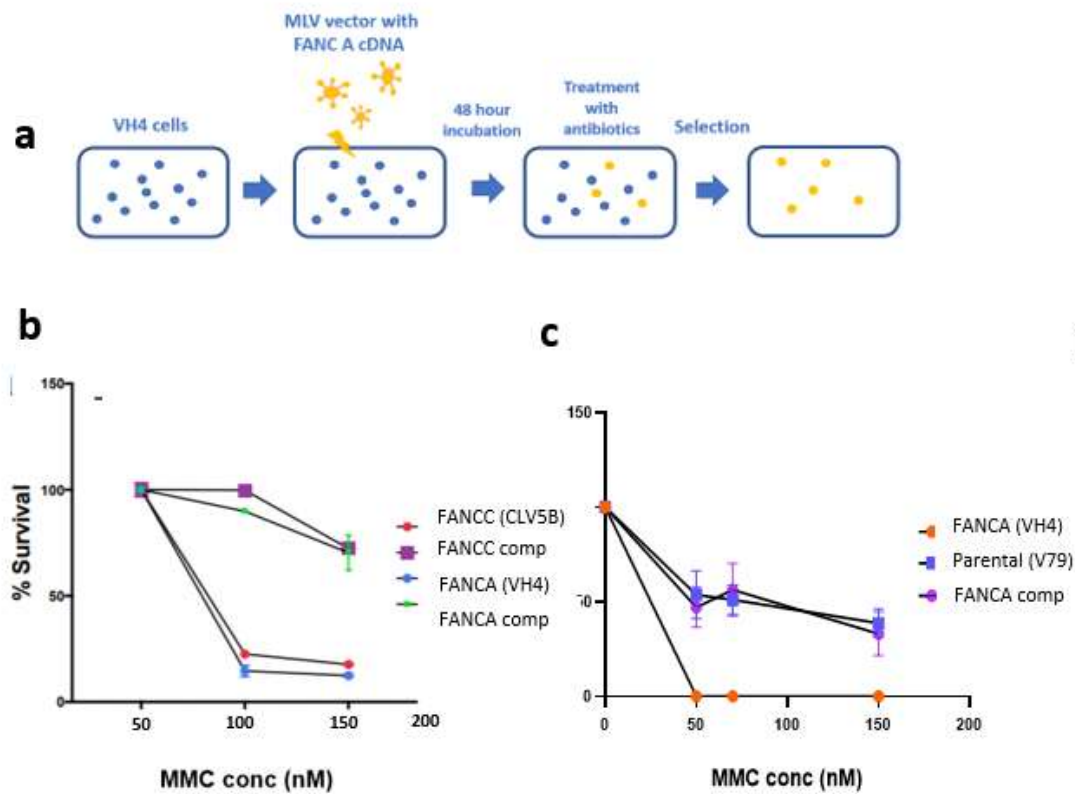
Following this series of assays, I concluded that 20,000 cells was an optimal seeding density for this cell line. Lower densities may make it more difficult to detect low-efficiency retrotransposition, and higher numbers may make it more difficult to accurately count colonies and therefore assess relative retrotransposition rates.

### 3.6 Optimising retrotransposition cell culture assay in FANCA complemented and NHEJ deficient cell lines.

A key experiment of this thesis is the comparison of retrotransposition in FANCA deficient cells and FANCA deficient cells complemented with FANCA. This control confirms that any unusual retrotransposition in FANCA deficient cells results from the absence of FANCA and not from secondary DNA damage associated with the FANC phenotype.

The FANCA cell line was generated by members of the Jose Garcia Perez lab, by infection with a Murine Leukaemia Virus (MLV)-derived vector expressing the mouse FANCA cDNA. 48 hours after infection, antibiotic selection was used to select for cells expressing FANCA.

For these experiments I also include CLV5B (a FANCC cell line which is also derived from the parental V79 cells, similarly to VH4 cells) for comparison, as well as CLV5B cells complemented with FANCC. As these cells share a parental line with VH4 cells, as well as a hypersensitivity to MMC (Morell 2016).



**Figure 3.5: Development of a cell line complemented with FANCA.**

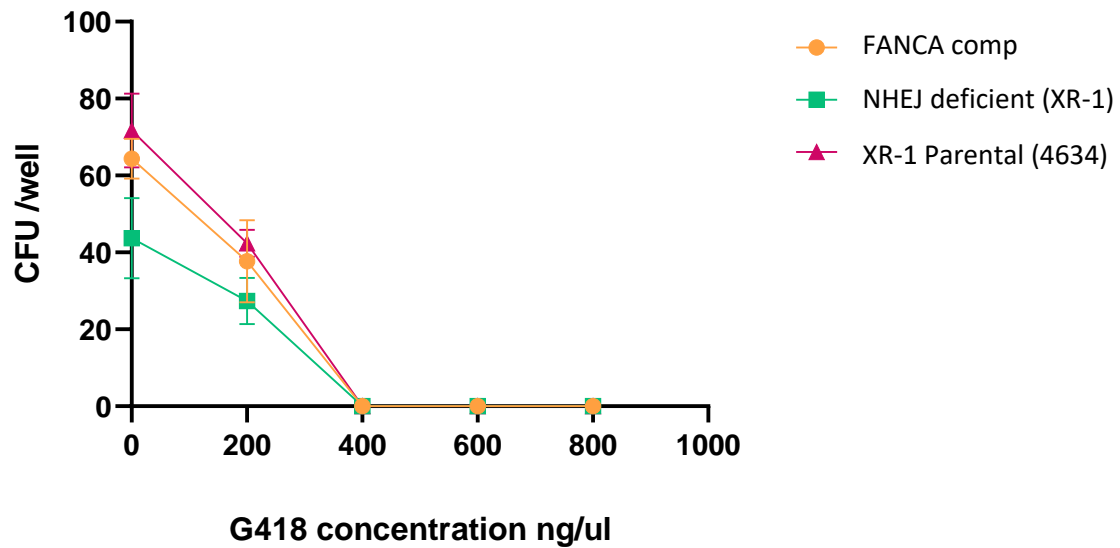
Error bars in all figures represent SEM and each experiment was performed three times

- Scheme of how FANCA complemented cell line was developed. This work was performed by members of the JGP lab.
- MMC sensitivity assay performed in FANCA (VH4) and FANCA comp cells after generation of FANCA complemented cell line in comparison with FANCC deficient (CLV5B) and FANCC complemented cell line.
- MMC sensitivity assay showing FANCA complemented cells in comparison with FANCA (VH4) and parental cell line.

Following generation of the cell line, the FANC phenotype was confirmed with a MMC sensitivity assay. FANCA complemented cells showed a reduced sensitivity to MMC comparable to other complemented cell lines, including FANCC deficient CLV5B (**Figure 3.5B**) and the parental cell line, V79.

As this cell line is part of key experiments to this thesis, I also optimised the retrotransposition assay in these cells. Alongside these experiments, I also optimised the assay in a NHEJ defective cell line, XR-1, and its parental line 4364. This CHO hamster cell line contains a mutation in XRCC4 which renders it defective in double strand break repair and V(D)J recombination (Giaccia *et al.* 1990; Zdzienicka *et al.* 1992; Pergola *et al.* 1993). This cell line is key in this thesis for distinguishing FANC specific effects from effects resulting from other defects in DNA damage repair, thus it was important to optimise the experiments in this cell line.

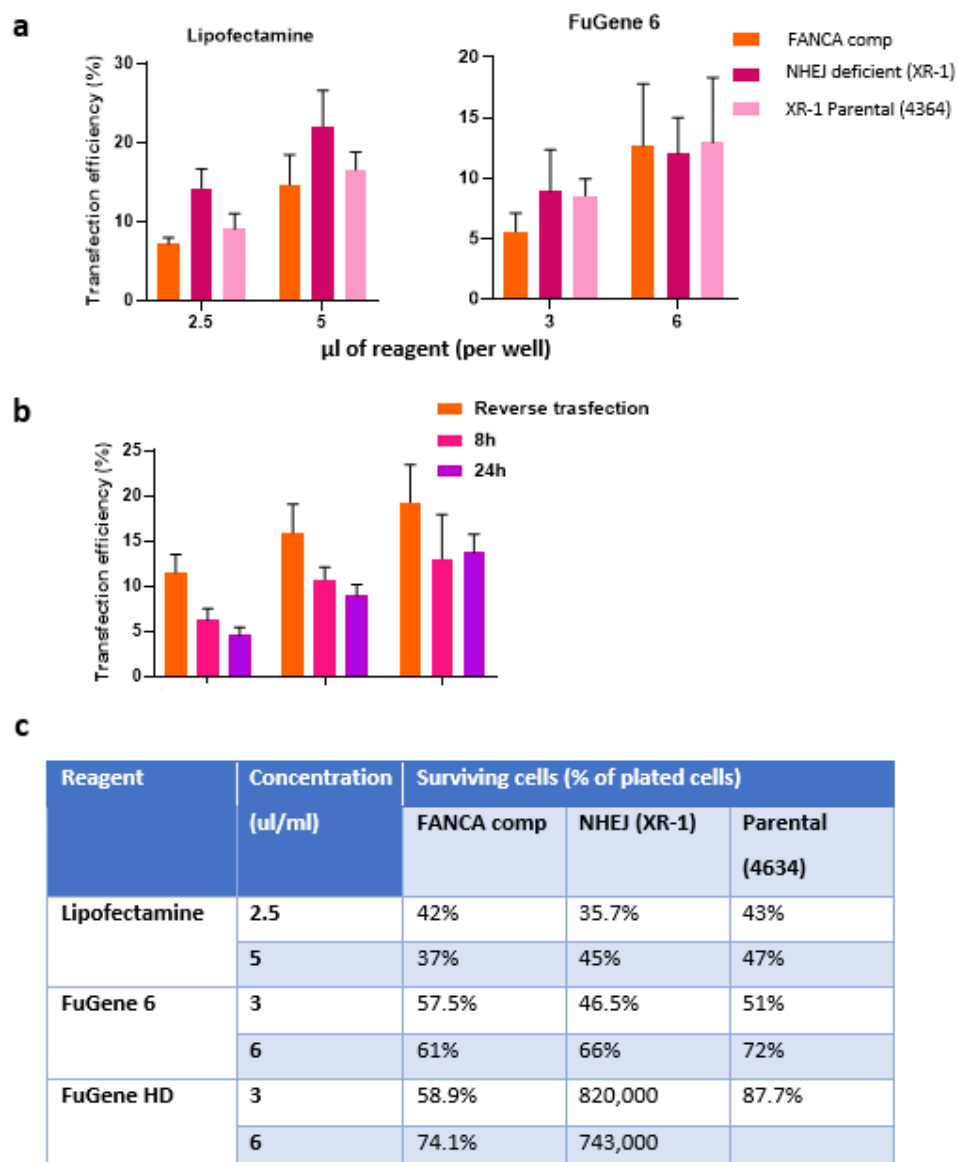
Firstly, I performed a death curve using the antibiotic G418, in order to determine which concentration is sufficient for the selection stage of the assay.



**Figure 3.6: Death curve to identify the optimal antibiotic concentration in FANCA complemented cells, NHEJ deficient (XR1) and parental (4364) cells.** XR1 and 4364 cell lines were treated with G418 for 12 days, when the number of colony forming units was counted. Error bars represent the SEM and this experiment was performed three times.

As previously observed with the VH4 and V79 cells, 400ng/ul is the lowest concentration of G418 that kills all plated cells within a 12-day selection period, and was therefore used for subsequent assays with these cell lines.

I also performed a series of experiments to maximise the transfection efficiency of these three cell lines.



**Figure 3.7: Optimisation of transfection efficiency in FANCA complemented, NHEJ deficient (XR-1) and parental (4364) cell lines.** Each experiment was performed three times and error bars represent the SEM.

- Transfection efficiency following the use of three different reagents, at two concentrations, in FANCA comp, NHEJ mutant (XR-1) and the XR-1 parental (4364) cell lines
- Transfection efficiency in FANCA comp, NHEJ mutant (XR-1) and the XR-1 parental (4364) cell lines with different time delays between plating and transfection.
- Cell count measured by FACS, following the use of three different reagents, at two concentrations in FANCA comp, NHEJ mutant (XR-1) and the XR-1 parental (4364) cell lines



As previously observed with the CHO cell lines VH4 and V79, Lipofectamine 2000 was the reagent that generated the highest transfection efficiencies. However, as can be seen from the cell count at FACS (**Figure 3.7b**) a low number of cells survived the transfection, indicating high toxicity. FuGene6 offers a good efficiency with a lower degree of toxicity to the cells. Notably, reverse transfection, where cells were transfected immediately after plating, increased the transfection efficiency in all three cell lines, particularly in XR-1 and 4364. In addition, 4364 consistently exhibited a lower transfection efficiency than the XR-1 cell line, emphasising the importance of normalising results from the retrotransposition assay by transfection efficiency when comparing between cell lines.

### 3.7 Main Conclusions

Primarily, this work has determined the optimal parameters for cell-based assays in five cell lines that are key to my thesis. These findings provide a solid basis for subsequent experiments, ensuring that the assay is sensitive enough to detect rare retrotransposition events and effectively quantify variations in L1 activity.

### 3.8 Discussion

In this chapter, I discuss the optimisation of the cell based retrotransposition assay in five cell lines that are key to this thesis. As cell-based assays form the basis of much of my experimental work, establishing the parameters to optimise their sensitivity is an important foundation for my thesis.

Initially, I observed low transfection efficiency with the VH4 and V79 cell lines. As previously mentioned, low transfection efficiencies can limit the sensitivity of the assay, as rare or low-efficiency retrotransposition events will only be recorded if a high number of cells express the L1 element. Reducing the time delay between plating and transfection to 8h increased the average transfection efficiency in these CHO cell lines, as did performing a reverse transfection.

I also observed some variation in how amenable different cell lines were to transfection. XR-1, for example, appeared to transfect more easily than the cell line it was derived from, 4364. This difference highlights the importance of normalising retrotransposition rate according to the transfection efficiency for each plasmid. To account for this and other differences between cell lines, for example

the rate of proliferation, measures of retrotransposition are taken relative to the retrotransposition rate of a wild type element in that cell line.

This series of experiments also showed the importance of seeding density to the outcome of the assay. As seen in **Figure 3.4**, a high seeding density results in the cells being very densely packed on the plate and can make counting distinct colonies difficult. For the cell lines tested in this chapter, seeding 20,000 cells appears to generate a good number of wild type colonies.

## Chapter 4

---

# **Retrotransposition of mutant L1 elements in a FANCA mutant background**



## 4.1 Introduction

For a L1 element to mobilise through canonical TPRT, both open reading frames, ORF1 and ORF2 are required. The reverse transcriptase activity of ORF2 is an absolute requirement for retrotransposition and typically the EN domain of ORF1p is also required (Mathias *et al.* 1991; Moran *et al.* 1996). Despite the role of ORF1 in TPRT not yet being fully established, it is known that the protein is a nucleic acid chaperone, and abolition of the chaperone activity also abrogates retrotransposition (Kolosha and Martin 2003; Martin *et al.* 2005). In addition to this, mutations which prevent ORF1p from binding back to its transcript to form the ribonucleoprotein particle also prevent efficient retrotransposition. A key paper by Kulpa and Moran (2005) demonstrated this, and established the RNP as a necessary retrotransposition intermediate.

In recent years, an alternative mechanism of retrotransposition called endonuclease independent retrotransposition (ENi), has been identified. In 2002, Morrish *et al.* demonstrated that LINE-1 elements can mobilise without needing a functional EN domain to create an incision in the genomic DNA. Experimentation using cultured cell retrotransposition assays showed that the frequency of ENi retrotransposition increases in cells which are deficient in DNA repair (For example XR-1 cells, which harbour XRCC4 mutations and lack DSB repair). L1 elements without a functional EN domain were able to mobilise with a relatively high efficiency in these cells, leading the group to hypothesise that the EN deficient elements were exploiting unrepaired gaps in the genomic DNA as insertion sites.

This work has demonstrated that in certain cellular contexts, different L1 domains might be required for retrotransposition. Not only this, but there is a body of work demonstrating that some DNA repair pathways might be directly be involved in the regulation of L1 retrotransposition (Mita *et al.* 2020; Liu *et al.* 2018). It is evident that the complex relationship between DNA repair proteins and L1 retrotransposition remains to be better established.

To further investigate the relationship between DNA repair and L1 retrotransposition in the human genome, we decided to explore the nature of retrotransposition in FA cells.

The FA pathway is primarily concerned with the repair of ICLs, and cells mutated in this pathway exhibit a hypersensitivity to crosslinking agents such as MMC. As well as this function, there is evidence that some of the 21 proteins in the FA pathway have additional roles. Mutations in the FANCA gene account for nearly 65% of all known FA cases (Morgan *et al.* 1999)(Solomon *et al.* 2015), though a broad range of mutations in the FA pathway have been identified in patients (Neveling *et al.* 2009).

Recently, a whole –genome siRNA screen using image based retrotransposition assays in HeLa cells, identified 15 members of the FA family, including FANCA, as potential regulators of L1 retrotransposition (Mita *et al.* 2018). Most FA factors operated as inhibitors of L1 activity, though a small number appeared to enhance retrotransposition levels. The group validated FANCA and 10 other bone fide FA proteins as inhibitors of retrotransposition. A previous genome-wide CRISPR Cas9 screen by Liu *et al.* (2017), had already identified several of these proteins, including FANCA, as inhibitors of retrotransposition.

I therefore wanted to investigate L1 retrotransposition in FA cells for several reasons: to test the possibility that FA proteins are regulators of L1 activity, to identify whether ENi retrotransposition takes place in FA cells, and to determine if there are other domains which are not required for retrotransposition in a FA context. The aim of these experiments is to improve our understanding of the relationship between L1 retrotransposition and FA repair proteins. As FANCA has been identified previously as a regulator of L1, and is mutated in a majority of FA patients (Morgan *al.*, 1999) I decided to test a series of L1 elements for their retrotransposition rates in FANCA cells and a parental line.

## 4.2 Research Aims

In summary, the main research aims of this chapter are threefold:

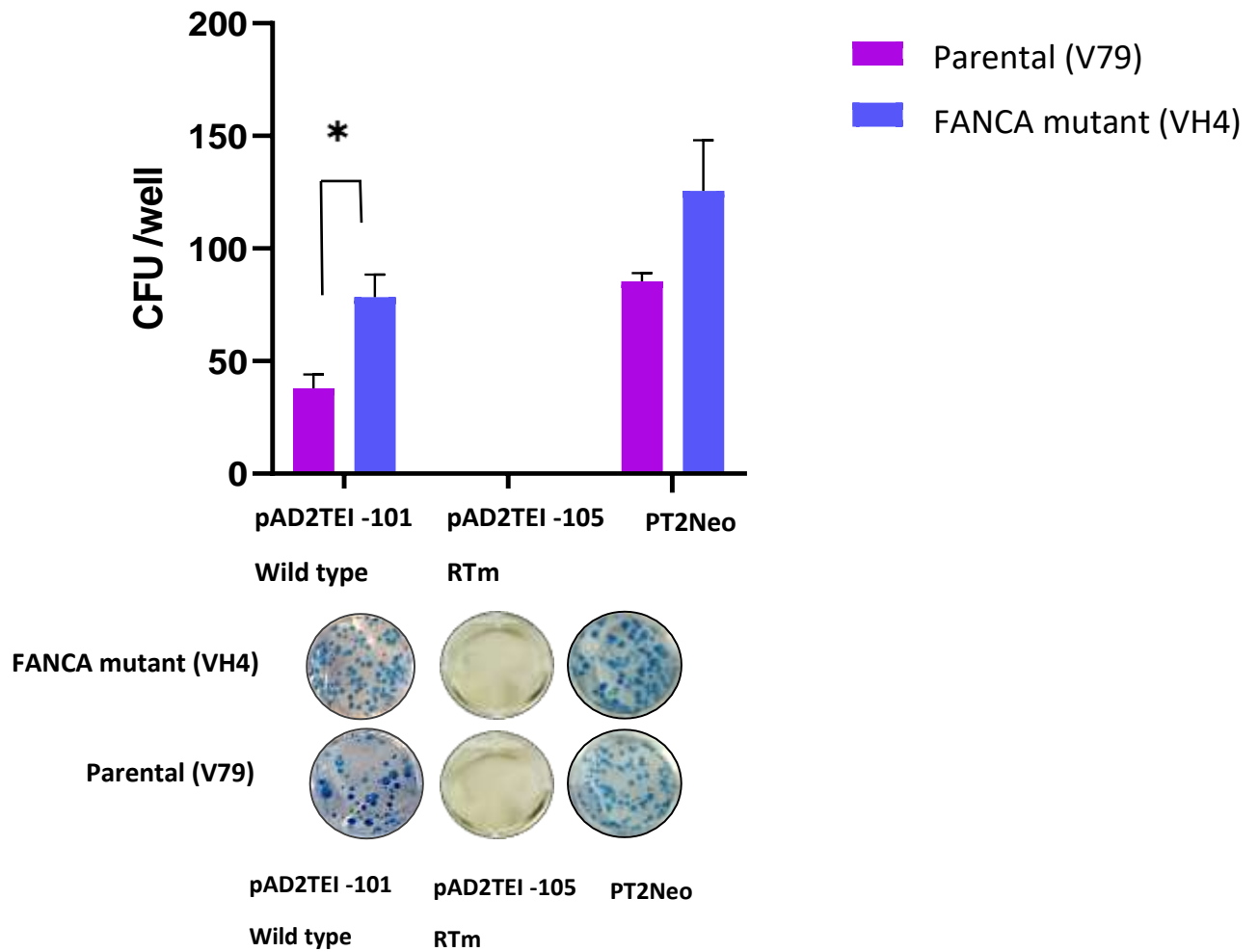
- A) To identify how loss of FANCA impacts RC-L1 retrotransposition
- B) To determine whether ENi takes place in a FANCA deficient context
- C) To identify which other L1 domains are required for retrotransposition in FANCA deficient cells.

### 4.3 Testing wild type retrotransposition in FANCA (VH4) and parental V79 cells.

As several studies have indicated that FANCA could be a potential suppressor of L1 retrotransposition (Liu *et al.* 2017; Tristan Ramos *et al.* 2020; Mita *et al.* 2020), I wanted to test how knocking out FANCA impacts the retrotransposition rate of a wild type L1 element. We hypothesised that deletion of FANCA would result in a higher rate of retrotransposition of a wild type element.

To test this hypothesis, I used a cultured cell retrotransposition assay to measure the activity of JM101/L1.3, a retrotransposition competent element cloned into a pCEP4 vector. The activity of the cassette was measured in two CHO cell lines: V79 and a FANCA deficient derivative, VH4. These cell lines were used because they support a high level of retrotransposition, and because having access to the parental cell line enables a better control for cellular differences.

The results are shown in **Figure 4.1**. After adjusting for variation in transfection efficiency between the two cell lines and between the different plasmids, the WT L1 element is approximately 45% more active in the FANCA mutant line than the WT parental. This suggests that FANCA can act as a negative regulator of L1 retrotransposition in Chinese hamster cells. During the preparation of this thesis, a study by Tristan Ramos *et al.* (2020) showed that depleting FANCA by siRNA significantly increased levels of L1 retrotransposition in U2OS cells, without increases in toxicity or clonability. Using a different method, Liu *et al.* (2018), also found that knocking out FANCA (using a CRISPR/Cas9 genome wide screen) lead to elevated L1 retrotransposition, this time in K562 cells (a human immortalised leukaemia cell line (Rutherford *et al.* 1981)). Therefore, my results support findings in the wider literature that FANCA acts as a suppressor of L1.



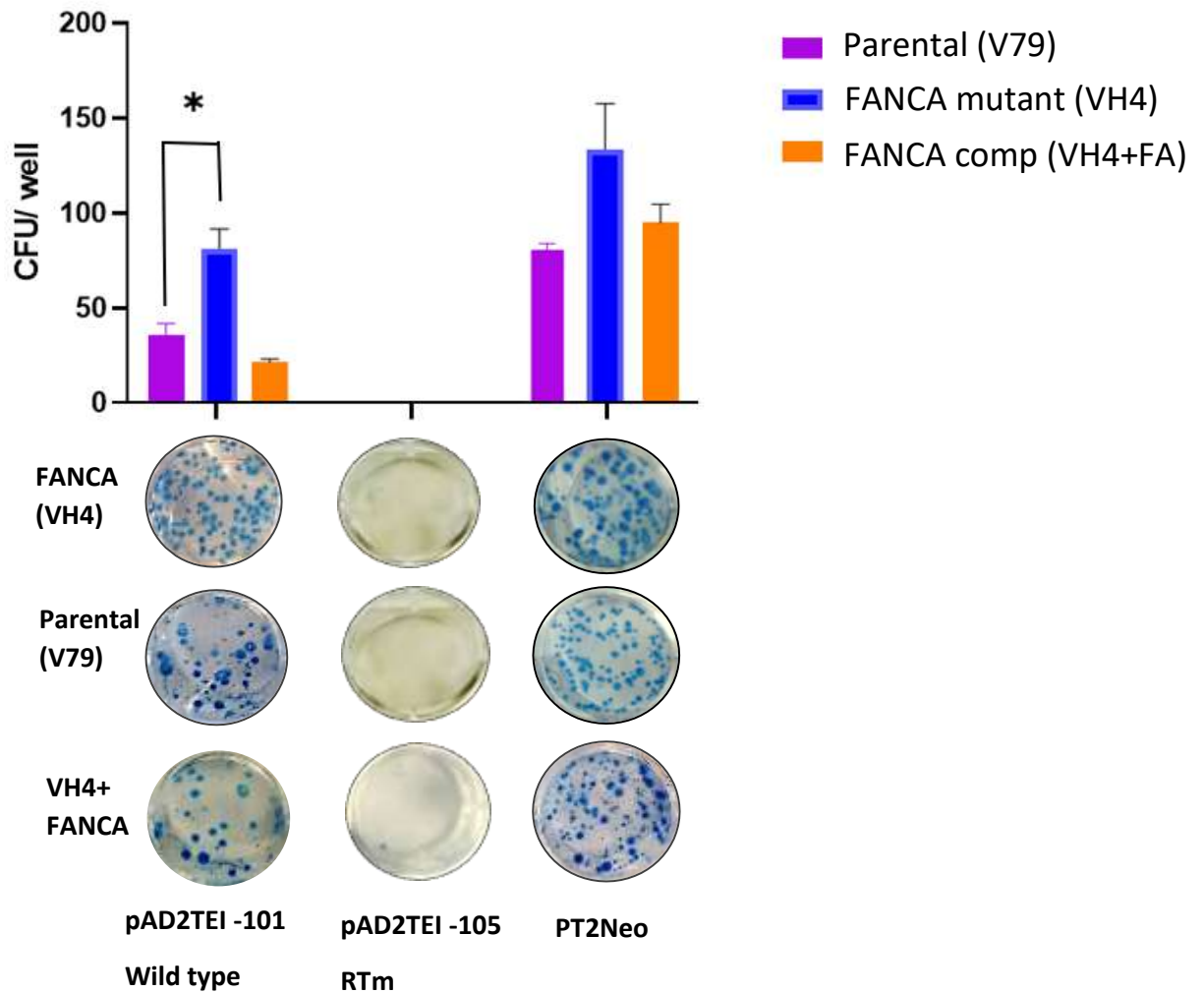
**Figure 4.1: Testing wild type L1 retrotransposition in FANCA (VH4) and parental (V79) cells**

Quantification of L1 retrotransposition in a Chinese hamster cell line (V79) and a derivative deficient in FANCA (VH4) demonstrated that FANCA cells support significantly higher levels of wild type retrotransposition. Quantification of retrotransposition of a synthetic WT L1 element (pAD2TEI -101), an RT mutant (RTm; pAD2TEI-105) and PT2Neo, a plasmid which includes only the reporter gene (without an L1 transcript) are shown. The number of colony forming units (CFU) per well was quantified and a mean was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times. \* indicates  $p < 0.05$ , Student's t-test. Representative plates showing neomycin-resistant colonies stained with crystal violet are shown below the graph.

To control for any secondary effects due to DNA damage resulting from the FA phenotype, I wanted to see if I could rescue the observed effect using VH4 cells that had been complemented with FANCA. The purpose of this experiment was to confirm that the elevated retrotransposition was specifically a result of loss of FANCA. To do this, I performed the same assay as previously described with the FANCA complemented cell line. The results are shown in **Figure 4.2**.



Complementation with FANCA significantly reduced retrotransposition of wild type L1, indicating that expression of the protein successfully rescued the elevated L1 retrotransposition shown in VH4 cells. This indicates that the effect is a direct result of FANCA loss and not a secondary effect of DNA damage.



**Figure 4.2: Complementation with FANCA rescues elevated retrotransposition in VH4 cells.**

The retrotransposition of wild type element L1.3 was quantified by a cultured cell retrotransposition assay in a Chinese hamster cell line (V79), a derivative deficient in FANCA (VH4) and VH4 cells complemented with FANCA. The number of colony forming units (CFU) per well was quantified and a mean was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times. VH4 and V79 data was plotted in Figure 4.1 and is shown again for comparison. \* indicates  $p < 0.05$ , Student's t-test. Representative plates showing neomycin-resistant colonies stained with crystal violet are shown below the graph.

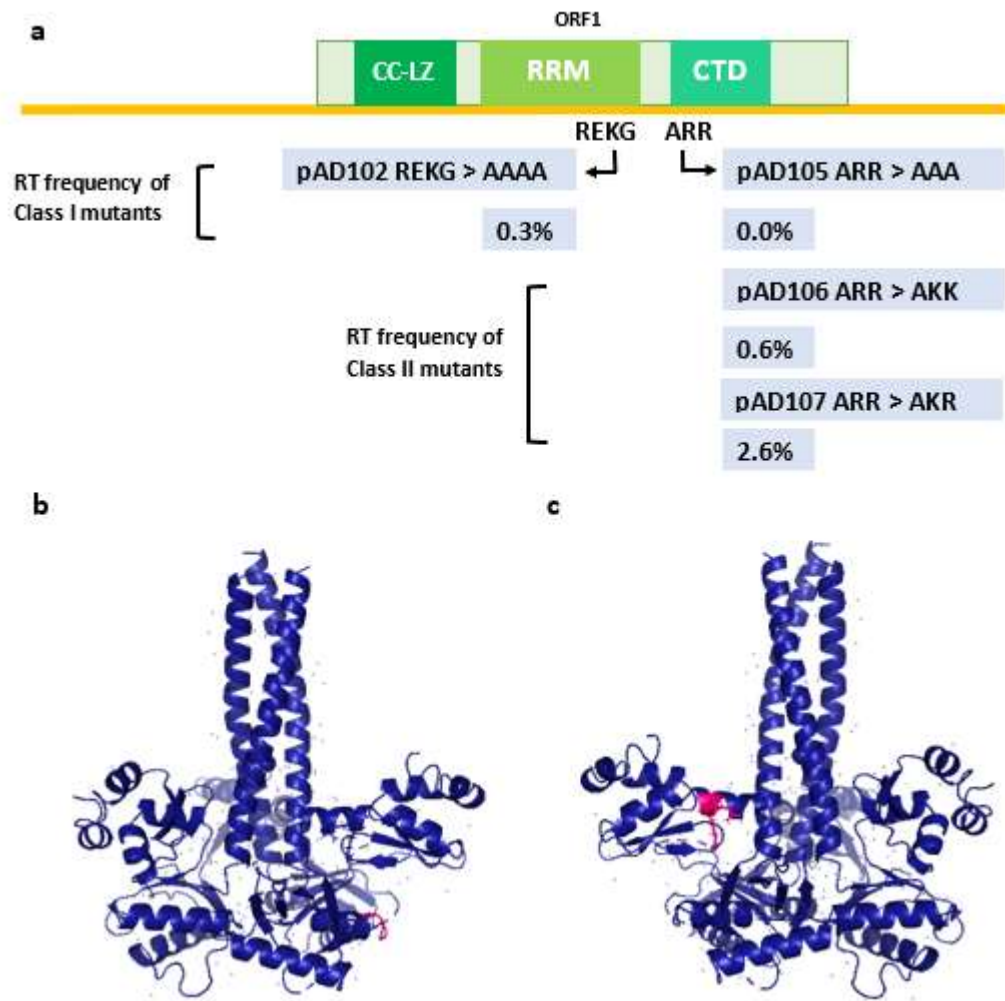
#### 4.4. Testing a battery of ORF1p mutants in FANCA mutant cells.

The first section of this chapter indicated that FANCA could serve as a negative regulator of wild type L1 retrotransposition. Next, I wanted to identify whether retrotransposition in a FA context required the same domains that are fundamental for TPRT. There is precedence for this, as key studies have shown that the mechanism of mobility can be altered in cell lines deficient in DNA repair proteins (Moran *et al.* 2007; Moran *et al.* 2005). As such, I was interested to see whether the deletion of FANCA would enable retrotransposition of elements which would otherwise be retrotransposition incompetent.

I began this line of investigation with a focus on ORF1. In 2005, Kulpa and Moran tested a series of ORF1p mutant L1 elements for their capacity to retrotranspose in HeLa cells. They identified two classes of mutants. Class I mutants derived from the REKG or ARR motif prevented the correct formation of the RNP. Following a process of differential centrifugation, ORF1p could no longer be detected in the centrifugal fraction which contained the particle. These mutations abrogated retrotransposition of the element, which aligns with data indicating that the RNP is an essential retrotransposition intermediate.

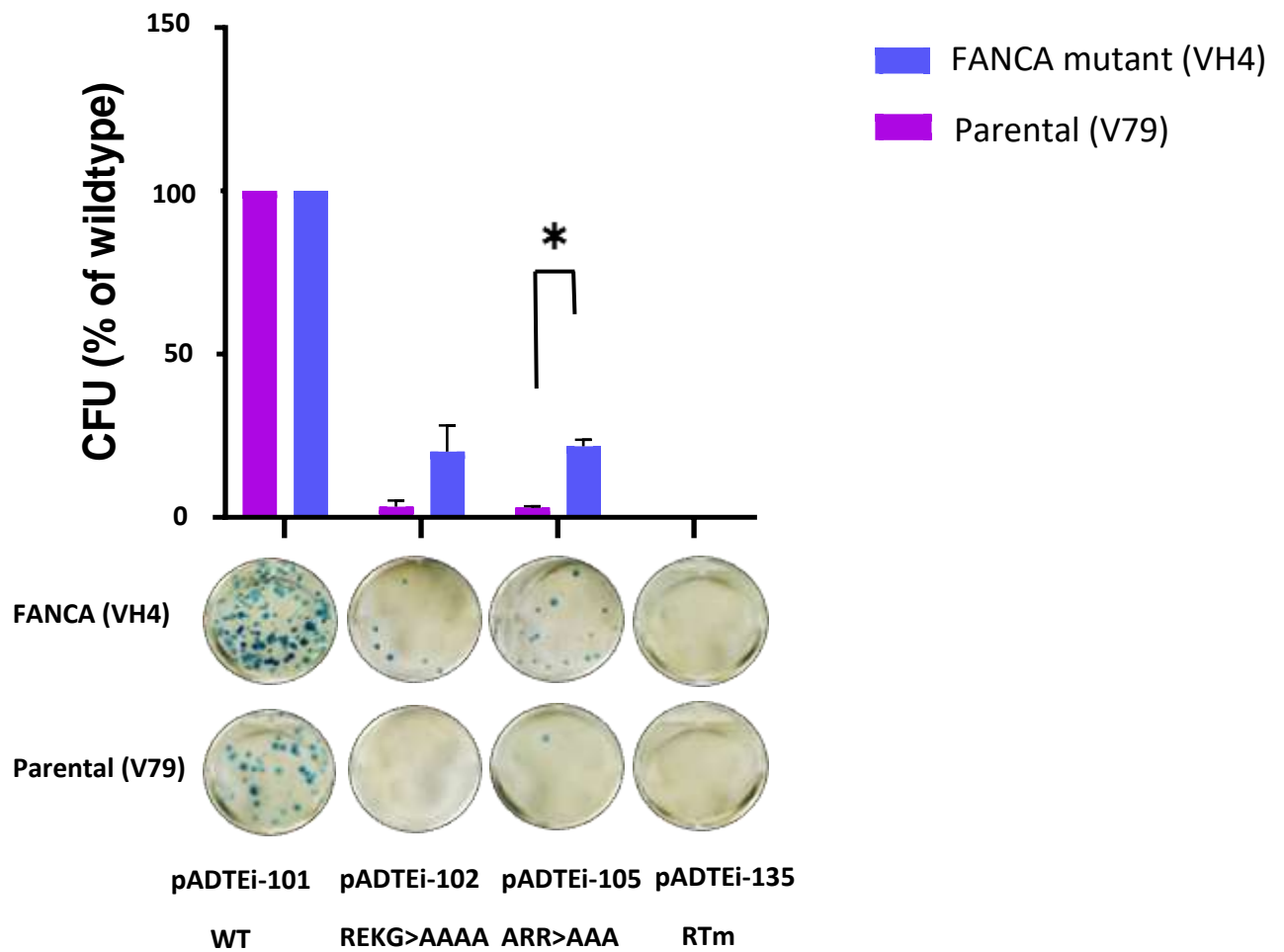
The mutants were generated from three highly conserved amino acid motifs in the carboxyl domain of ORF1 (REKG235-238, RR261-262 and YPAKLS282-287). Because the impacts of these ORF1 mutations on canonical retrotransposition have been well characterised, I wanted to see how they would impact retrotransposition in a FANCA deficient context. A scheme describing the ORF1p mutants I tested is shown in **Figure 4.3**. I tested this series of mutants in the VH4 and V79 cell line **Figure 4.4**.

It is worth noting that in Figure 4.4 the CFU of generated by each L1 plasmid in a particular cell line are normalised against the CFU of the wild type L1 in the same cell line. This is to account for differences which may be accounted for by variation between cell lines rather than between cell lines. Thus, in each cell line the retrotransposition frequency of each mutant L1 is shown as a percentage of its wild type activity. This method will be used in each experiment where the mobilisation of different mutant L1s is compared. However, as this calculation masks differences between the wild type retrotransposition frequency in the different cell lines, Figure 4.3, and other figures focusing on wild type elements, do not employ this normalisation.



**Figure 4.3: Scheme of previously characterised ORF1 mutants**

- A)** A series of ORF1 mutations and their retrotransposition rates as previously characterised in HeLa cells (Kulpa and Moran 2005). Kulpa and Moran (2005) separated these mutants into two categories, Class I and Class II (as depicted above), where Class I mutants abrogate retrotransposition and Class II mutants reduce retrotransposition. Two conserved motifs in ORF1 are depicted, the mutations derived from them and the RT frequency of those mutants. Percentages indicate the frequency of retrotransposition of these mutants relative to the wild type L1 element.
- B)** Diagram depicting the ARR motif
- C)** Diagram depicting the REKG motif which is altered in the pAD2TEI ORF1 mutant



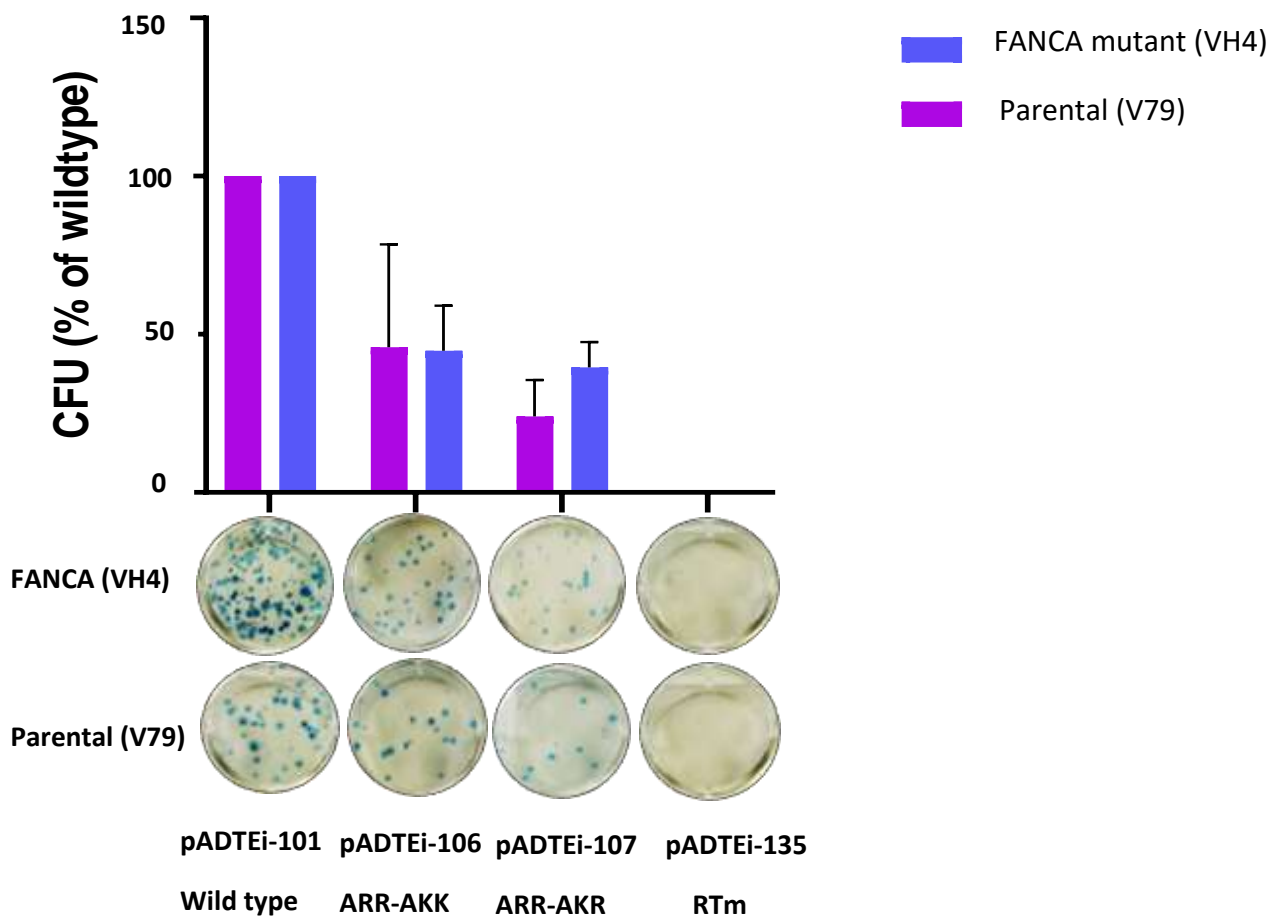
**Figure 4.4: Testing type 1 ORF1 mutations in FANCA deficient and parental cell lines**

The retrotransposition of two severe ORF1 mutants was quantified by a cultured cell retrotransposition assay in a Chinese hamster cell line (V79) and a derivative deficient in FANCA (VH4). The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times. \* indicates  $p < 0.05$ , Student's t-test. Representative plates showing neomycin-resistant colonies stained with crystal violet are shown below the graph.

Interestingly, the class I mutants showed significantly different rates of retrotransposition between the VH4 and V79 cell lines. In the V79 cells, retrotransposition rates were comparable to those found by Kulpa and Moran (2005) in HeLa cells (0-0.5% of wild type levels). In that experiment, ORF1p was undetectable in a RNP isolated by differential centrifugation, implying that the mutation prevented the localisation of the protein to the RNP retrotransposition intermediate. Interestingly, in the absence of FANCA, retrotransposition was significantly higher, around 20% of wild type levels. This

suggests that FANCA can support some retrotransposition of mutant elements which harbour severe ORF1 mutations and likely have diminished capacity to form an RNP.

Of the class II mutants, there was a less distinct difference between the two cell lines. The ORF1 mutant (ARR-AKK) jumped at a similar rate in VH4 and V79 cells. pAD107, mobilised on average at 40% of wild type levels in a FANCA knock out context, which is substantially higher than the 2.6% recorded by Kulpa and Moran (2005) in HeLa cells. However, this was not significantly less than the rates detected in the parental cell line. These findings imply that the loss of FANCA has a more substantial impact on the capacity of severely mutated L1 elements to retrotranspose.



**Figure 4.5: Testing type II ORF1p mutations in FANCA deficient (VH4) and parental (V79) cell lines**

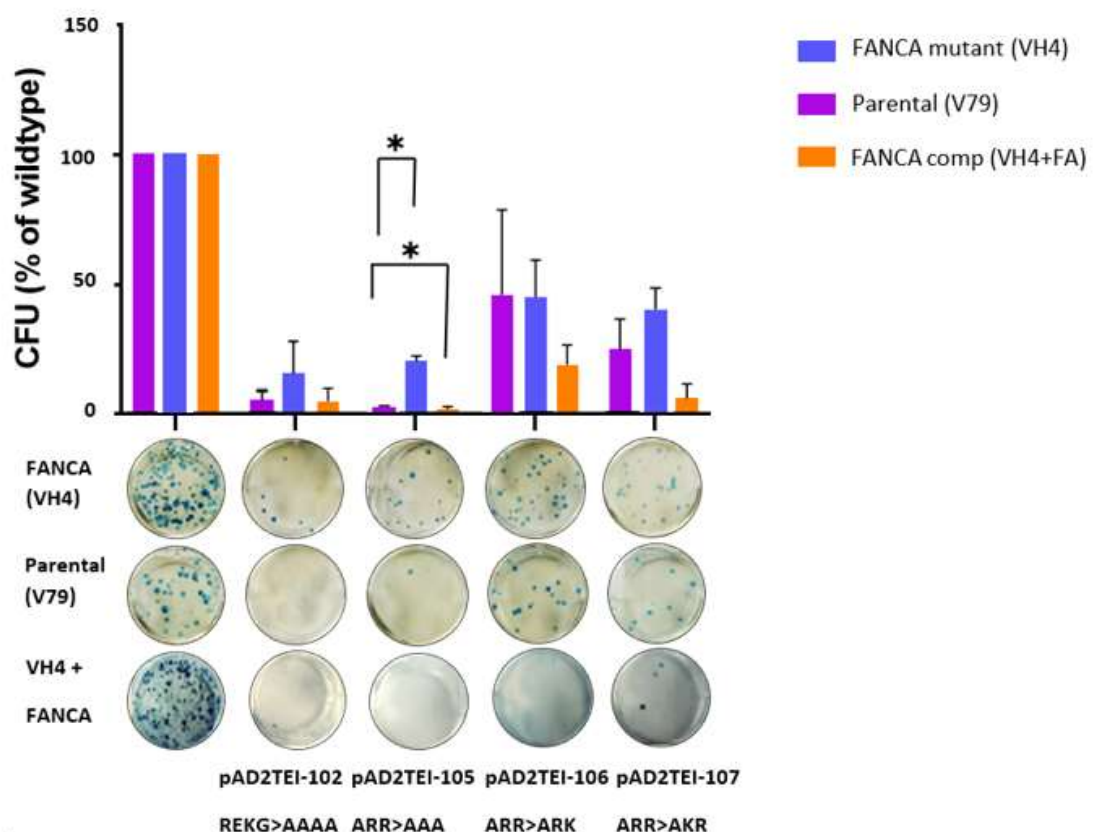
The retrotransposition of two more ORF1 mutants was quantified by a cultured cell retrotransposition assay in a Chinese hamster cell line (V79) and a derivative deficient in FANCA (VH4). The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times. Representative plates showing

neomycin-resistant colonies stained with crystal violet are shown below the graph. Note that wild type panels for VH4 and V79 were previously shown in **Figure 4.4** but is displayed again here for comparison purposes.

#### 4.5. Rescuing unusual retrotransposition in VH4 cells with complementation using FANCA.

As the retrotransposition of severe REKG>AAAA and ARR>AAA ORF1p mutants in VH4 cells was unexpected, we wanted to identify whether it was directly a result of FANCA deficiency and not a secondary effect of DNA damage resulting from the FA phenotype. I therefore repeated the above assay with FANCA complemented cells to see if I could rescue the increased retrotransposition of the mutant elements.

As seen in **Figure 4.6**, complementation by FANCA restored retrotransposition rates to levels more comparable with the v79, parental cell line. This confirms that the increased retrotransposition is likely a direct result of the loss of FANCA.



**Figure 4.6: Rescuing increased retrotransposition of ORF1 L1 elements with FANCA complementation.**

The retrotransposition of a battery of ORF1 mutants was measured in VH4, V79 and VH4 cells complemented with FANCA. The number of colony forming units (CFU) per well was quantified for each element as a percentage

of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times. The VH4 and V79 values were previously plotted in **Figure 4.4** and **Figure 4.5** and are shown here for comparison. \* indicates  $p < 0.05$ , Student's t-test. Representative plates showing neomycin-resistant colonies stained with crystal violet are shown below the graph.

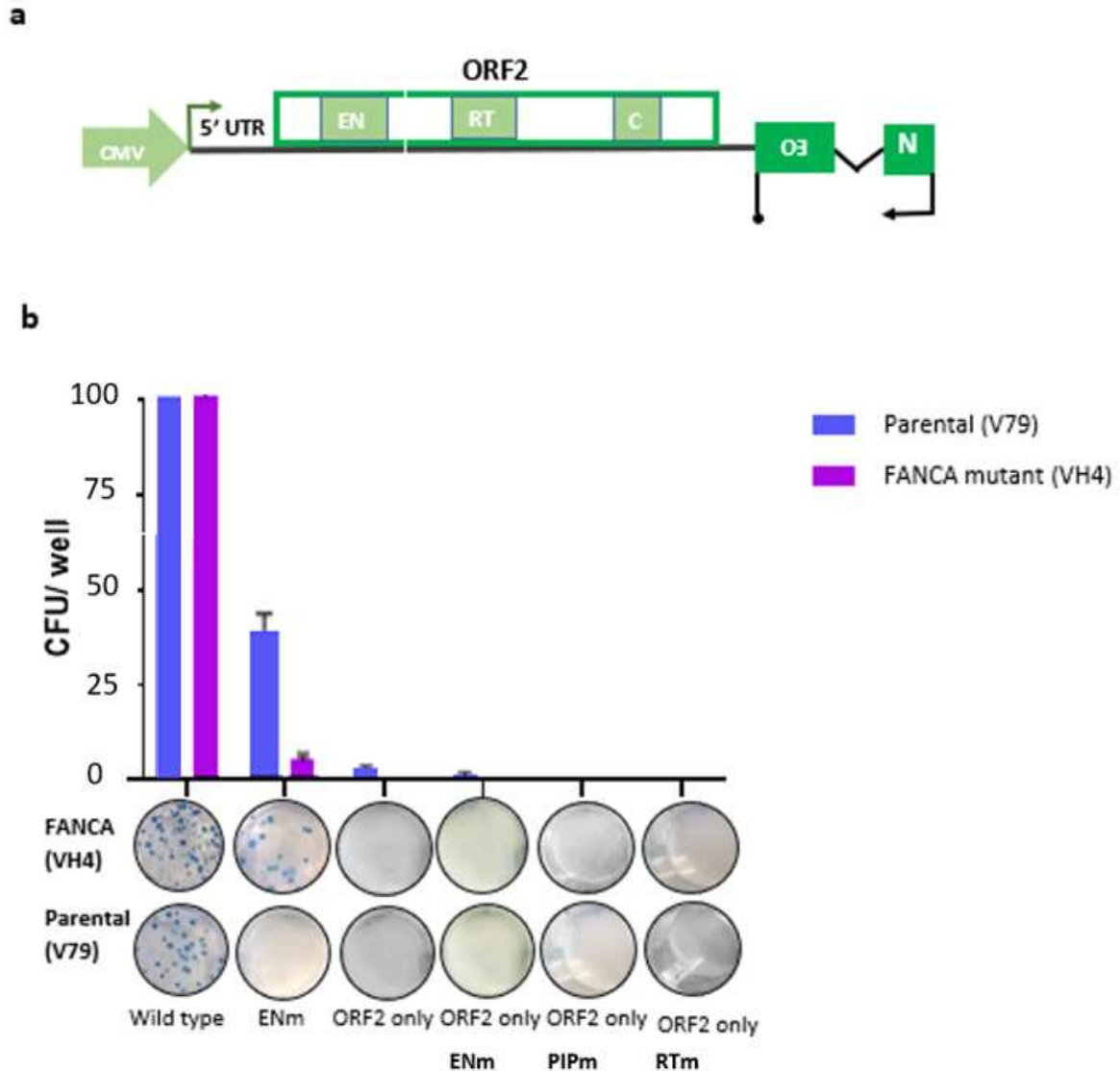
#### 4.6 Testing ORF2 only retrotransposition in FANCA (VH4) and parental (V79) cells.

The mobilisation of the two Class I ORF1 mutant elements in VH4 cells (REKG>AAAA and ARR>AAA) is particularly interesting, because the ORF1p of these elements is reported to have a compromised ability to localise to the RNP (Kulpa and Moran 2005). As the RNP is a vital intermediate of conventional TPRT, this raised the possibility that VH4 cells may support an alternative mechanism of retrotransposition. As previously discussed, the role of ORF1p in retrotransposition has not been fully elucidated. It is not clear, for example, whether ORF1 is essential for retrotransposition outside of its role in the formation of the RNP. Therefore, I decided to test whether VH4 cells could support retrotransposition of a series of ORF1p only plasmids.

In this assay, I tested a series of plasmids derived from the pSA500 cassette (**Figure 4.7**). This series includes four ORF1p only plasmids with several different mutations. As overexpression of ORF2 can lead to high levels of toxicity due to the activity of the endonuclease domain (Gasior *et al.* 2006), I tested two endonuclease mutants to determine whether reducing toxicity of ORF1p would alter the degree of retrotransposition. I also tested a RT mutant as a negative control and a double mutant, which contained alterations in both the EN and the RT domains. As a positive control, the wild type JM101/L1.3 plasmid was used. Finally, an endonuclease mutant derived from JM101/L1.3 was tested as a positive control for ENi retrotransposition in this assay. The results are shown in **Figure 4.7**.

Consistent with previous findings, the wild type element showed good levels of retrotransposition in both cell lines, with a slightly higher rate of retrotransposition in VH4 cells. An endonuclease mutant element also mobilised significantly more in VH4 cells, supporting the theory that FA mutant cells can support endonuclease independent retrotransposition. However, the ORF2 only plasmids, derived from the pSA500 cassette, showed no retrotransposition activity. Notably, the EN domain mutant, which should have reduced toxicity, also did not support any retrotransposition.

These findings indicate that ORF1p is still required for retrotransposition in VH4 cells. This is interesting in light of my previous data, which suggests that severe mutants of ORF can still mobilise in this context, suggesting that ORF1 may have an additional role in TRPT to localisation to the RNP.



**Figure 4.7 Testing ORF2 only retrotransposition in FANCA (VH4) and parental (V79) cells.**

**A)** The scheme of the pSA500 cassette that was used to measure ORF1p only retrotransposition

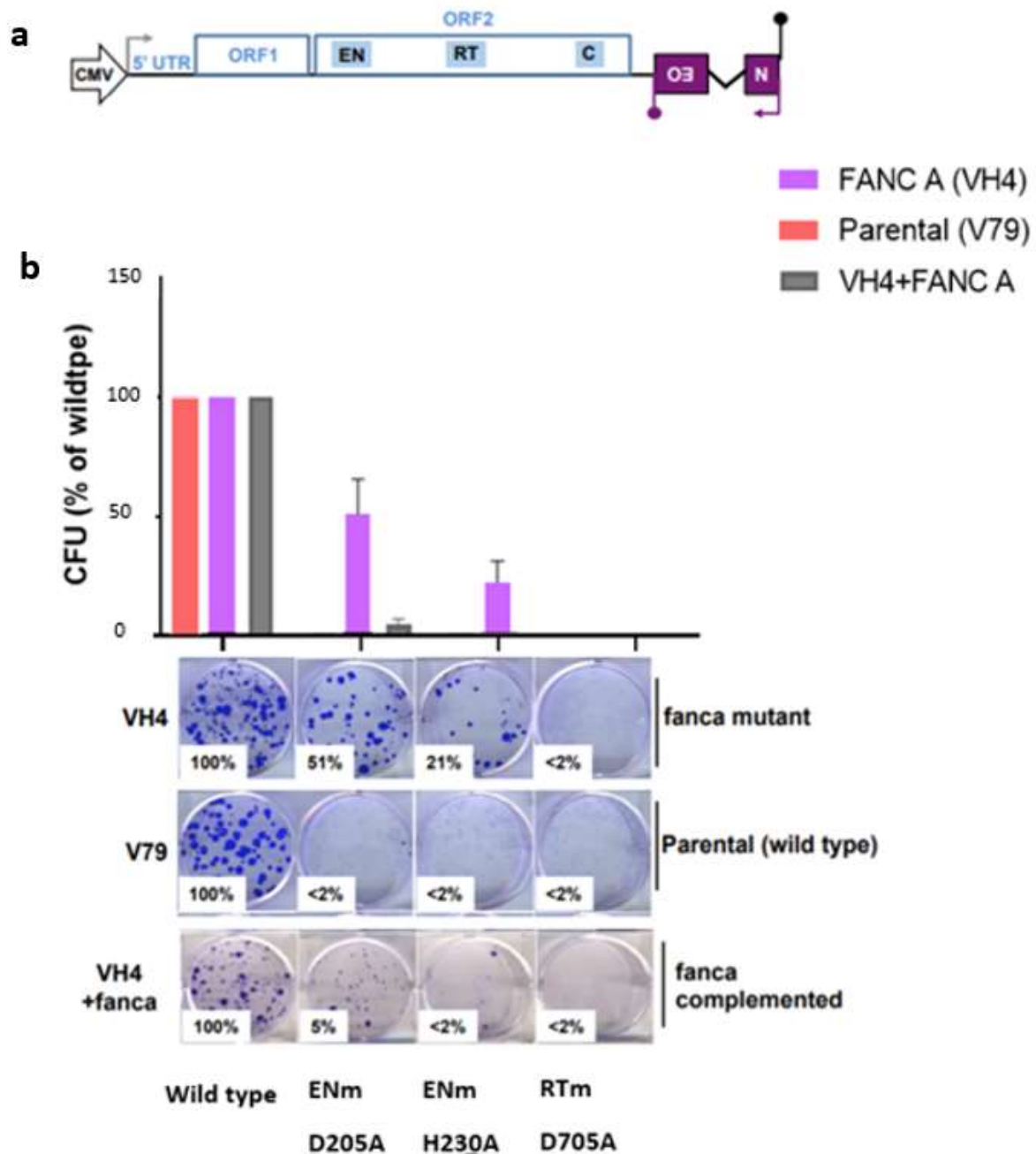
**B)** Retrotransposition of a series of plasmids containing only ORF2 and either an endonuclease mutant (ENm: H230A), a PIP box mutant (PIPm: YY414/15AA) or a reverse transcriptase mutant (RTm :D702A) was measured in FANCA (VH4) and parental (V79) cells. The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times. Representative plates showing neomycin-resistant colonies stained with crystal violet are shown below the graph.



#### 4.7 Testing EN mutants in FANCA deficient cells.

As previously discussed, several studies have found evidence for endonuclease independent retrotransposition including Morrish *et al.* (2002), Morrish *et al.* (2007) and Sen *et al.* (2007). Mutations of the endonuclease domain in wild type cells appear to reduce retrotransposition to approximately 0.8-5% of wild type levels. In cells which are deficient in DNA repair, however, retrotransposition can reach up to 90% of wild type levels. The rate of ENi retrotransposition in a FA context has not yet been identified, however. Therefore, I wanted to investigate EN independent retrotransposition in FANCA mutant cells.

Previous studies investigating endonuclease independent retrotransposition have focused on two elements which harbour missense mutations in the active site of the endonuclease domain: ENm-L1, D205A/L1.3 and ENm-L1, H230A/L1.3. Both mutations have been previously reported to disrupt *in vitro* endonuclease activity, in HeLa cells reducing retrotransposition to 0.5% of wild type levels (Morrish *et al.* 2002). In these assays, I also include elements with the RT mutant D702A, L1 elements which harbour catalytic mutants in the reverse transcriptase domain are retrotransposition defective and primarily serve as a negative control. As shown in **Figure 4.8**, VH4 cells supported increased endonuclease independent retrotransposition of both EN mutants, with higher levels of retrotransposition by the D205A/L1.3 mutant.



**Figure 4.8: Testing Endonuclease mutants in VH4 and parental cells.**

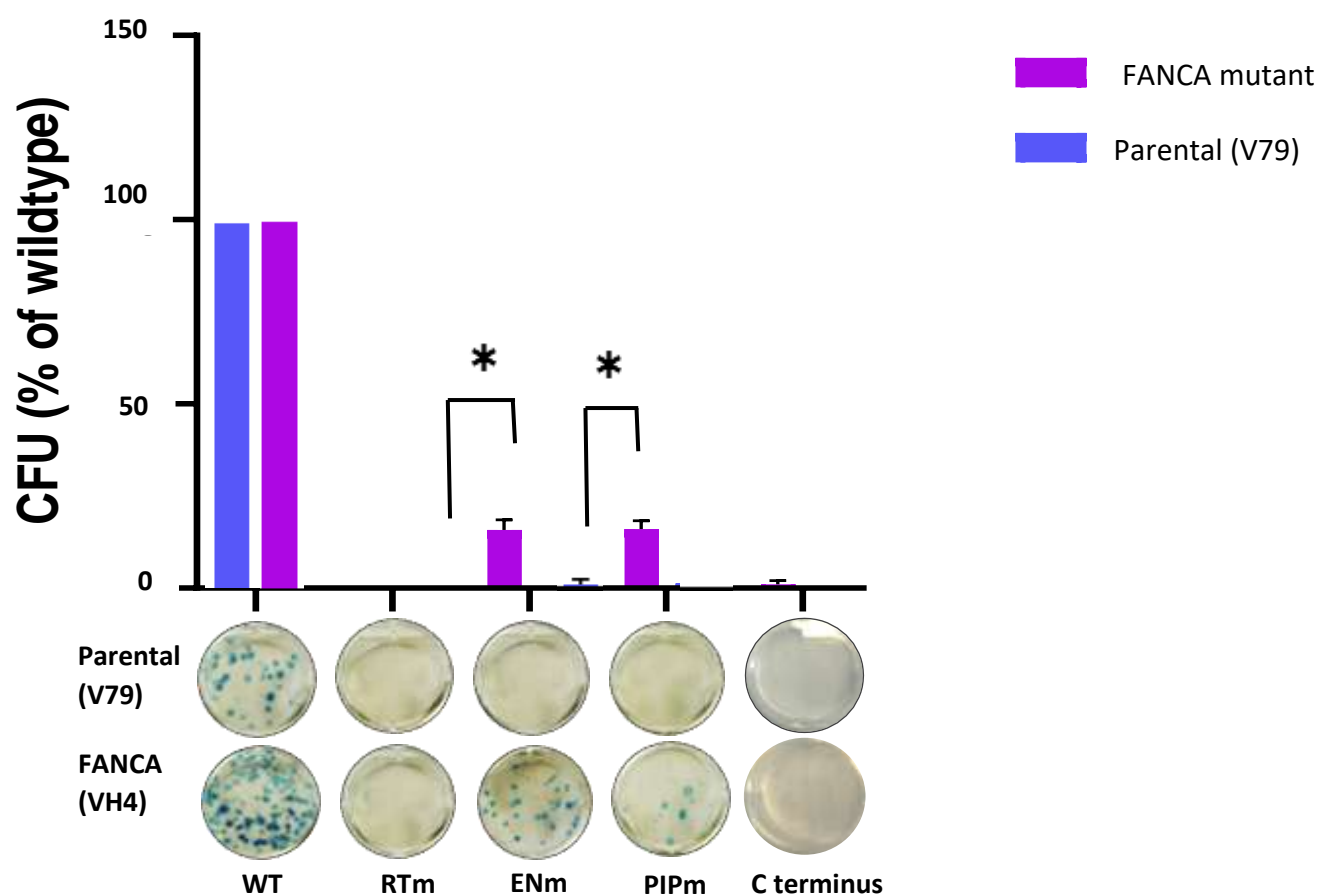
- The scheme of the L1.3 cassette that was used to measure ORF1p only retrotransposition
- Retrotransposition of a series of plasmids containing two different EN mutants was measured in VH4 and V79 cells. The number of colony forming units (CFU) per well was quantified and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times.

#### 4.8 Testing other ORF2 mutants in FANCA deficient cells.

As well as the battery of ORF1p mutants and ORF1p only retrotransposition, there are several domains in ORF2 that were of interest to investigate. ORF2 has several key domains which have been experimentally determined to be required for retrotransposition (Mathias *et al.* 1991), including the EN domain, the PIP domain, the RT domain and the C terminal domain (Moran *et al.* 1996).

In 2013, using an affinity proteomics study, Taylor *et al.* demonstrated that PCNA, the polymerase-delta-associated sliding DNA clamp, interacts with ORF1p via a PIP box motif, during or shortly after TPRT. This binding of PCNA to ORF1p was not affected by the loss of ORF1. Cultured cell assays revealed that the PIP domain was required for retrotransposition of a wild type L1 element in both HeLa and HEK293T cells. Line-1 elements mutated in the PIP domain have not been tested in FANCA deficient cells, therefore I wanted to investigate whether this domain is required in a FANCC context. The PIP mutant used has been previously found to obstruct the interaction between PCNA and L1.

In addition, I wanted to investigate movement of elements with mutations in the C terminus of the ORF1p domain. This domain contains three conserved C residues and a H residue. Previous studies have shown that the C terminus is important for retrotransposition (Moran *et al.* 1996), though the exact function of this domain in TPRT is not yet known. Using plasmids derived from pAD2TEI, I measured the retrotransposition of a PIP mutant, an RT mutant and a C terminal mutant. The results are shown in **Figure 4.9**



**Figure 4.9 Measuring retrotransposition of ORF2 mutants in FANCA (VH4) and parental (V79) cells**

The retrotransposition of several ORF2 mutants: RTm (D702A), ENm (H230A), PIPm (YY414/15AA) and C terminus (C1143S, C1147S) was quantified by a cultured cell retrotransposition assay in a Chinese hamster cell line (V79) and a derivative deficient in FANCA (VH4). The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times. \* indicates  $p < 0.05$ , Student's t-test. Representative plates showing neomycin-resistant colonies stained with crystal violet are shown below the graph. Note that wild type panels were previously shown in **Figure 4.4** and **4.5** but are shown here again for comparison.

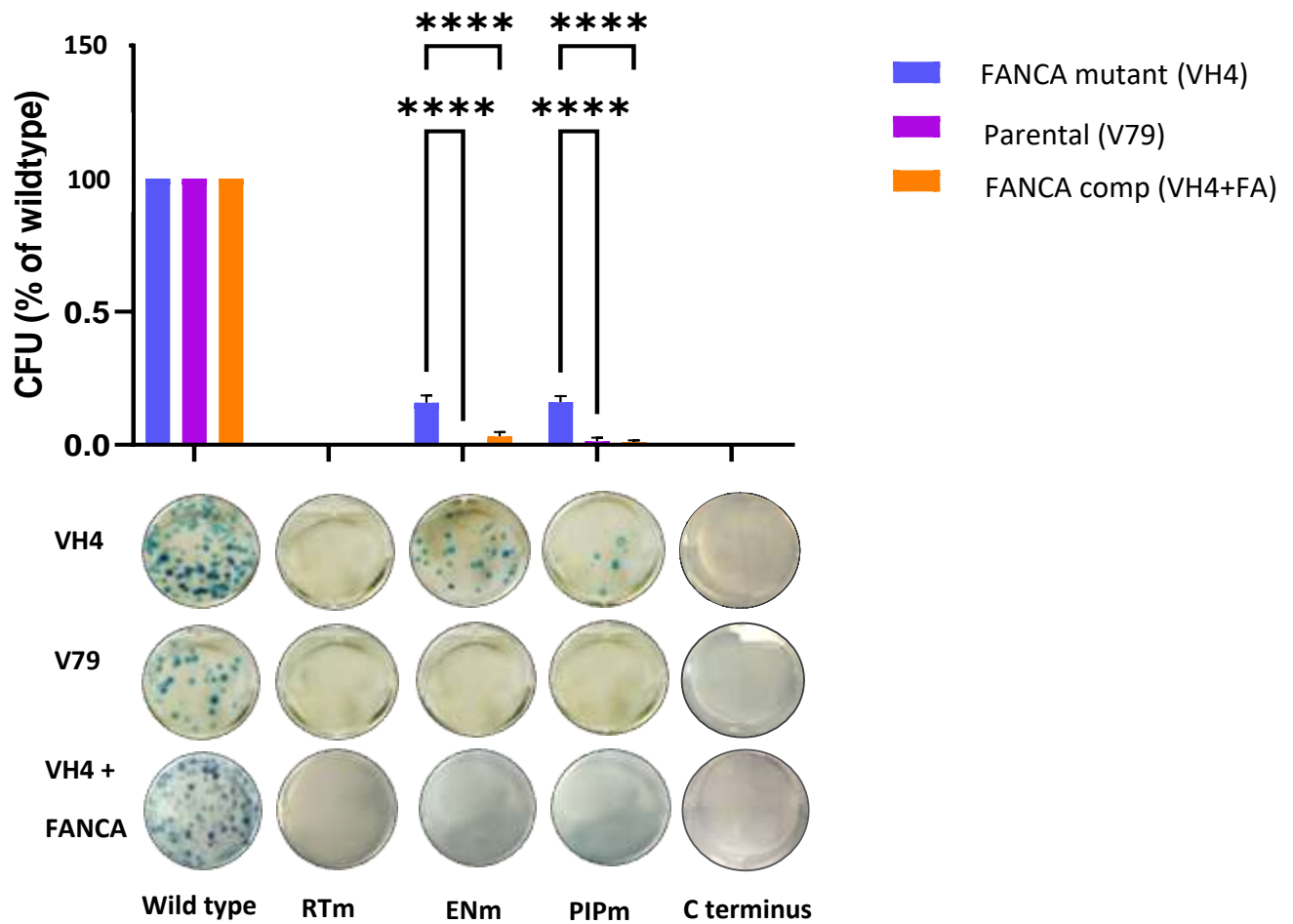
As expected, no retrotransposition was observed in the reverse transcriptase mutant in either cell line. Interestingly though, significantly higher levels of retrotransposition were observed in the VH4 cells

of the PIP mutant. This suggests that a proportion of retrotransposition events are happening in the absence of FANCA which do not require a functional PIP domain, suggesting that the involvement of PCNA may be altered during retrotransposition in a FA context.

The C terminal mutant, which harboured alterations in the conserved residues C1143S, C1147S, did not retrotranspose either in VH4 or V79 cells. Providing further evidence that these residues are important even in cell lines deficient in DNA repair proteins.

As confirmed in previous experiments, VH4 cells also supported increased retrotransposition of the H230A endonuclease mutant.

Together, these experiments suggest that some domains of ORF1p, including the PIP and EN domain, are not required for retrotransposition in FANCA deficient cells. To substantiate these findings, I repeated the assay in the VH4 cells complemented with FANCA. The results are shown in **Figure 4.8** Complementation with FANCA successfully rescued the abnormal retrotransposition observed in VH4 cells, indicating that this effect is a direct result of loss of FANCA, and not an indirect result of secondary DNA damage resulting from the FA phenotype.



**Figure 4.10 Increased retrotransposition of ORF2 mutants is rescued by complementation with FANCA.**

Retrotransposition of several ORF2 mutants was measured in three cell lines, FANCA mutant (VH4), parental (V79) and FANCA complemented (VH4 + FANCA). The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times. FANCA mutant and Parental data were previously shown in **Figure 4.9** but are displayed again here for comparison purposes. \*\*\*\* indicates  $p < 0.001$ , Student's t-test

## 4.9 Main conclusions

To conclude, the main findings from this chapter are as follows. Firstly, I have demonstrated that FANCA can serve as a negative regulator of wild type L1 retrotransposition. This effect can be rescued using complementation with FANCA.

Secondly, FANCA mutant cells can support retrotransposition of several ORF1 mutants. In particular VH4 cells support significantly higher levels of a severe ORF1 mutant than a parental cell line. This finding is of particular interest, as the ORF1 mutant has been shown to abrogate formation of the RNP, which is a required retrotransposition intermediate for TPRT. Complementation with FANCA rescues this unusual retrotransposition.

Thirdly, my results show that ORF1 is still required as a whole for retrotransposition in VH4 cells, as demonstrated with a retrotransposition assay using variations of an ORF2 only plasmid.

Finally, I have shown that VH4 cells support retrotransposition of several ORF2 mutants. This includes two endonuclease mutants and a PIP mutant. This effect is rescued by complementation with FANCA.

## 4.10 Discussion

The first notable finding from the work in this chapter is the increase in wild type retrotransposition following deletion of FANCA. This finding is supported by previous reports which have proposed FANCA as a potential regulator of L1 retrotransposition (Mita *et al.* 2018; Liu *et al.* 2018; Tristan Ramos *et al.* 2020). The mechanism by which FANCA could modulate retrotransposition is not yet known. The protein forms part of the core Fanconi complex, along with seven other FA proteins. This complex primarily functions as an E3 ligase, catalysing the ubiquitination of FANCD2. It is not known whether FANCA uniquely serves as a regulator of L1, or if this capacity is shared by other FA proteins, potentially those in the core complex. Indeed, genome wide screens have identified multiple FA proteins as interactors (Mita *et al.* 2018; Liu *et al.* 2018). In addition, Cherubini *et al.* (2011) found that the FA pathway is activated by adenoviral infection, independently of the canonical DNA repair response. It is therefore possible that the FA pathway could be triggered in a similar manner by retrotransposition intermediates.

Several proteins in the FA pathway have demonstrated individual additive functions outside of their contribution to the pathway, and the regulatory effects of FANCA could therefore be additional to its function in the core complex. Notably, Yuan *et al.* (2012) have shown that purified FANCA has an

intrinsic affinity for nucleic acids. The protein was observed to bind to both ssDNA and dsDNA, with a significantly higher affinity for single stranded transcripts. It is therefore conceivable that FANCA could modulate L1 activity through some form of direct interaction, though the details of this remain to be elucidated.

Interesting questions moving forward from this finding would therefore be 1) is the regulation of L1 by FANCA shared by other FA proteins? 2) Does FANCA bind the L1 transcript?

A second discussion point from this body of work is the abnormal retrotransposition reported in VH4 cells. Overall, this cell line supported increased retrotransposition from a range of mutants, including a severe ORF1 mutant which has been shown to prevent localisation of ORF1p to the RNP, usually an essential retrotransposition intermediate (Kulpa and Moran 2005). Notably, deletion of ORF1 entirely abrogated retrotransposition, both in VH4 cells and the parental line, indicating that there is more complexity to the function of ORF1 in retrotransposition than facilitating formation of the RNP. Though unlikely, as the modified amino acid motifs are highly conserved, there is a slim possibility that these mutants do not behave as previously reported in VH4 and V79 cells as in other cell lines, and a future experiment to confirm this would further consolidate this conclusion. Nonetheless, the retrotransposition of the ORF1p mutant suggests that an alternative mechanism of mobility to TPRT could be taking place in these cells.

In addition, the VH4 cells also supported retrotransposition of a PIP mutant, raising the possibility that the role of PCNA may be altered in these retrotransposition events. The rescue of these abnormalities with FANCA complementation suggests that these findings are a genuine result of deletion of the protein. This rescue is a critical control, as FA cells harbour large amounts of secondary DNA damage due to the loss of the pathway. Thus, it is very important to demonstrate that the loss of the protein is caused the observed effect, and not unrelated DNA damage.

It is also worth noting that the FA mutant cells support ENi retrotransposition, similar to that observed by Morrish *et al.* (2002) and (2007). This indicates that L1 may be able to exploit other types of DNA lesion than the double stranded breaks found in NHEJ deficient cells, and shows that ENi retrotransposition might be possible in a variety of cell lines.

Investigation into the characteristics of these retrotransposition events is of interest, as these domains are usually vital for efficient mobilisation via TPRT. Structural hallmarks in these insertions could provide more insight into the mechanism of L1 mobilisation taking place, and whether this is common between mutant and wild type L1 elements against a FA background.



Previously it was thought that V79 Chinese hamster cells were devoid of p53 activity and that p53 inactivation was required for efficient ENi, however these studies were mostly performed in HCT116 colorectal cancer cell lines and NHEJ-deficient XR-1 hamster cell lines and extrapolated to V79 cells (Moro *et al.* 1995; Chuang *et al.* 1997). Experiments in Jose Garcia Perez's lab based on my results with V79 and VH4 cells suggest that V79 and VH4 cells are proficient for p53 activation based on phosphorylation of p53 serine-15 in response to etoposide, and that co-expression of a dominant-negative form of p53 (s.Morell unpublished) in the L1 retrotransposition assay in V79 and VH4 cells did not significantly alter L1 mobility. Thus, it seems likely that ENi retrotransposition in V79 and VH4 cells is independent of p53 activity



## Chapter 5

---

# **Investigating the activity of mutant L1 in different DNA damage response contexts**



## 5.1 Introduction

Fanconi Anaemia (FA) is caused by the biallelic inactivation of one of 22 genes that form the FA/BRCA pathway of DNA repair (Moreno *et al.* 2021). In response to DNA lesions during replication, most FA proteins associate to form the “core complex”, which is required to signal/activate DNA repair (reviewed in Ceccaldi *et al.* 2016). In the previous chapter, I had identified that FANCA mutant cells support unusual pathways of L1 retrotransposition, suggesting that FANCA, part of the “core complex”, could be involved in the regulation of L1 activity. Notably, FANCA is the most frequently mutated gene in patients (>65% of all characterised; Levrán *et al.* 2005). These data further suggest that L1 mobilization might be inherently deregulated in FA patients. However, from these experiments it is not clear whether FANCA might be unique in this respect. Thus, I next explored whether a similar phenotype might be associated with additional members of the FA “core complex”, with members of FA subcomplexes, or with alternative DNA repair pathways.

Remarkably, several FA proteins are known to have additional roles outside the FA/BRCA pathway of DNA repair, which is critical for the repairing of interstrand DNA crosslinks (reviewed in Ceccaldi *et al.* 2016). Therefore, it's possible that any L1 regulatory function of FANCA could be additional to its role in the core FA complex. However, recent genome wide screens have also identified other FA proteins as potential L1 interactors/regulators, suggesting that other proteins in the pathway could share this regulatory capacity (Liu *et al.* 2018; Mita *et al.* 2021). The FA pathway and the overlap of functionality between these different proteins is still being elucidated.

Additionally, I also wanted to determine whether the alternative Endonuclease independent (ENi) pathway of L1 integration activated in FANCA cells is mechanistically related to ENi retrotransposition reported in Non-Homologous End Joining (NHEJ) mutant cells (Morrish *et al.* (2002); Morrish *et al.* (2007)). Furthermore, it is currently unknown if additional pathways of alternative L1 retrotransposition activated in FA cells (i.e., ORF1p- and PIP-independent retrotransposition) could also occur in NHEJ mutant cells.

## 5.2 Aims

The experiments included in this chapter was designed to answer the following research questions:

1. Do NHEJ mutant cells support the unusual pathways of L1 retrotransposition characterised in FANCA mutant cells (ORF1p-, EN-, and PIP-independent integration)?
2. Can other FA mutant cells support the unusual pathways of L1 retrotransposition characterised in FANCA mutant cells (ORF1p-, EN-, and PIP-independent integration)?

## 5.3 Several FA proteins have been identified as potential L1 regulators.

I was interested to determine whether deletion of other FA proteins would result in a similar retrotransposition phenotype, similar to those found in the previous chapter using FANCA mutant cells. As previously discussed, several large-scale studies seeking to identify L1 regulators have identified several FA proteins (Liu *et al.* 2018; Mita *et al.* 2020). It is therefore plausible that other proteins within the FA/BRCA pathway of DNA repair could also regulate retrotransposition, and that their deletion could activate the mobilisation of LINE-1 elements using non-canonical pathways (i.e., ORF1p-, EN- and PIP-independent retrotransposition). The FA proteins which have been identified in previous studies as regulators of L1 retrotransposition are shown in Table 5.1. Intriguingly, while most FA proteins were shown to negatively impact L1 retrotransposition (i.e., repressors), other FA members, like FANCI and C, were shown to positively regulate retrotransposition (i.e., activators, Table 5.1).

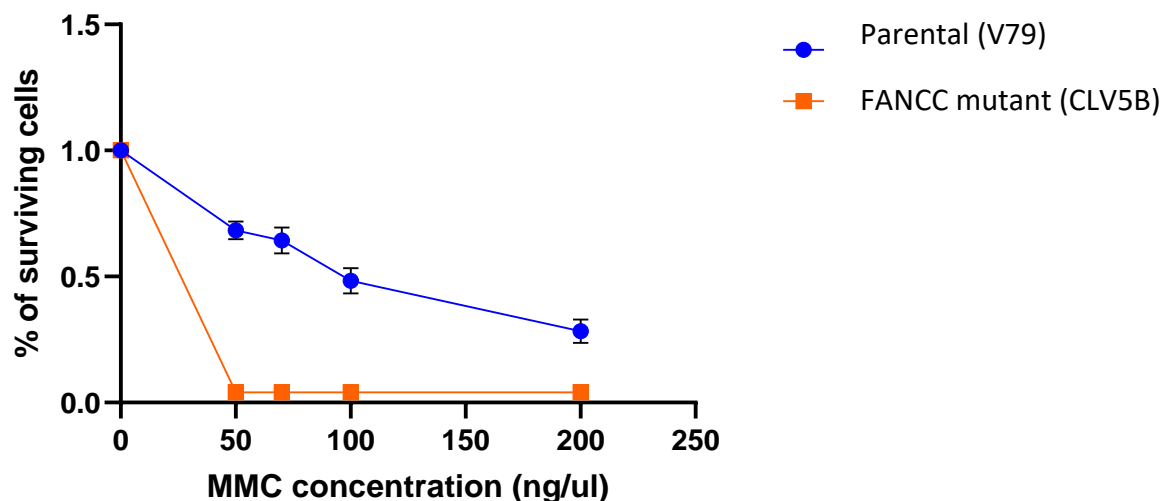
**Table 5.1 FA proteins previously identified as L1 regulators.**

Protein	Study	Positive/Negative regulator
FANCI	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	Positive
FANCL	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	Negative
FANCB	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	Negative
FANCD2	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	Negative
BRCA1	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	Negative
FANCC	<i>Liu et al. 2018</i>	Positive
FANCG	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	Negative
FANCE	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	Negative
FANCF	<i>Liu et al. 2018</i>	Negative
FANCM	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	
BRCA2	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	Negative
UBE2T	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	Positive
FANCA	<i>Liu et al. 2018</i> <i>Mita et al. 2020</i>	Negative
FANCI	<i>Mita et al. 2020</i>	
PALB2	<i>Mita et al. 2020</i>	Negative

One candidate FANCC member which was identified by Mita *et al.* (2020) and validated as a negative regulator of retrotransposition, was FANCC. Like FANCA, FANCC assembles into the core complex

which facilitates the monoubiquitination of FANCD2. Therefore, I decided to investigate whether L1 retrotransposition is deregulated in FANCC mutant cells. To do that, I used hamster lung FANCC KO (CLV5B) and parental (V79) cells, and the battery of WT and mutant L1 elements described previously.

Prior to this series of experiments, I performed a Mitomycin C (MMC) sensitivity assay on CLV5B cells alongside their parental line. The results are shown in **Figure 5.1**, demonstrating that, as expected and similar to VH4 cells, CLV5B cells exhibit hypersensitivity to MMC (i.e., a classical FA phenotype).



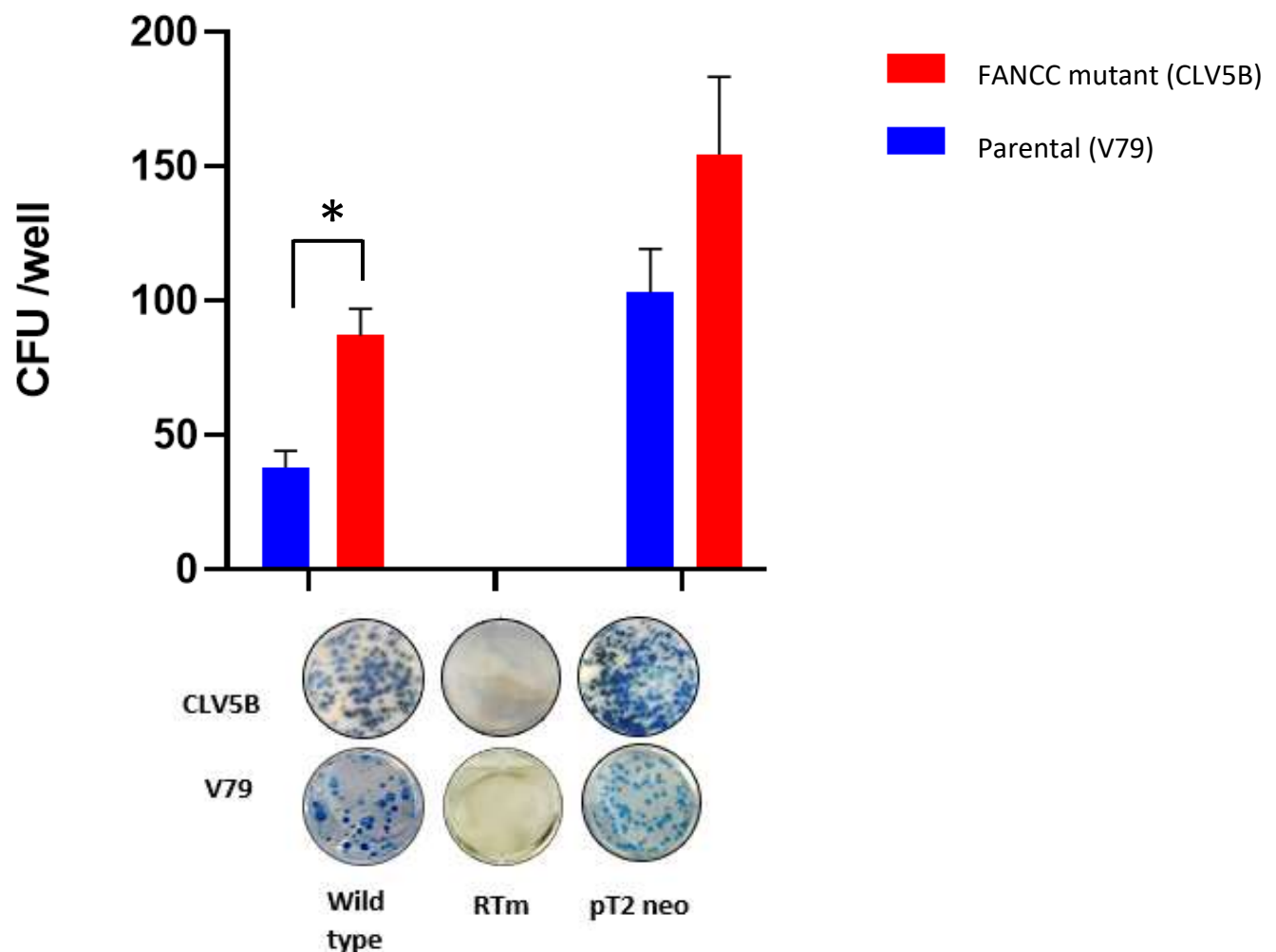
**Figure 5.1 Mitomycin C (MMC) sensitivity assay in FANCC (CLV5B) and parental (V79) cells.**

CLV5B and V79 cells were tested for their relative sensitivity to the crosslinking agent MMC. After a 24-hour treatment using different concentrations of MMC, surviving cells were fixed, stained and quantified using a spectrophotometer. Error bars indicate SEM and this experiment was performed three times.

Once confirmed as FA cells, I first tested whether FANCC also acts as a negative regulator of WT L1 retrotransposition. To do this, I measured WT L1 retrotransposition in CLV5B cells in comparison to parental V79 cells, using plasmids tagged with *mneol* (i.e., activate resistance to G418). I tested the retrotransposition of three plasmids: pAD2TE1-101/L1.3, a WT active human L1 element (Sassaman *et al.* 1997), pAD2TE1-135/L1.3, an RT-mutant allelic vector (acting as a negative control), and pT2 Neo, a plasmid that constitutively expresses the *NEO* reporter gene, serving as a positive control (i.e.,



allows to control for differences in the formation of antibiotic-resistant foci). The results of these experiments are shown in **Figure 5.2**. Notably, I observed that the WT L1 element retrotransposed significantly more in CLV5B cells, indicating that, similarly to FANCA, FANCC has the capacity to restrict WT L1 retrotransposition. Indeed, these results align with findings by Tristan Ramos (2020) suggesting that depletion of FANCC with siRNAs increases wild type L1 retrotransposition. This experiment should also help to deduce whether FANCA regulation of L1 is related to the role of FANCA in the FA pathway



**Figure 5.2 FANCC regulates WT L1 retrotransposition**

Human L1 retrotransposition in FANCC mutant (CLV5B) and parental cells (V79). In the assay, I used engineered L1 vectors containing epitope tags in L1-ORF1p and L1-ORF1p (T7 and TAP, respectively; both epitopes were cloned in the C-terminus) and that were tagged with *mneol*; pAD2TE1-101, WT L1; pAD2TE1-135, RT mutant. To control for the formation of G418-resistant foci in FANCC and parental cells, I used pT2Neo, a plasmid which express the NEO reporter gene. The number of colony forming units (CFU) per well was quantified for each element and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times (shown are average +/-

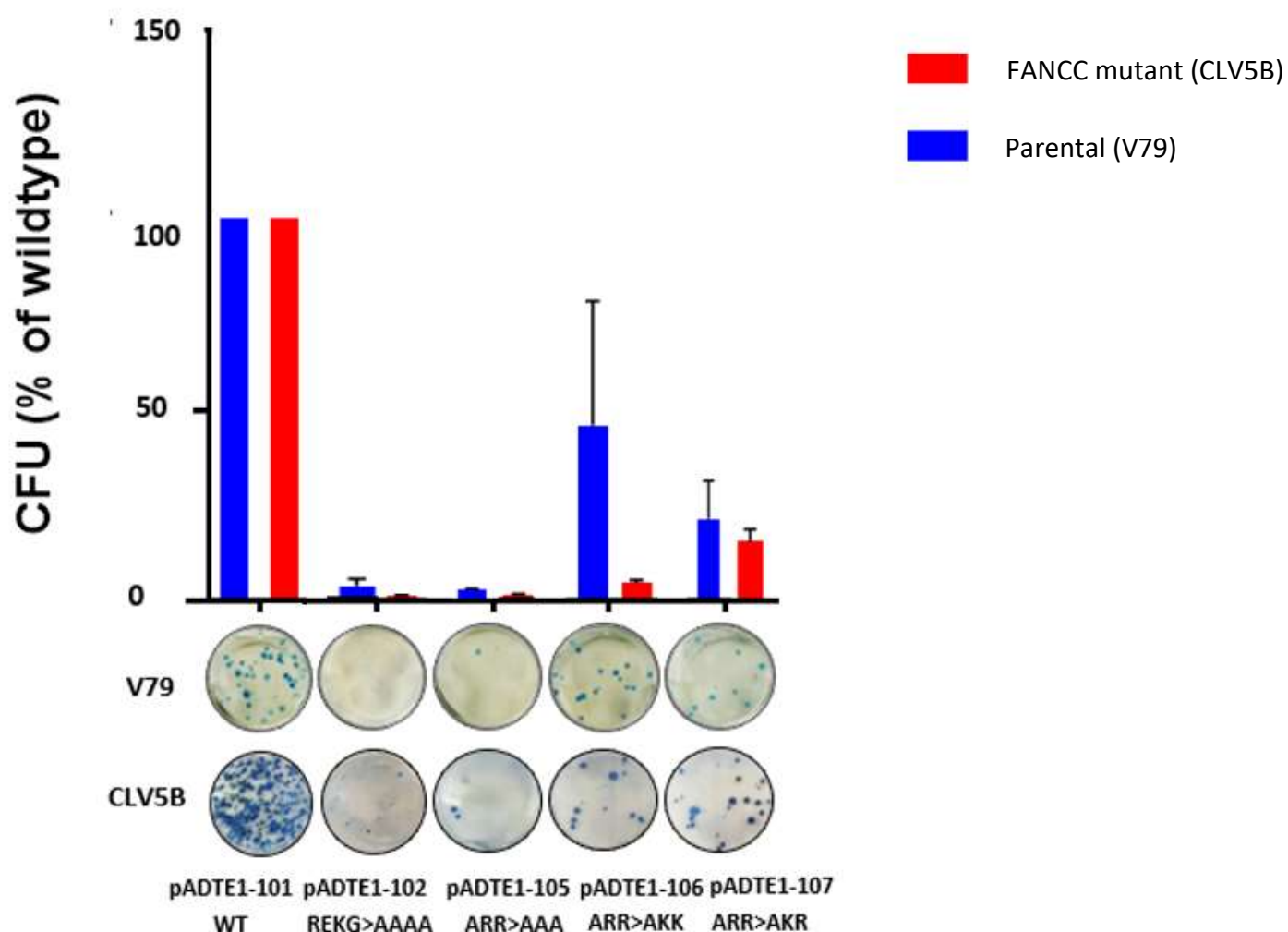
SD). \* indicates  $p < 0.05$ , Student's t-test. The V79 data has been previously show but is displayed again here for comparison purposes.

#### 5.4 Testing the retrotransposition of L1-ORF1p mutant elements in FANCC deficient cells.

A prominent finding in Chapter 4 was the retrotransposition of severe L1-ORF1p mutants in FANCA deficient cells. As FANCC deficient cells showed a similar increase in WT L1 retrotransposition, I wanted to test whether it would also support the unexpected retrotransposition of L1-ORF1p mutants as might be expected if this form of retrotransposition is a general features of FA proteins. Therefore, as part of the core FA complex, FANCC was a good candidate for these experiments.

Thus, I transfected the same battery of L1-ORF1p mutants (tested in FANCA deficient cells, see **Figure 5.2**) in CLV5B and parental V79 cells. As above, the L1-ORF1p mutant vectors, and the WT-L1, were cloned in the pAD2TE1 backbone (contains T7 and TAP epitope tags in L1-ORF1p and L1-ORF1p, respectively, and that were tagged with *mneol*); the series of plasmid tested include pAD2TE1-101/L1.3, the WT L1 element, pAD2TE1-102/L1.3 and pAD2TE1-105/L1.3, two severe **class I** L1-ORF1p mutants, as well as pAD2TE1-106/L1.3 and pAD2TE1-107/L1.3, which are less severe **class II** L1-ORF1p mutant elements. As internal controls, cells were also transfected with pAD2TE1-135/L1.3 (RT-mutant L1 element) and pT2neo (constitutive NEO expression).

In contrast to what I observed in FANCA mutant hamster cells (VH4 cells), there was not a significant difference between the retrotransposition of L1-ORF1p mutant elements between CLV5B (FANCC mutant) and parental cells (**Figure 5.3**). Indeed, I found that retrotransposition of two class I L1-ORF1p mutants (pAD2TE1-102/L1.3 and pAD2TE1-105/L1.3) was negligible in both CLV5B and V79 cells (approximately 1% and 3% of L1-WT, respectively **Figure 5.3**). More similar to what I reported in FANCA mutant cells, I found that the two class II L1-ORF1p mutants (pAD2TE1-106/L1.3 and pAD2TE1-107/L1.3, **Figure 5.3**) retrotransposed at similarly low levels in FANCC deficient and parental cells (approximately -50% of L1-WT, **Figure 5.3**). Altogether, these data suggest that the ability of FANCA mutant cells to support retrotransposition of severe L1-ORF1p mutant elements (i.e., L1-ORF1p-retrotransposition) is not a general feature of FA proteins.



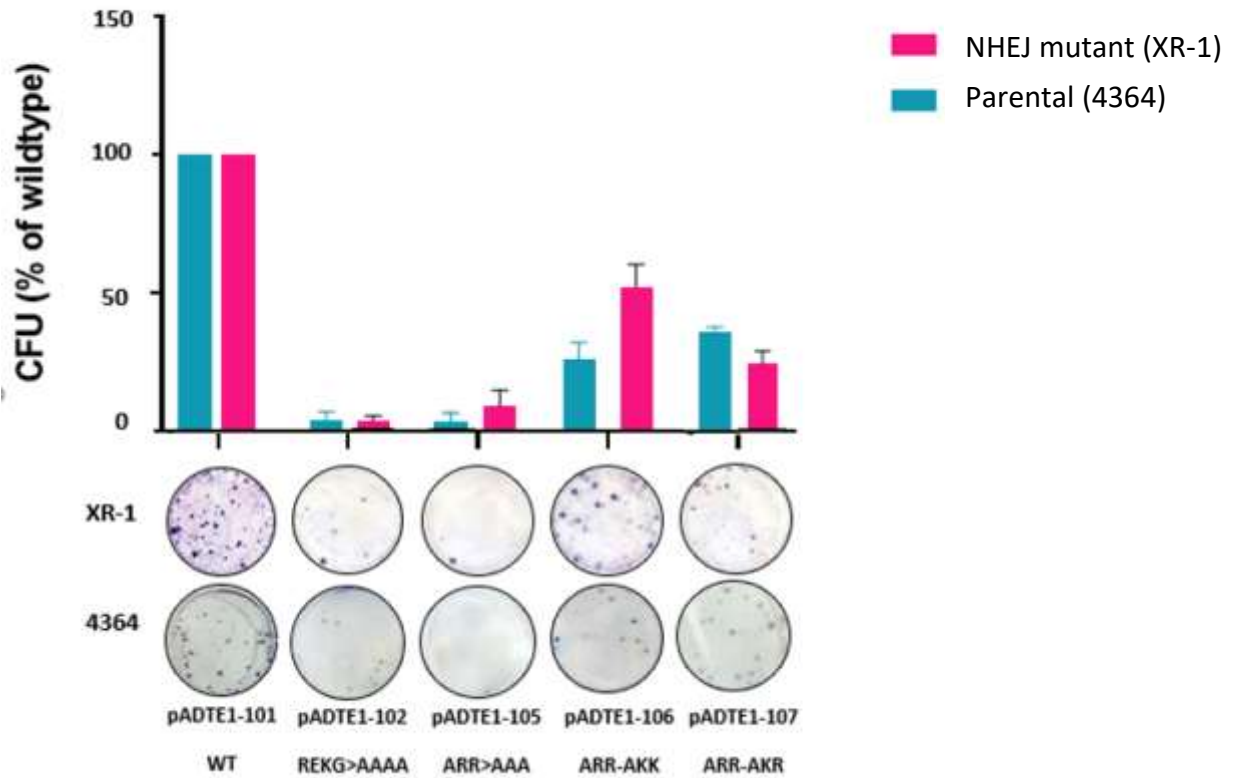
**Figure 5.3 Testing a series of L1-ORF1p mutants in FANCC deficient cells.**

The retrotransposition of a WT-L1 (pAD2TE1-101) and of several class I (pAD2TE1-102 & 105) and class II (pAD2TE1-106 & 107) L1-ORF1p mutants was quantified using the cell culture retrotransposition assay, using parental hamster lung cells (V79) and a derivative deficient in FANCC (CLV5B). The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times (shown are mean  $\pm$  SD).

## 5.5 Testing the retrotransposition of L1-ORF1p mutant elements in NHEJ deficient cells.

As discussed, a prominent finding in Chapter 4 was the discovery of L1-ORF1p-retrotransposition in FANCA deficient cells (that is that severe L1-ORF1p mutants could retrotranspose in FANCA deficient cells). One of the key questions in this chapter was to investigate whether this phenotype was unique to the FA pathway or could be observed when other DNA repair pathways are inactivated. I therefore decided to examine whether L1-ORF1p-retrotransposition could be detected in cells deficient in the Non-Homologous End Joining (NHEJ) pathway of DNA repair. To conduct this key experiment, I used hamster XR-1 cells (Stamato *et al.* 1983), Chinese hamster cells deficient in XRCC4 and lacking DNA ligase IV activity, two proteins which are necessary for NHEJ (Stamato *et al.* 1983). As a control, I used the parental cell line of XR-1 cells, 4364 cells (Stamato *et al.* 1983). Notably, previous studies using XR-1 and additional NHEJ deficient cell lines (i.e., DNAPKcs deficient cells) revealed that that NHEJ deficient cells could support elevated levels of ENi L1-retrotransposition (Morrish *et al.* 2002; Morrish *et al.* 2007), so whether they could also support retrotransposition of severe L1-ORF1p mutants was a matter of interest.

As above, I transfected the same series of engineered L1 vectors (pAD2TE1 series) in 4364 and XR-1 cells, and I selected cells harbouring de novo insertions using G418 selection. Interestingly, as observed with FANCC mutant cells, there were no significant differences between the retrotransposition of L1-ORF1p mutant elements in NHEJ-mutant cells and the parental cell line (**Figure 5.4**). Controls revealed that EN-mutant L1s could retrotranspose in XR-1 cells but not in parental cells (data not shown but see below). In XR-1 cells, L1-ORF1p mutants retrotransposed at 0.6-2% of WT-L1 levels, similar to the rates reported in HeLa cells by Kulpa and Moran (2005). These data suggests that loss of activity of Ligase IV and XRCC4 does not have the same impact on retrotransposition as deletion of FANCA. Thus, the capacity to support L1-ORF1p-retrotransposition does not appear to be generalisable to cells with defects in DNA repair in general.



**Figure 5.4 Testing a series of L1-ORF1p mutants in NHEJ deficient cells.**

The retrotransposition of a WT-L1 (pAD2TE1-101) and of several class I (pAD2TE1-102 & 105) and class II (pAD2TE1-106 & 107) L1-ORF1p mutants was quantified using the cell culture retrotransposition assay, using parental hamster ovary cells (4364) and a derivative deficient in XRCC4/DNAIlgase IV (XR-1). The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times (shown are mean +/- SD).

## 5.6 Testing the retrotransposition of L1-ORF1p mutant elements in human FA deficient cells.

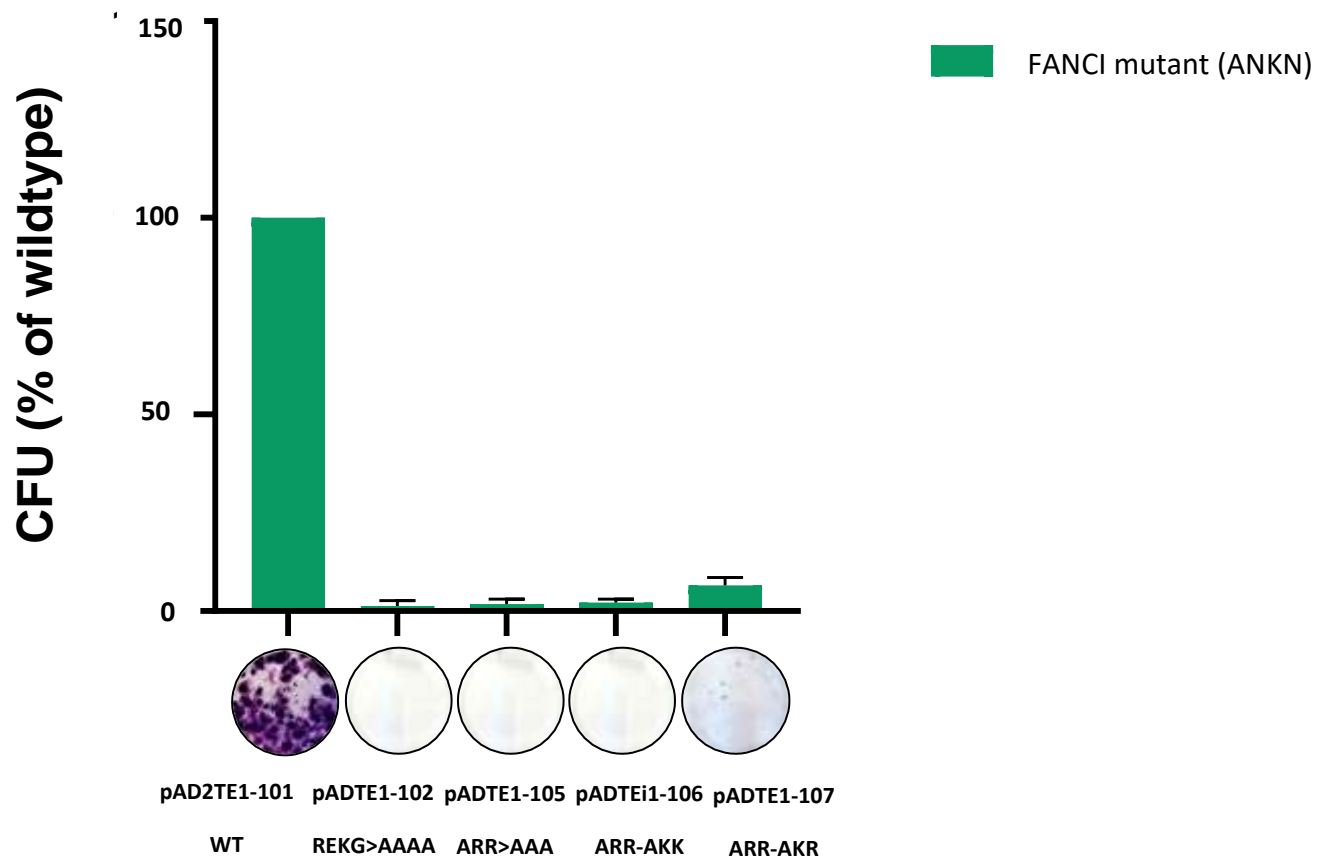
Having tested the mobilisation of L1-ORF1p mutants in hamster NHEJ and FA deficient cells, and to confirm that L1-ORF1p-retrotransposition is unique to FANCA deficient cells, I decided to investigate other proteins of the FA pathway, but using human cells. So far, all the cell lines used in this thesis were derived from hamster, which support high levels of human L1 retrotransposition; however, the use of model organisms allow to isolate mutant cells but also parental cells, which provide a specific control. However, to examine potential differences between human and rodent cells, I next analysed L1 retrotransposition (of WT and L1-ORF1p mutant elements) in two FA cell lines derived from human patients: ANKN (deficient in FANCI) and PD20F cells (deficient in FANCD2).

Notably, FANCI and D2 proteins, which are not part of the “FA core” complex, are interesting to investigate because they play unique and key roles in the FA pathway, outside of the core complex (Ceccaldi *et al.* 2016). In response to DNA damage, the FA core complex monoubiquitinates FANCD2, and this activated isoform of the protein forms a complex with FANCI (Smogorzewska *et al.* 2007). The FANCI:FANCD2 complex is loaded onto chromatin in a coating that is thought to provide protection for stalled DNA (Ceccaldi *et al.* 2016). Consistently, both proteins share a similar sequence and are likely to be derived from a common ancestor (Smogorzewska *et al.* 2007). Notably, FANCI is also monoubiquitinated, and studies have shown that ubiquitination of each protein is important for maintaining ubiquitination of the other (Smogorzewska *et al.* 2007).

Interestingly, a recent genome wide Loss of Function (LOF) CRISPR screen by Liu *et al.* (2018) identified both FANCI and FANCD2 as potential suppressors of L1 activity, as did an siRNA screen by Mita *et al.* (2021; **Table 5.1**). I therefore wanted to investigate L1-ORF1p-retrotransposition in these cell lines, using the battery of mutant elements that I’ve previously tested in other cell lines. Because these human cell lines are fibroblasts isolated from human patients, I do not have a parental line. However, we can still measure the rate of retrotransposition of mutant elements as a percentage of WT-L1 retrotransposition in the same cell line. Additionally, cDNA complementation can be used to confirm phenotypes. The results of these experiments are shown in **Figures 5.5** and **5.6**; remarkably, we observed similar retrotransposition rates in both cell lines, potentially due to their similar sequence and overlapping function. In fact, while we observed elevated levels of WT-L1 retrotransposition in FANCI and D2 deficient cells (**Figure 5.5 & 5.6**), all the L1-ORF1p mutants tested, class I or class II, retrotransposed at background levels, ranging from 0.02-0.06% the level of the WT-L1 (**Figure 5.5 &**

**5.6).** Consistently, for the class II L1-ORF1p mutants, these rates were lower than in any of the other cell lines tested.

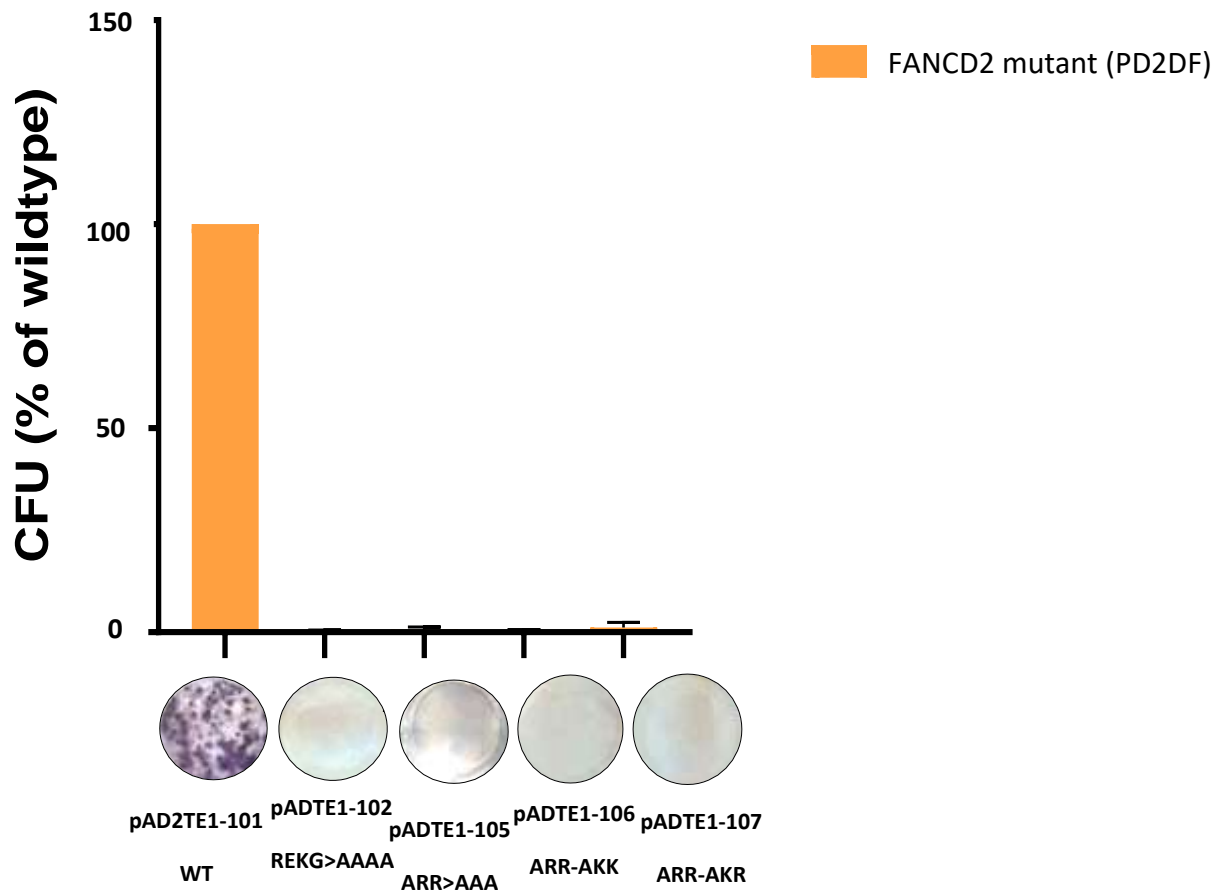
In summary, data included in this thesis have revealed that FANCA and FANCC deficient cells accommodate elevated levels of WT-L1 retrotransposition. Additionally, these data further revealed that FANCA deficient cells are unique among the cell lines tested thus far in accommodating elevated levels of L1-ORF1p mutant elements (i.e., L1-ORF1pi-retrotransposition).



**Figure 5.5 Testing a series of L1-ORF1p mutants in FANCI deficient cells.**

The retrotransposition of a WT-L1 (pAD2TE1-101) and of several class I (pAD2TE1-102 & 105) and class II (pAD2TE1-106 & 107) L1-ORF1p mutants was quantified using the cell culture retrotransposition assay, using FANCI deficient human cells. The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times (shown are mean $\pm$  SD).





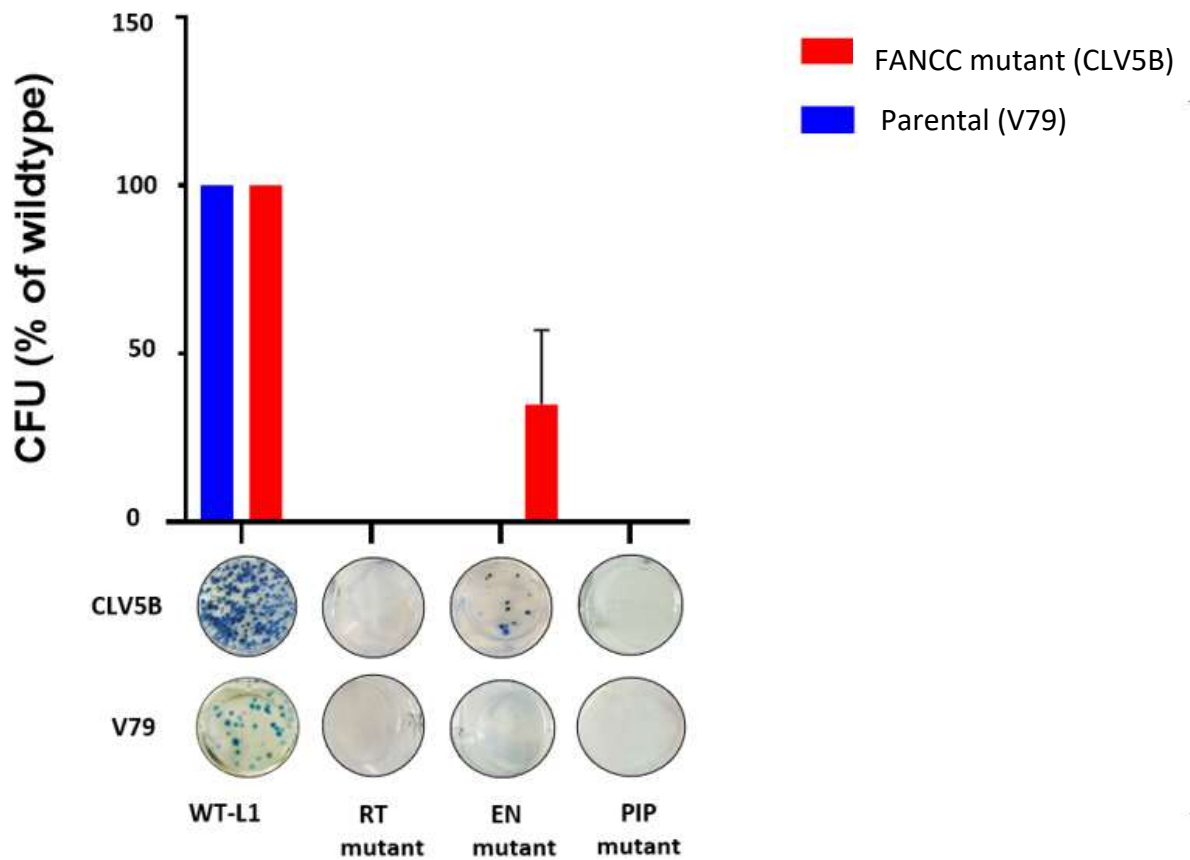
**Figure 5.6 Testing a series of L1-ORF1p mutants in FANCD2 deficient cells.**

The retrotransposition of a WT-L1 (pAD2TE1-101) and of several class I (pAD2TE1-102 & 105) and class II (pAD2TE1-106 & 107) L1-ORF1p mutants was quantified using the cell culture retrotransposition assay, using FANCD2 deficient human cells. The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times (shown are mean +/- SD).

## 5.7 Testing the retrotransposition of L1-ORF1p mutants in FANCC mutant cells.

To further explore the extent of L1 deregulation in FA mutant cells, I next investigated whether L1-ORF2 mutant elements could retrotranspose in FANCC deficient cells. As described in Chapter 4, I found that elements carrying mutations in the EN and PIP domains of L1-ORF1p retrotransposed at a high level in FANCA deficient cells, and that these phenotypes (i.e., ENi- and PIPi-retrotransposition) could be rescued upon complementing cells with rodent FANCA cDNA (see Chapter 4).

To explore ENi and PIPi-retrotransposition in FANCC deficient cells, I used engineered pAD2TE1 L1 vectors (containing T7 and TAP carboxyl-terminal epitope tags in L1-ORF1p and L1-ORF2p, respectively, and the *mneol* retrotransposition indicator cassette) carrying missense inactivating mutations in the EN and PIP domains of L1-ORF1p (D/HMoreA and YY414/15AA, respectively). As controls, FANCC deficient (CLV5B) and parental (V79) cells were transfected with a WT-L1, and RT-mutant allelic plasmid or with pT2neo (**Figure 5.7**). As previously observed, I detected significant higher retrotransposition of WT-L1s in FANCC deficient cells when compared to parental cells (**Figure 5.7**). Notably, I found that FANCC deficient cells accommodate elevated levels of ENi-retrotransposition, at similar levels as those previously observed in FANCA deficient cells (approximately 35-40% of WT-L1, **Figure 5.7**). Interestingly, the L1-ORF1p PIP-mutant element retrotransposed at background levels in FANCC deficient and parental cells (**Figure 5.7**), in stark contrast with FANCA deficient cells. In sum, upon exploring if FANCC deficient cells might support the unusual retrotransposition of L1-ORF1p mutant elements, I found that only ENi-retrotransposition seems to be prevalent in FANCC mutant cells.



**Figure 5.7 Testing a series of L1-ORF1p mutants in FANCC deficient cells.**

The retrotransposition of a WT-L1 (pAD2TE1-101) and of several L1-ORF1p mutants (EN-mutant, D/H230A; PIP-mutant, YY414/15AA; and RT-mutant, D702A) was quantified using the cell culture retrotransposition assay, using FANCC deficient (CLV5B) and parental (V79) hamster cells. The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times (shown are mean  $\pm$  SD). \* indicates  $p < 0.05$ , Student's t-test

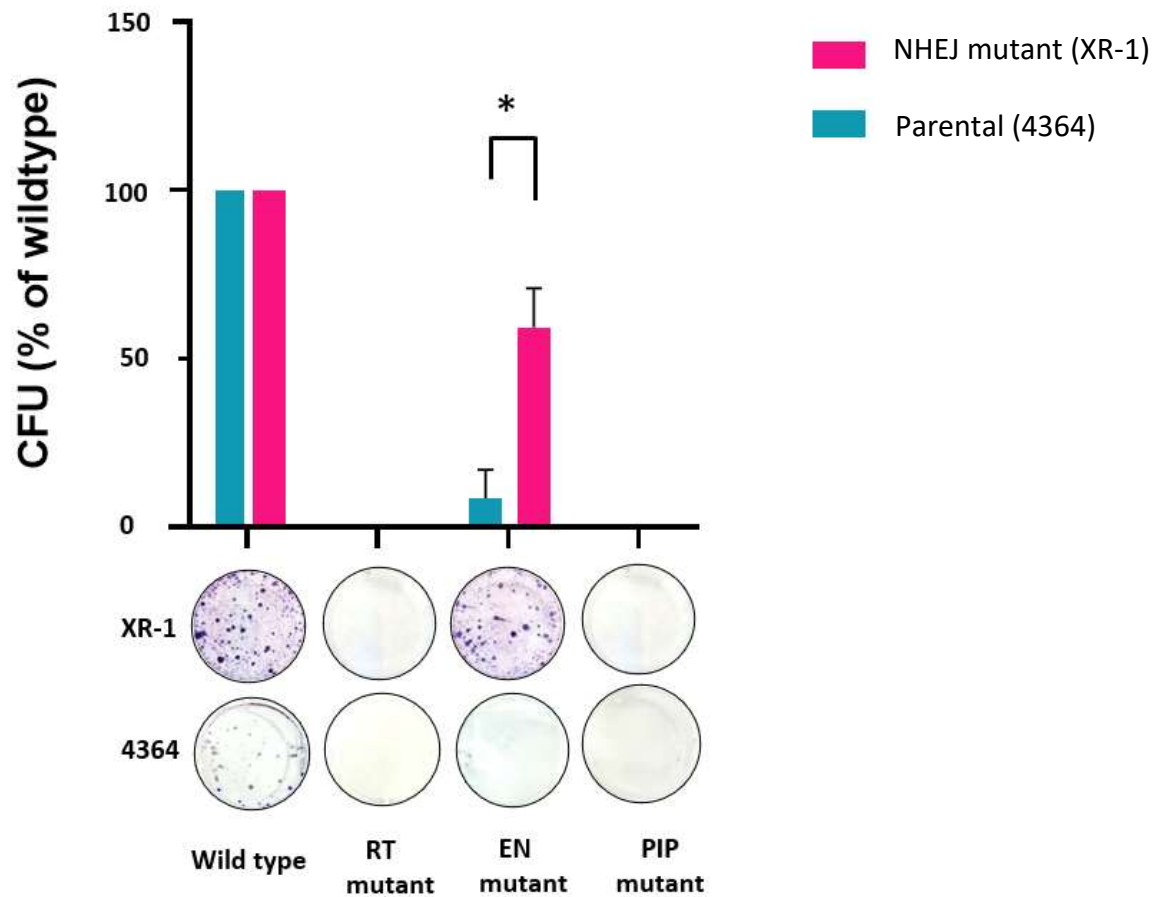
## 5.8 Testing the retrotransposition of L1-ORF1p mutants in NHEJ cells.

Collectively, the data in this thesis indicates that FANCA deficient cells are permissive for multiple alternative pathways of L1 integration, while FANCC, D2 and I deficient cells support ENi-retrotransposition but not L1-ORF1pi- or PIPi-retrotransposition. To compare the extent of ENi-retrotransposition with NHEJ mutant cells, I next tested how my series of L1-ORF1p mutant elements behave in XR-1 cells (deficient in XRCC4/DNALig4), using parental cells (4364 cells) as a control. These and other NHEJ-mutant cells were previously shown to support ENi-retrotransposition (Morrish *et al.* 2002; Morrish *et al.* 2007); however, whether PIPi-retrotransposition can occur in NHEJ-mutant cells remains to be determined.

Using pAD2TE1 L1 vectors, I found that NHEJ deficient cells supported elevated levels of ENi-retrotransposition (**Figure 5.10**). However, retrotransposition of the PIP mutant L1 in these cells was negligible, with values <1% of WT-L1 (**Figure 5.10**). Thus, consistent with previous reports (Morrish *et al.* 2002; Morrish *et al.* 2007), deletion of XRCC4/DNA Ligase IV appears to enable retrotransposition of EN mutant elements, but not other types of mutant elements tested in this thesis.

Collectively, my data suggest that FANCA deficient cells can accommodate L1-ORF1pi-, ENi- and PIPi-retrotransposition, while ENi-retrotransposition is observed in several other FA complementation groups (C, D2 and I), and in NHEJ mutant cells (XRCC4/DNALig\_4 and DNAPKcs mutant cells; Morrish *et al.* 2002; Morrish *et al.* 2007). Indeed, based on these findings, I speculate that PIPi- and L1-ORF1pi-retrotransposition take place by a fundamentally different mechanism than ENi-retrotransposition.

Notably, my data extend previous studies (Morrish *et al.* 2002; Morrish *et al.* 2007; Flasch *et al.* 2019) and further confirm a clear functional interaction between DNA repair pathways and ENi-retrotransposition. However, it is worth noting that the level of ENi-retrotransposition reported in FA deficient cells is proportionately less than in NHEJ mutant cells (XR-1 cells), which supported ENi retrotransposition at around 60% of WT-L1 levels. I further speculate that this difference may relate to the types of DNA lesions accumulated in the different cell types, and how efficiently L1 can exploit these lesions to initiate retrotransposition

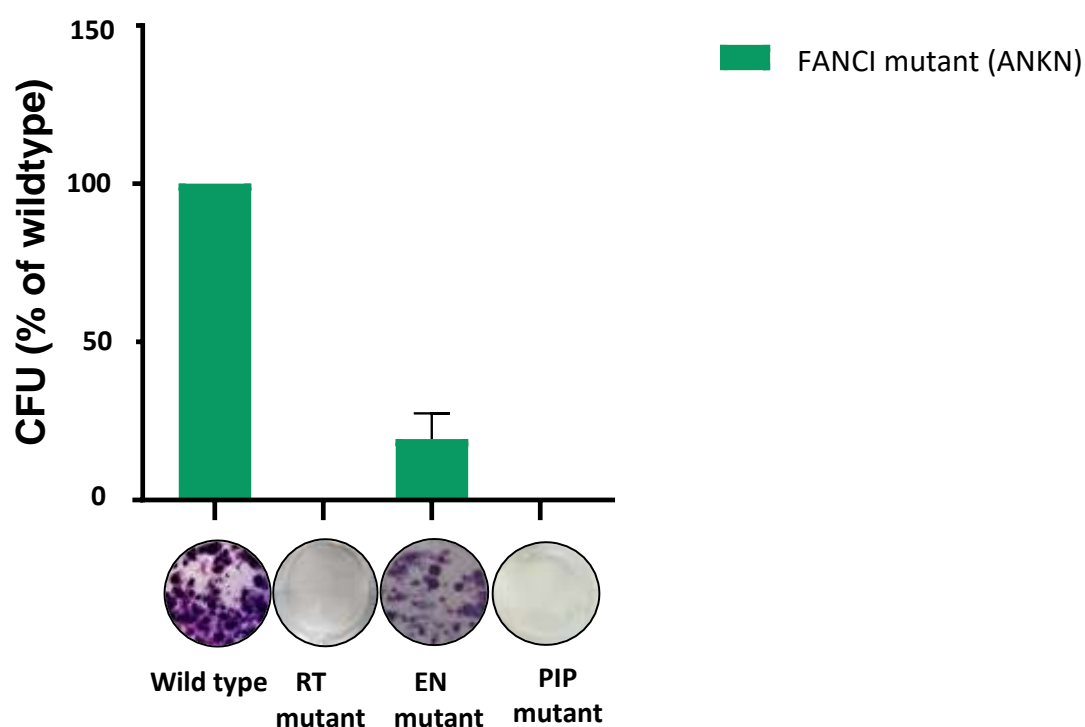


**Figure 5.10 Testing a series of L1-ORF1p mutants in NHEJ deficient cells.**

The retrotransposition of a WT-L1 (pAD2TE1-101) and of several L1-ORF1p mutants (EN-mutant, D/H230A; PIP-mutant, YY414/15AA; and RT-mutant, D702A) was quantified using the cell culture retrotransposition assay, using hamster NHEJ deficient cells (XR-1, deficient in XRCC4/DNALig\_4). The number of colony forming units (CFU) per well was quantified and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times (shown are average +/- SD). \* indicates  $p < 0.05$ , Student's t-test

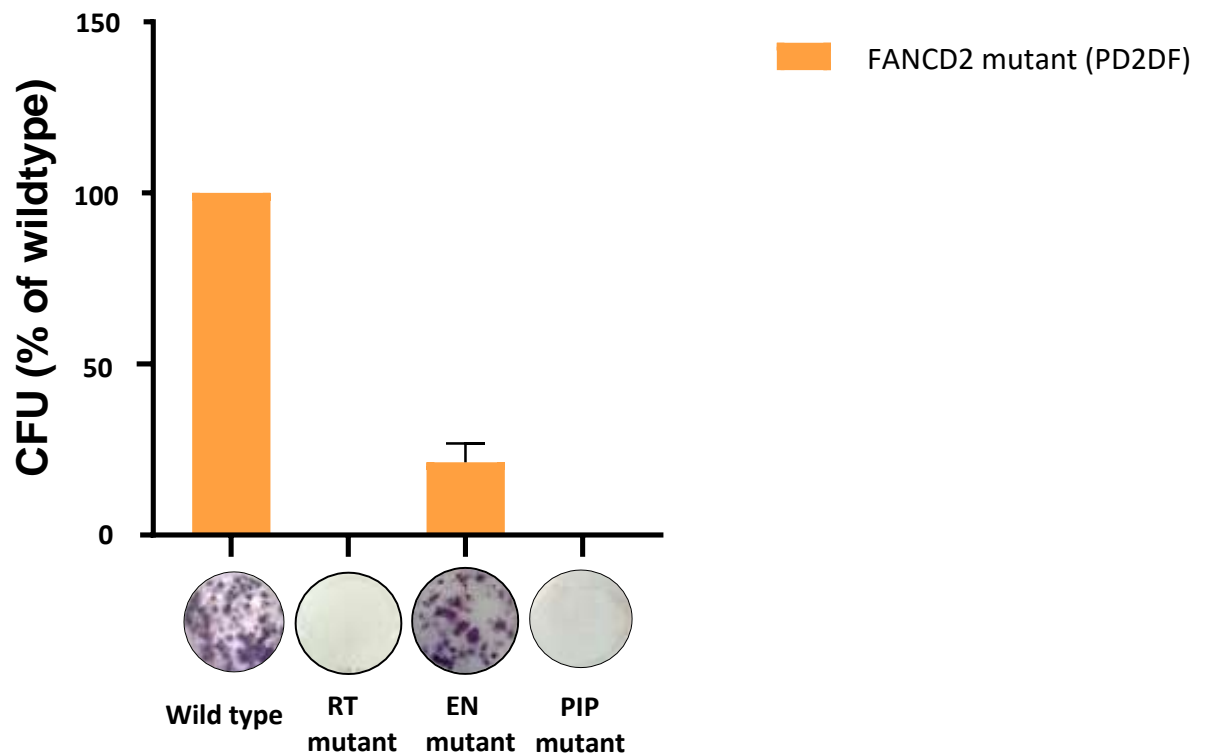
## 5.9 Testing the retrotransposition of L1-ORF1p mutants in FA deficient human cells.

Given that FANCA, FANCA and DNA repair pathway mutant cell lines support ENi-retrotransposition, I next wanted to test whether the patient cell lines with mutations in the non-core FA complex components FANCI and FANCD2 that I used earlier also support ENi-retrotransposition. Notably, and similar to the other cell lines deficient in DNA repair proteins tested in this thesis, FANCI and D2 deficient cells supported elevated ENi retrotransposition (**Figure 5.8 & 5.9**). The retrotransposition rates of the endonuclease mutant element were approximately 20% of WT/L1 levels (**Figure 5.8 & 5.9**). Remarkably, as observed in hamster FANCC deficient cells, no PIPi-retrotransposition was observed in human FA cells, with retrotransposition levels <1% of WT-L1 levels (**Figure 5.8 & 5.9**). Thus, I conclude that while ENi-retrotransposition is elevated in all FA deficient lines tested, L1-ORF1pi- and PIPi-retrotransposition is restricted to FANCA cells.



**Figure 5.8 Testing a series of L1-ORF1p mutants in FANCI deficient cells.**

The retrotransposition of a WT-L1 (pAD2TE1-101) and of several L1-ORF1p mutants (EN-mutant, D/H230A; PIP-mutant, YY414/15AA; and RT-mutant, D702A) was quantified using the cell culture retrotransposition assay, using human FANCI deficient cells. The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times (shown are average  $\pm$  SD).



**Figure 5.9 Testing a series of L1-ORF1p mutants in FANCD2 deficient cells.**

The retrotransposition of a WT-L1 (pAD2TE1-101) and of several L1-ORF1p mutants (EN-mutant, D/H230A; PIP-mutant, YY414/15AA; and RT-mutant, D702A) was quantified using the cell culture retrotransposition assay, using human FANCD2 deficient cells. The number of colony forming units (CFU) per well was quantified and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three times (shown are average  $\pm$  SD).

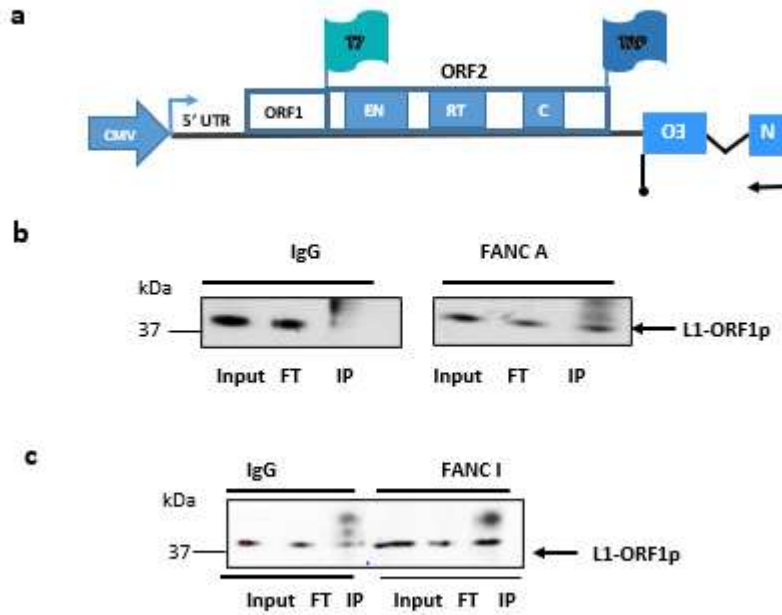


## 5.10 Testing for an interaction between FA proteins and L1 encoded proteins

From this work so far, there is evidence to suggest that FANCA is a regulator of L1 retrotransposition, and loss of the protein enables a unique mechanism of retrotransposition which does not appear to require the formation of a canonical L1-RNP, or the enzymatic activities of the EN and PIP domains of L1-ORF1p. In this chapter, my results demonstrate that these results do not extrapolate to NHEJ deficient cells or to other FA proteins, even FANCC, which also forms part of the core complex. Several large-scale screens have identified FANCA as a potential regulator of L1 but it is not known whether FANCA regulate L1 retrotransposition through direct interaction with L1 or by more indirect mechanisms. As deletion of FANCA enables mobilisation of L1-ORF1p mutant L1s, I wanted to test if FANCA could interact with L1-ORF1p. Alongside this experiment, I investigated whether another protein, FANCI, which is not part of the “core FA” complex, would interact L1-ORF1p. FANCI has also been identified as a potential L1 regulator, but in contrast to FANCA, FANCI deficient cells can only support ENi-retrotransposition.

To examine interaction of FA proteins with L1-encoded proteins, I used co-immunoprecipitation and specific antibodies to endogenously expressed FA proteins. Notably, the engineered L1 vectors used in this Chapter, pAD2TE1 series, contain T7- and TAP-epitope tags in L1-ORF1p and L1-ORF1p, respectively (see scheme in **Figure 5.11a**), allowing me to use robust and specific antibodies to detect L1-ORF1p on immunoprecipitates, using western-blotting. Thus, I transfected HEK293T cells with the WT-L1 plasmid where the C-terminus of L1-ORF1p bears a T7 tag. I then used anti-FANCA, anti-FANCI or control non-specific IgG antibodies to immunoprecipitate these proteins from transfected cells, and Western blotted the immunoprecipitates with anti-T7 antibodies to assess whether T7-tagged L1-ORF1p co-immunoprecipitates with these FA proteins

T7-tagged L1-ORF1p appeared to be more enriched in anti-FANCI or anti-FANCA immunoprecipitates than in control IgG immunoprecipitates (**Figure 5.11b&c**), indicating that there is some interaction between L1-ORF1p and the FA proteins. I noticed a minor and non-specific binding of L1-ORF1p either to the control IgG or to protein G beads in WCLs incubated with the IgG control in both co-IPs (Figure 5.11b&c). Whether the observed interaction is dependent on the formation of a L1-RNP, or whether the interaction occur through RNA and/or accessory proteins, remains to be determined. However, when considering the distinct retrotransposition phenotypes observed in FANCA and I deficient cells, I speculate that the interaction of FA proteins with L1-ORF1p might not be relevant to mechanistically distinguish ENi-retrotransposition from L1-ORF1p- or PIPi-retrotransposition.



**Figure 5.10 Co-immunoprecipitation of FA proteins and T7 tagged L1-ORF1p.**

**a)** Scheme of pAD2TE1-101/L1.3. See text for additional details.

**b&c)** Co-immunoprecipitation of FANCA (b) and I (c) with L1-ORF1p. Co-Ips were performed in HEK293T cells transfected with pAD2TE1-101/L1.3 using an antibody to FANCA (b), FANCI (c) or an IgG antibody (**b&c**) as a negative control. Samples were resolved using SDS-PAGE, transferred and blotted with an anti-T7 antibody.

## 5.11 Main conclusions

To conclude, this chapter provides further evidence that FANCA deficient cells support unique retrotransposition pathways of integration, such as L1-ORF1pi- and PIPi-retrotransposition. Indeed, the L1-ORF1p mutants which mobilised in FANCA cells were not active in NHEJ deficient cells, or in other cells deficient in FA proteins. Neither was the PIP mutant element. As expected, EN mutant elements were consistently active in cell lines deficient in DNA repair proteins.

## 5.12 Discussion

In this chapter, building from data presented in the previous chapter, I examined how penetrant was the L1 phenotype described in FANCA deficient cells. Although I found that there is a unique phenotype associated with FANCA cells and a more generic phenotype in FANCC, D2 and I cells, as well as in NHEJ mutant cells, this series of experiments brought several interesting conclusions. Firstly, I was able to establish that FANCC also limits L1 retrotransposition. This aligns with data by Tristan Ramos (2020), which demonstrated restriction of L1 retrotransposition by FANCC using siRNAs. Interestingly, a genome wide siRNA screen by Mita *et al.* 2020 identified FANCC as a positive regulator of L1. Thus, the nature of the relationship may be context dependent or require further experimentation to elucidate.

I was also able to establish that the capacity to support retrotransposition of severe L1-ORF1p mutant elements is not shared by cells deficient in NHEJ proteins, or indeed other proteins in the FA pathway, including FANCC, D2 and I. This suggests that, unlike ENi-retrotransposition, ORF1pi-retrotransposition is not a generic feature of cells deficient in DNA repair or FA cells. Similarly, I also established that PIPi-retrotransposition is unique to FANCA cells, not being neither a generic feature of cells deficient in DNA repair or FA cells.

During ENi retrotransposition, it is theorised that elements are exploiting lesions resulting from the lack of DNA repair proteins. Mobilising into these DNA breaks means that the element does not need its own endonuclease to create an incision in the DNA to initiate integration. (Morrish *et al.* 2002) It is possible, then, that FA related lesions could be exploited by the element in a similar way, explaining why these elements can move without domains which are usually required for retrotransposition. As the FA pathway plays a key role in preventing replication fork collapse, FA cells are characterised by the delayed repair kinetics of double stranded breaks as well as accumulations in stalled replication

forks and R loops (Houghtaling *et al.* 2005; Garcia-Rubio *et al.* 2015; Renaud *et al.* 2016). Indeed, I speculate that the marked differences in how penetrant is the L1 phenotype in FA cells might be directly related to the main type of DNA lesion accumulated in FA cells and in NHEJ cells.

R-loops are transcription associated DNA: RNA hybrids, which contribute to genome instability in FA cells (Garcia-Rubio *et al.* 2015). Some FA proteins, including FANCD2, interact directly with R-loops. One possibility from my findings is that LINE-1 elements could somehow exploit R-loops to facilitate the abnormal retrotransposition I observed here. However, in this case we would expect to see the abnormal retrotransposition in each of the cells with a FA phenotype, as these should all exhibit increased R-loops. Experimentation in FANCC mutant cells demonstrated an increase in wild type retrotransposition, but not an increase in ORF1p<sup>-</sup> or PIP<sup>-</sup> retrotransposition. Thus, it is possible that FANCC also works as a negative regulator of LINE-1 activity, but does not support a mechanism of retrotransposition which negates the need for the PIP domain or formation of a canonical RNP.

An average human genome contains 80-100 active L1s, and their activity continues to impact the human genome. Notably, patients carrying biallelic inactivating mutations in FANCA cells may have a significantly higher pool of active L1s in their genomes, as elements carrying mutations in L1-ORF1p and some domains of L1-ORF1p (EN and PIP) could retrotranspose in patients. Intriguingly, FANCA is by far the most commonly mutated gene in FA patients (>65% of all patients) and it seems paradoxical that this particular mutation is associated with the highest rate of endogenous retrotransposition, which can drive mutagenesis. Interestingly, ENi-retrotransposition is thought to represent an ancestral mechanism of L1 insertion, used prior to the capture/evolution of an EN domain by LINEs (Malik *et al.* 1999), as a basic mechanism of DNA repair that can repair DSBs (i.e., how a band aid is used on a cut; Eikbush *et al.* 2002). Accurate and error-free DNA repair is instrumental through genome evolution, and DNA repair is a heavily regulated process. However, the accumulation of DSBs has been associated with genomic instability, lethality and carcinogenesis (Peirce *et al.* 2001), and repairing DSBs through retrotransposition definitively represents a better scenario than accumulating DSBs. Thus, the apparent paradoxical activation of a large pool of L1 elements in cells from FANCA patients might be associated with DNA-repair processes, explaining why FANCA might be the most common mutation characterized in patients. However, the rarity of the disease and its genetic complexity (i.e., reported founder effects (Garcia de Teresa 2019), are factors that could explain the high prevalence of FANCA mutations, and future studies are warranted to solidify this hypothesis.

Finally, my research supports previous proteomic work suggesting that FANCI can interact with L1-ORF1p. Additionally, here I demonstrated that antibodies to FANCA and FANCI can co-IP exogenously expressed L1-ORF1p, consistent with several large screens that have identified both of these proteins

as potential L1 regulators. At a mechanistic level, the fact that FANCA and FANCI can both interact with L1-ORF1p but the associated L1 phenotypes are so distinct, suggests the possibility that FANCA and FANCI interact with different domains of ORF1p and regulate retrotransposition through distinct ORF1p-dependent mechanisms.



## Chapter 6

---

# **Towards a mechanism for L1 retrotransposition in a FA context**





## 6.1 Introduction

The use of mutant L1 elements in retrotransposition assays revealed that retrotransposition in FA deficient cells can take place via a different mechanism/s to canonical TPRT. On one hand, cell based retrotransposition assays in cell lines deficient in FANCA, C, D2 and I revealed that mutant L1s lacking an ENdonuclease (EN) domain can retrotranspose using an EN-independent (ENi) mechanism, but that this mechanism is distinct from ENi-retrotransposition reported in Non-Homologous End Joining (NHEJ) deficient cells (Morrish *et al.* 2007; Morrish *et al.* 2002). In fact, ENi-retrotransposition in FA cells does not require p53 inactivation and the hallmarks associated with *de novo* retrotransposition events are similar for EN-proficient (i.e., Wild type L1s) and EN-deficient L1s (EN mutant L1s) (S.Morell, unpublished). Notably, throughout the characterization of >100 *de novo* L1 insertions from WT- and EN-mutant L1s, our lab found that insertions in FA cells are often associated with long Target Site Duplications (TSDs, >100bp) flanking the integrated L1, DNA deletion processes at the site of integration and can be associated with complex DNA rearrangements at the site of integration (i.e., presence of palindromic sequences captured at the 5' end of *de novo* insertions). In contrast, ENi-retrotransposition in NHEJ cells requires p53 inactivation (Coufal *et al.* 2011) and retrotransposition hallmarks of EN-proficient (i.e., Wild type L1s) and EN-deficient L1s (EN mutant L1s) are very different in structure: while WT-events are canonical (have a polyA tail and normal size TSDs(2-20bp)), ENi-events are often 3' truncated and are associated with alterations at the site of integration (deletions and inversions).

Remarkably, besides ENi-retrotransposition, the data included in this Thesis also uncovered that FANCA deficient cells can accommodate the retrotransposition of L1-ORF1p and PCNA Interaction Protein (PIP)-mutant L1s, suggesting that the main L1 retrotransposition mechanisms are different in FANCA deficient cells than in other FA complementation groups or in NHEJ-mutant cells. In sum, while FANCC, D2 and I mutant cells, as well as NHEJ mutant cells, can accommodate ENi-retrotransposition, FANCA cells can accommodate ENi-, L1-ORF1p- and PIPi-retrotransposition. Notably, FANCA is the most commonly mutated gene in patients (>65% of patients), further suggesting that L1 deregulation could be an important determinant for the reported genomic instability of FA patients.

In this Chapter, I decided to investigate the mechanism of L1 deregulation in FA cells, focusing on FANCA cells because of its dramatic retrotransposition phenotype. I speculate that L1 elements could use a particular kind of DNA lesion accumulated in FA cells to promote their integration, similarly to

how EN-mutant L1 elements are thought to exploit increased numbers of unrepaired Single and Double Strand Breaks (SSBs/DSBs) in NHEJ mutant cells as sites of integration (Morrish *et al.* 2007; Morrish *et al.* 2002). Consistent with this hypothesis, FA cells are known to accumulate several particular lesions which are characteristic of the FA phenotype. One example are stalled replication forks. Replication forks naturally stall as they come across DNA lesions, and in particular in the encountering of interstrand cross-links (ICLs), which are typically overproduced in FA cells (Renaudin and Rosselli 2020). The FA pathway plays a key role in restarting stalled forks and as such the depletion of the pathway results in a build-up of stalled forks. This safeguarding activity of the FANC/BRCA pathway of DNA repair is believed to take place through cross talk with other pathways, including the mTOR pathway (Nolan *et al.* 2021) and the translesion synthesis pathway (Renaud and Rosseli 2013). A recent paper by Mita *et al.* (2020) measured L1 retrotransposition frequency in cells treated with different concentrations of aphidicolin (APH) which is known to slow and stall fork progression (Ikagami *et al.* 1978). They observed an increase in L1 insertions as the APH caused cells to build up at S phase, potentially indicating that L1 is recruited to stalled forks and is able to utilise them as integration substrates.

Similar to ICLs, FA cells also accumulate R-loops, a DNA lesion made up from an RNA: DNA hybrid and a displaced ssDNA (Garcio Rubio *et al.* 2015). Thus, it is possible that R-loop structures would represent a good substrate to promote L1 retrotransposition, particularly as they leave stretches of exposed ssDNAs (i.e., exposed 3'OH groups).

While my work throughout this Thesis has highlighted unusual features of retrotransposition in FANCA cells, I have not, to this point, investigated the potential mechanism/s of L1 deregulation. In this chapter, I applied several experimental approaches to further investigate mechanistic aspects of L1 retrotransposition in FANCA deficient cells. First, I decided to use Proteomics to further unveil L1 regulation in FA cells; briefly, I used L1-ORF1p epitope-tagged L1 constructs to perform mass spectrometry analyses to identify candidate host factors which may facilitate a different mechanism of retrotransposition in FA cells (comparing with parental). Interestingly, these analyses uncovered L1-ORF1p specific interactors associated with the metabolism of R-loops, a DNA lesion frequently accumulated in FA cells. Thus, in the second part of this chapter, I also wish to investigate the relationship between L1 retrotransposition regulation and the metabolism of R-, to explore one potential mechanism behind the altered frequencies of retrotransposition in FANCA cells. In short, my research questions for this chapter are as follows:

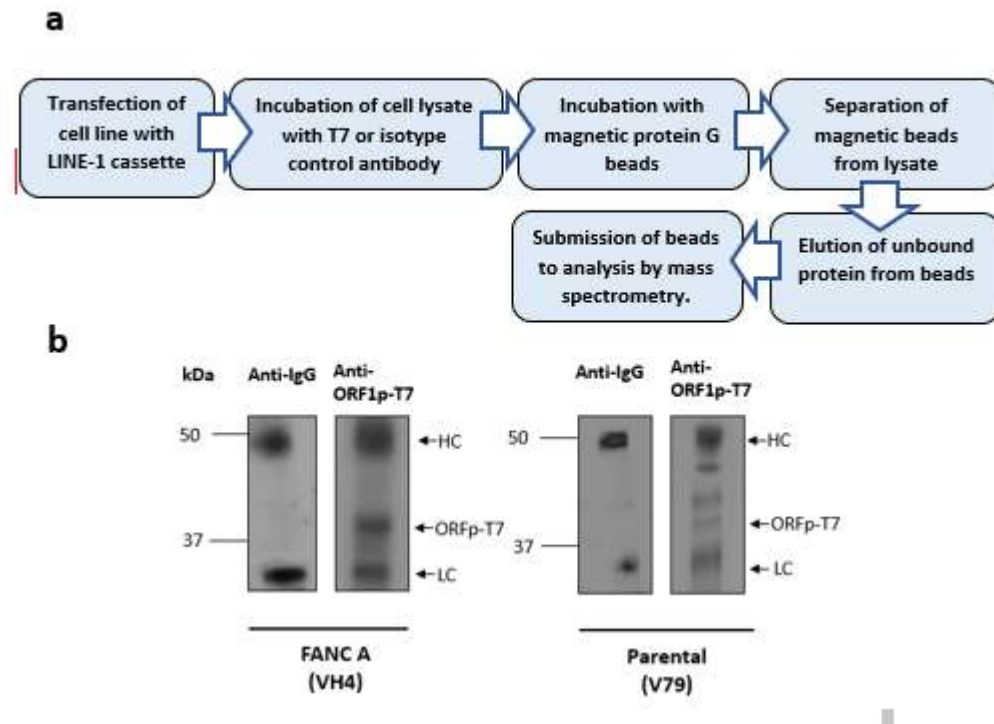
Aims:

1. Do different networks of host factors interact with L1-ORF1p in FANCA deficient (VH4) and parental (V79) cells?
2. Does L1 retrotransposition occur at accumulated R-loops in FANCA deficient cells? (i.e., does the alteration of R-loop levels influence the rate of L1 retrotransposition in FA cells?).

## 6.2 L1-ORF1p proteomic studies in FANCA deficient cells: identification of host interactors with L1-ORF1p.

Throughout this Thesis, I exploited several engineered L1 vectors, the pAD2TE1 vector series, to investigate L1 deregulation in a panel of DNA repair deficient cell lines. Notably, in pAD2TE1 vectors, L1-encoded ORFs contain unique T7 (L1-ORF1p) and TAP (L1-ORF1p) epitope tags at their carboxyl terminus, allowing me to use these same vectors on proteomic studies. Notably, previous studies revealed how L1 proteomic approaches represent a valid approach to identify a wide array of L1 regulators (Taylor *et al.* 2016; Taylor *et al.* 2018; Mita *et al.* 2018; Dai *et al.* 2014; Taylor *et al.* 2013). However, these same studies described a significant limitation of proteomic approaches, as only L1-ORF1p is expressed at a level compatible with the requirements of immunoprecipitation/mass spectrometry protocols, at least when proteins are expressed from the natural full-length bicistronic L1 RNA. In fact, this limitation is associated with how poorly L1-ORF1p is translated from the bicistronic L1 RNA (Alisch *et al.* 2006). However, I reasoned that I would be able to use pAD2TE1 vectors to identify L1-ORF1p interactors in FA cells. Intriguingly, I demonstrated that FANCA deficient cells could support retrotransposition of L1-ORF1p mutant elements, further supporting that the identification of L1-ORF1p interactors in FANCA deficient cells might clarify the mechanism by which L1 retrotransposition occurs in VH4 cells. I therefore tested whether L1-ORF1p could interact with a different set of proteins in FANCA deficient cells when compared with the parental cell line (**Figure 6.1**). To do this, I transfected FANCA (VH4) and parental (V79) cells with the pAD2TE1 vector containing a WT-L1 (pAD2TE1/101-L1.3) and then performed co-immunoprecipitation using the anti-T7 specific antibody (i.e., pull down of L1-ORF1p using a T7 antibody, **Figure 6.1**). Western-blot controls confirmed successful immunoprecipitation of L1-ORF1p, and of associated host factors, in FANCA deficient and parental cells (**Figure 6.1**)

After the completion of QC experiments, the sample containing L1-ORF1p and bound interactors was then analysed using mass spectrometry (**Figure 6.1**). In the mass spectrometry analyses, L1-ORF1p and interactors were analysed in triplicate, using isotypic IgG (triplicate) as an internal negative control.



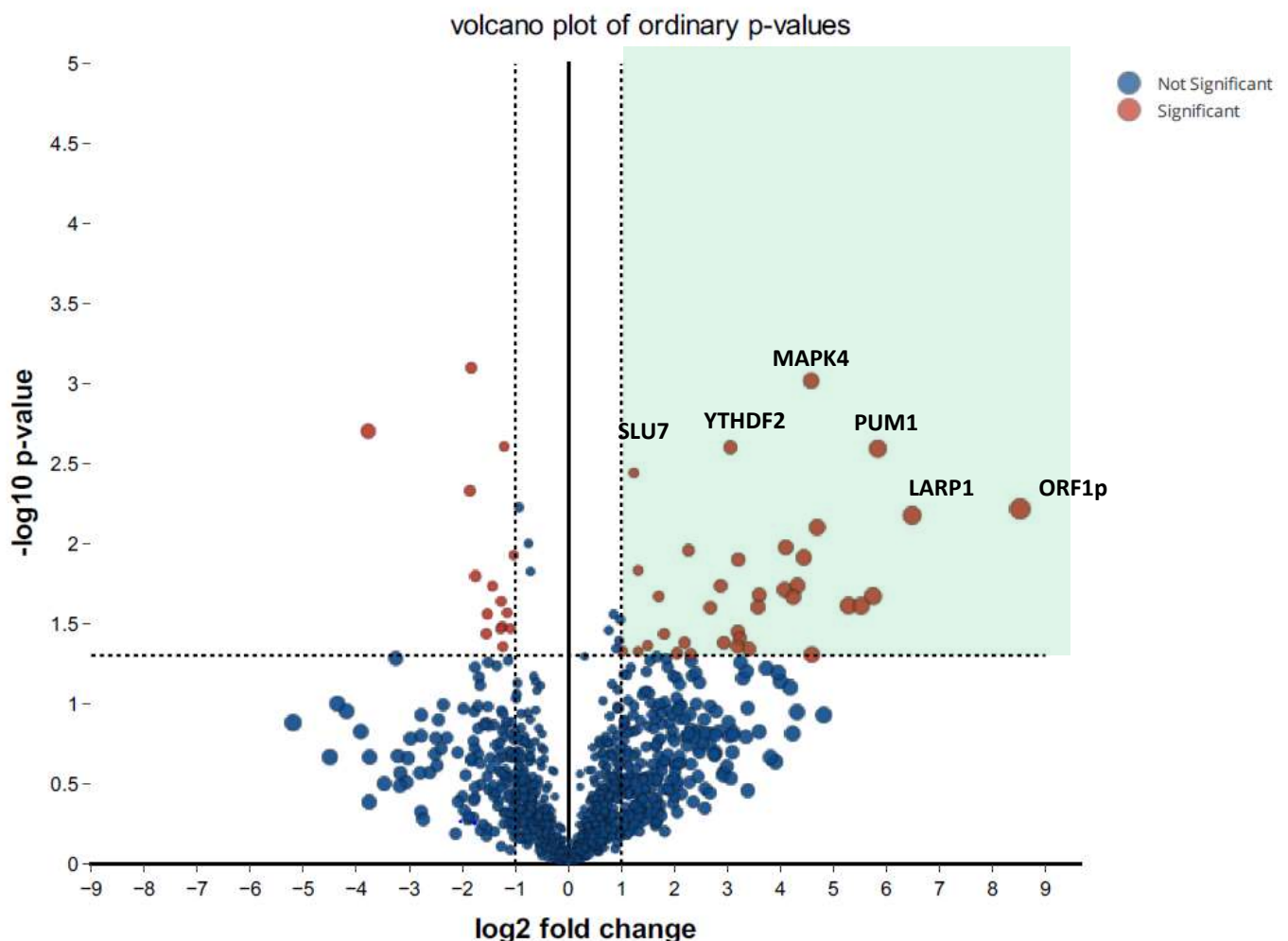
**Figure 6.1: Immunoprecipitation of L1-ORF1p-T7 interactors**

- Flow chart of the methodology used to immunoprecipitate L1-ORF1p using a T7 tag and an anti-T7 antibody. Upper and lower bands are presumed to be heavy chain (HC) and light chain (LC) of the antibody. See text for additional details.
- Western Blot demonstrating the presence of T7 tagged ORF1p in IP on protein G beads following pull down.

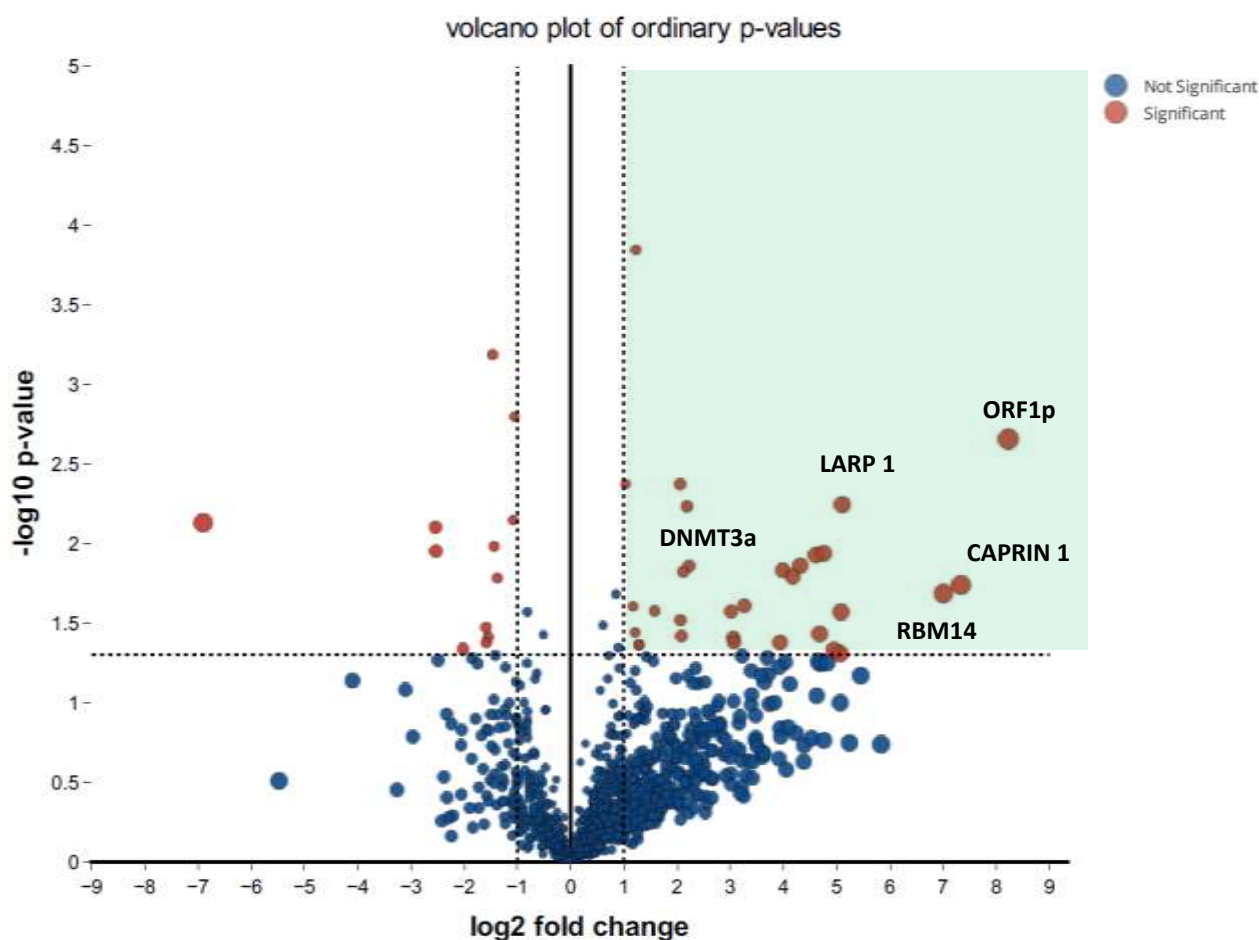
Notably, bioinformatic analyses using the proteomic datasets revealed a good correlation among the biological replicates for L1-ORF1p samples, both in VH4 and V79 cells.

To define the pool of specific L1-ORF1p interactors in FANCA deficient (VH4) and parental (V79) cells, I used a pipeline that included several inclusion criteria steps. Briefly, interactors were defined as proteins of a log Fold Change (log<sub>2</sub>FC) of more than 1, and with a p-value of <0.05.

Remarkably, proteomic experiments in FANCA deficient (VH4) and parental (V79) cells identified a selection of proteins which were significantly enriched in the L1-ORF1p samples (i.e., T7 antibody, **Figure 6.2&6.3**) as opposed to the negative control (i.e., Ig G antibody, **Figure 6.2&6.3**). As expected, L1-ORF1p was heavily enriched in the sample containing the T7 antibody in comparison to the negative control, together with additional host factors (**Figure 6.2&6.3**).



**Figure 6.2: Volcano plot showing MS analysis of proteins following immunoprecipitation of L1-ORF1p-T7 in FANCA (VH4) cells.** P-value of <0.05 ( $-\log_{10}$ ), indicating significance of the hit, is plotted against fold enrichment ( $\log_2$ ) of the protein in a sample, representing results from 3 independent experiments. Several of the most enriched proteins (shown as red dots) are indicated in the graph. See text for additional details.



**Figure 6.3: Volcano plot showing MS analysis of proteins following immunoprecipitation of L1-ORF1p in parental (V79) cells.** P-value of  $<0.05$  ( $-\log_{10}$ ), indicating significance of the hit, is plotted against fold enrichment ( $\log_2$ ) of the protein in a sample, representing results from 3 independent experiments. Several of the most enriched proteins (shown as red dots) are indicated in the graph. See text for additional details.

A list of the proteins which was found to be significantly enriched (defined as having a log Fold Change ( $\log_2\text{FC}$ ) of more than 1, and with a p-value of  $<0.05$ ) after a pull down with a T7 antibody is shown in **Table 6.1**. Of these interactors, 36 were common between the two cell lines, 24 were unique to FANCA cells and 31 were unique to the parental cell line.

Interestingly, several of the interactors shared between both cell lines were previously identified as L1-interactors in human cell lines. (Taylor *et al.* 2013; Dai *et al.* 2014; Taylor *et al.* 2016; Taylor *et al.* 2018; Mita *et al.* 2018)

**Table 6.1 List of significantly enriched ORF1p-T7 interactors in FANCA and parental cells**

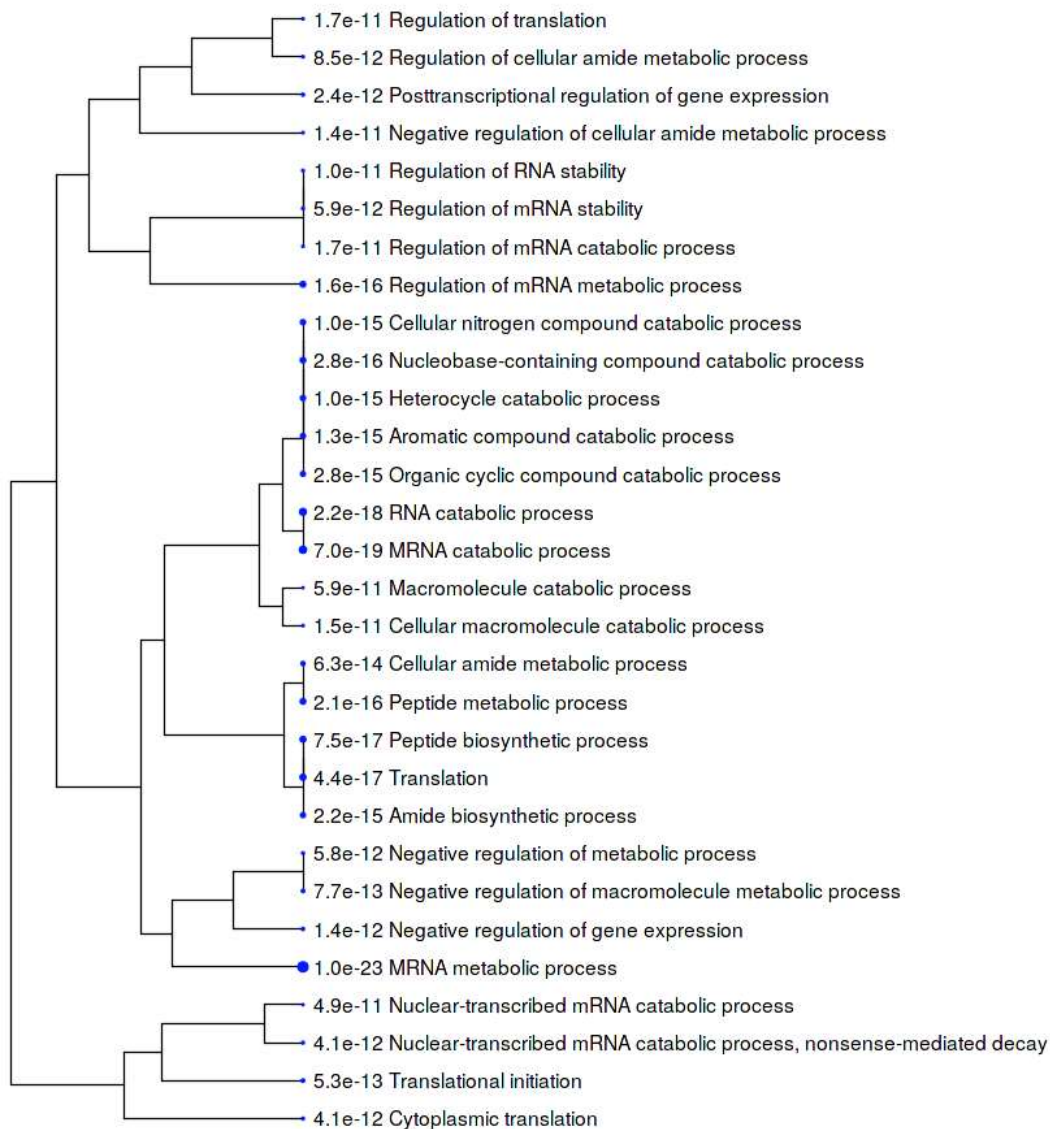
ORF1p interactors					
FANCA (VH4)		Parental (V79)		Common	
UPF1	KHSRP	RPL35	SMARCC2	LARP1	PABPC4
MAP4K4	DDX5	RPL19	RPS11	PUM1	SF3B2
CREBBP	ARHGEF2	G3BP2	TARDBP	ANKRD17	SF1
SEC23B	OTUD4	ARID1A	SMARCB1	ATXN2L	EIF2S2
NUFIP2	U2AF2	TFG	KHDRBS1	RPS21	RPL26
CALML5	CFL1	RPL18	TCP1	RPL15	RBM14
NFAT5		HNRNPA1	ABCF1	RPS27L	G3BP1
PABPC1		SRSF6	TSR1	ATAD3A	RPL21
ZC3H11A		HNRNPM	DNMT3A	EIF4G1	RPL6
WDR33		RPL39P5	DNAJC2	CAPRIN1	SNRNP70
SRSF2		SMARCA4	NCOA5	FAM120A	RPL9P9
YTHDF2		NONO	SERPINH1	SERBP1	SMARCC1
EEF2		ALYREF	PSMD2	CHD8	FXR2
GAPDH		CHD8		FXR1	HNRNPK
TUBA1B		RPS27		SF3B1	RPS26
YBX1		SMARCE1		BAIAP2	ATAD3B
DNAJA1		DDX3X		HSPA9	DDX17
RPL8		CAV1		SEC24C	RPL29

Overall, proteomic experiments in FANCA deficient cells identified 61 significantly enriched L1-ORF1p interactors. However, ~60% of them were identified in previous L1 proteomic studies using other cell lines, confirming the reliability of our approach and data (Taylor *et al.* 2013; Dai *et al.* 2014; Taylor *et al.* 2016; Taylor *et al.* 2018; Mita *et al.* 2018). Notably, GO term analyses of these proteins revealed that 31 of these are involved in mRNA metabolic processes, while 21 are involved in catabolic mRNA



processes (enriched functional pathways identified in these analyses are shown in **Figure 6.4**). It is possible that some of these genes could be involved in the regulation of the stability of the L1 transcript. LA related protein 1 (LARP1), for example, which is one of the most significantly enriched proteins in the sample, is known to regulated the stability and translation of mRNAs by binding the cap of the mRNA and preventing the binding of the translation initiation factor eIF4F (Lahr *et al.* 2017). Another enriched protein, PUM1, acts as a post transcriptional repressor by binding the 3' UTR or mRNA targets (Liu *et al.* 2018).

As expected, several of these proteins were also identified as L1 interactors in MS analyses using the parental cell line, which indicates that they are unlikely to be associated with the unusual retrotransposition associated with FANCA deficient cells. Consistently, previous proteomic studies in human cells revealed how the cellular background can influence the number and nature of L1-ORF1p interactors (Goodier *et al.* 2007; Mita *et al.* 2020).



**Figure 6.4: GO term analysis of functional pathways enriched in L1-ORF1p interactors identified in FANCA cells.** A GO term analysis of enriched pathways for L1-ORF1p interactors was performed using the online tool ShinyGO. *Cricetulus griseus* (Chinese hamster) was used as the reference species.

### 6.3 A comparative analysis of host factors previously identified in L1 proteomic studies.

As previously mentioned, a large number of the factors identified in this study had been previously reported as potential L1 interactors. **Table 6.2** highlights the number of factors which had been previously identified in the experiments described in this chapter. Table 5.3 lists which previous study identified which host factors

**Table 6.2 Previously identified L1 interactors from the current study.**

Number of VH4 interactors previously identified	36 (59%)
Number of V79 interactors previously identified	40 (60%)
Number of common interactors previously identified	24 (68%)

**Table 6.3 Overlapping L1 interactors listed by study**

Taylor <i>et al.</i> (2013)		Goodier <i>et al.</i> (2013)	Liu <i>et al.</i> (2018)	Taylor <i>et al.</i> (2018)	Mita <i>et al.</i> (2021)	Found in FANCA mutant cells only
CALML5	EEF2	DDX5	UPF1	PABPC1	NCOAS	Found in parental cells only
OTUD4	GAPDH	YBX1	PABPC1	UPF1		Found in both cell lines
DDX5	YBX1	HNRNPA1	PABPC4	PABPC4		
CFL1	RPL8	DDX17	SMARCC1			
ZC3H11A	KHSRP	LARP1	DNMT3A			
PABPC1	NONO	SRSF6				
RPL35	DDX3X	HNRNPK				
RPL19	RPS11	FAM120A				
RPL18	PSMD2	SF3B1				
RPS27	TARDBP	SNRNP70				
HNRNPA1	KHDRBS1	SERBP1				
HNRNPM	TCP1					
NCOA5	PABPC4					
LARP1	RPS21					
SF3B2	RPS27L					
RPL15	FRX1					
ATAD3A	BAIAP2					
SF3B1	RPL21					
RBM14	SNRNP70					
RPL6	SERBP1					

RPL9P9	HSPA9				
SMARCC1	ATAD3B				
FRX1	HNRNPK				
RPL29					

In contrast, I also identified proteins of high biological interest which were uniquely enriched in the FANCA mutant (VH4) sample, including YTHDF2 (**Figure 6.2 and Table 6.1**). Indeed, YTHDF2 is functionally involved in mRNA metabolism and catabolism (Du *et al.* 2016), together with DDX5; DDX5 is an RNA helicase found in P-bodies and stress granules, and is thought to be involved in the degradation and suppression of mRNAs (Xing *et al.* 2019). Consistently, there is ample literature demonstrating a clear connection between P-bodies and the metabolism of L1-RNPs (Goodier *et al.* 2007; Doucet *et al.* 2010).

YTHDF2 is involved in a number of biological processes but primarily, it is concerned with the reading of the N6-methyladenosine (m6A) modification (Wang and Lu 2021). m6A is a reversible post-translational modification which regulates the degradation of RNA transcripts. Of the >150 known modifications which take place on cellular RNAs, m6A is thought to be the most prevalent (Wang and Lu 2021) and is a dynamic and reversible modification.

YTHDF2 has been previously found to be involved in the regulation of R-loops (i.e., genomic RNA: DNA hybrids); additionally, and consistently, YTHDF2 has been found to associate with R-loops *in vitro*. Remarkably, increasing experimental evidence has demonstrated that R-loops are over accumulated in cells exhibiting a FA phenotype (Okamoto *et al.* 2019). Thus, I therefore decided to further explore the interaction between YTHDF2 and L1-ORF1p, with the ultimate goal of testing whether R-loops might represent potential sites for L1 retrotransposition in FA cells. However, and prior to further investigate the role of YTHDF2 and R-loops in the regulation of L1 retrotransposition, I decided to analyse whether YTHDF2 and R-loops are relevant for other DNA repair cells (i.e., NHEJ cells, which also support ENI-retrotransposition, see next section).

## 6.4 L1-ORF1p proteomic studies in NHEJ-deficient cells: identification of host interactors with L1-ORF1p.

To investigate the mechanism of L1 deregulation that I uncovered in FA deficient cells, I used proteomics to investigate the collection of host factors interacting with L1-RNPs in FANCA mutant cells, using parental cells as controls. I reasoned that by exploring host factors associated with L1-RNPs in FA deficient cells, and because we used parental cells as a control, I could use proteomics to identify host factors that specifically associate with L1-RNPs in the context of this DNA repair cellular deficiency, ultimately allowing me to investigate the mechanism of L1 deregulation in FA patients. In doing that, I uncovered host factors that have never been associated with L1-ORF1p in previous proteomic studies (around 10 studies published prior to the writing of my Thesis (Taylor *et al.* 2013; Moldovan and Moran 2015; Mita *et al.* 2018; Taylor *et al.* 2018; Ardeljan *et al.* 2019; Vuong *et al.* 2019; Briggs *et al.* 2021). A limitation of the above approach is that part of the L1 deregulation phenotypes described in FA deficient cells has also been reported for cells containing mutations in a different DNA repair pathway: that is ENi-retrotransposition, reported in NHEJ-mutant and FA mutant cells (Morrish *et al.* 2002; Morrish *et al.* 2007). Thus, and although the L1 phenotypes associated with FANCA mutant cells goes beyond ENi-retrotransposition, I wanted to identify whether the proteins of interest I identified in FA cells would also be identified as L1-ORF1p interactors in a different cellular background, which is also a DNA repair cellular deficiency.

To do this, I used the same optimised immunoprecipitation/MS protocol using T7-tagged L1-ORF1p (i.e., pAD2TE1 vectors) in hamster XR-1 cells, NHEJ mutant cells carrying mutations in XRCC4/DNA Lig IV, and using their parental cell line, 4364, as a control (Stamato *et al.* 1983; Li *et al.* 1995). At difference with hamster FANCA deficient cells (VH4), XR-1 is a hamster cell line deficient in NHEJ DNA repair that I previously demonstrated that it does not support retrotransposition of L1-ORF1p or L1-PIP mutant elements. Thus, seeing which interactors are common between XR-1 and VH4 cells would allow to highlight L1-ORF1p interactors which might be associated with deficiency in DNA repair but not specifically in a FA background.

**Table 6.4** List of significantly enriched ORF1p-T7 interactors in NHEJ deficient (XR-1) and parental (4634) cells

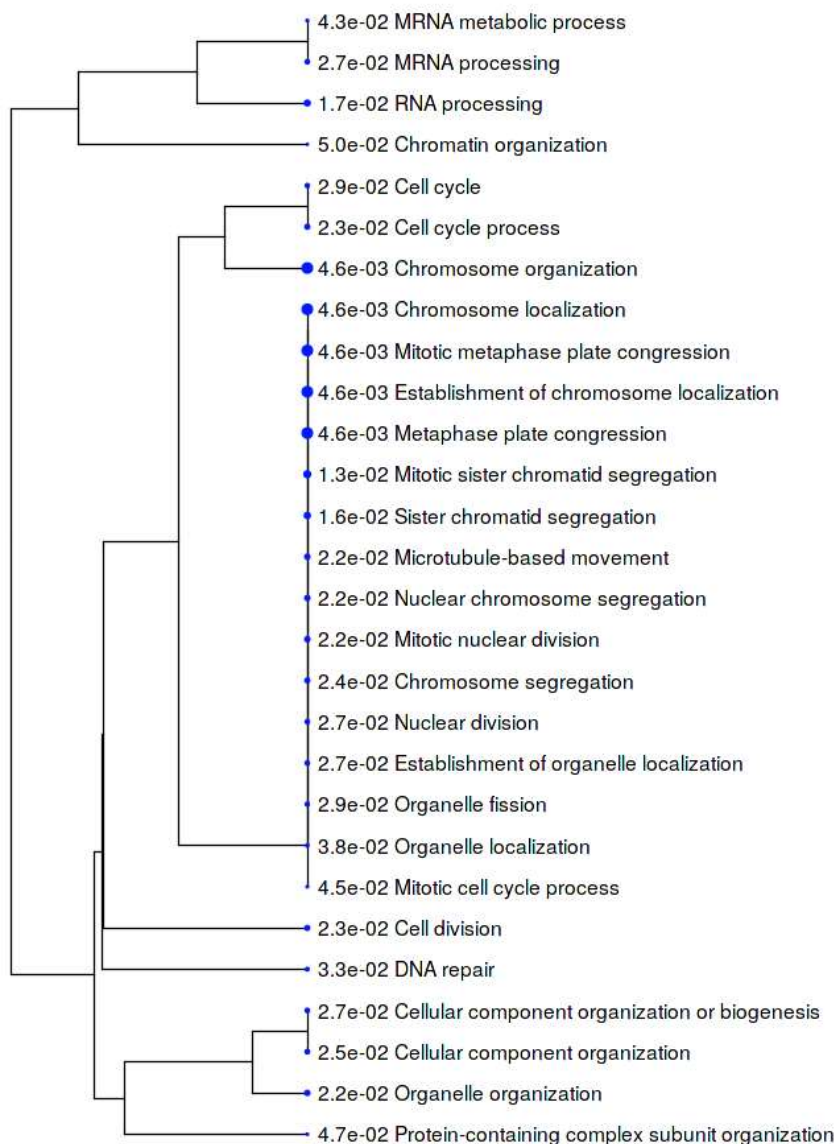
ORF1p-T7 interactors				
Parental (4364)	Previously Identified	NHEJ deficient (XR-1)	Previously Identified	Common
SFPQ		LRRC47		RPL13
RPS3		NCBP3		SMARCC2
AHCY		UBR4		RPS11
SMARCB1	V79 screen ( <b>Table6.1</b> )	CFL1	VH4 screen ( <b>Table6.1</b> )	TSR1
SMARCC2	V79 screen ( <b>Table6.1</b> )	SMARCC2	V79 screen ( <b>Table6.1</b> )	
EIF4B	Goodier et al 2013	CDC42BPA		
POLR1A		ALDH2		
MCM3	Liu et al 2018	POLR1A		
RPS12	Taylor et al 2013	DDX23	Goodier et al 2013	
RPS11	V79 screen ( <b>Table6.1</b> ) Taylor et al 2013	RPS11	V79 screen ( <b>Table6.1</b> ) Taylor et al 2013	
DNAJB1		KIF2C		
ATP5F1A		FUS		
PSMA1		GKAP1		
PSMA3		LDHA		
MSN		HSP90AB4P		
TSR1	V79 screen ( <b>Table6.1</b> ) Taylor et al 2013	TSR1	V79 screen ( <b>Table6.1</b> ) Taylor et al 2013	
DDX6	Taylor et al 2018, Taylor et al 2013	SF3B1	Taylor et al 2013, Goodier et al 2013	
DNMT1	Liu et al 2018	SRP72	Liu et al 2018	
U2AF2	VH4 screen ( <b>Table6.1</b> )	RBBP7		
RPL13	Taylor et al 2013	LARP7	Taylor et al 2018	
TLN1		HP1BP3		
RBM19		PABPC4	Liu et al 2018	
MYO5A		RPL19	V79 screen ( <b>Table6.1</b> )	
CRYBG1		TUBA1B	VH4 screen ( <b>Table6.1</b> )	
RPL19	V79 screen ( <b>Table6.1</b> ) Taylor et al 2013	RPL13	Taylor et al 2013	
RPL35	V79 screen ( <b>Table6.1</b> ) Taylor et al 2013			
RPL18	V79 screen ( <b>Table6.1</b> ) Taylor et al 2013			
RPL39P5	V79 screen ( <b>Table6.1</b> )			

As above, I used 3 biological replicas from XR-1 and 4364 cells in MS experiments, for T7 and isotypic IgG IPs; next, I used the same bioinformatic pipeline to analyse L1-ORF1p interactors in NHEJ mutant cells (and parental cells as controls). In fact, I found 25 and 28 host-factors that specifically interact with L1-RNPs in XR-1 and 4364 cells, respectively. Of the 25 proteins found in the NHEJ deficient cells, 8 had been identified in previously published screens (**Table 6.4**) (Taylor *et al.* 2013; Goodier *et al.* 2013; Liu *et al.* 2018; Taylor *et al.*, 2018). There were 8 proteins in the list which were previously identified by my previous screen in V79 cells (the parental line of VH4; **Table 6.1**).

Of the 28 host factors proteins found in the parental line, 11 had been identified in previously published screens (**Table 6.4**) Taylor *et al.* 2013; Liu *et al.* 2018; Taylor *et al.*, 2018). There were 4 proteins which were previously identified in my previous screen in V79 cells (**Table 6.1**).

Notably, the vast majority of interactors identified in the FANCA (VH4) screen were not identified in NHEJ deficient (XR-1) cells and vice versa, suggesting that ORF1p is likely interacting with a slightly different set of proteins in these two cell lines. By contrast, the screens in NHEJ cells and the parental line had a number of shared hits with V79, the parental CHO line of VH4. A GO term analysis for enriched functional pathways in NHEJ deficient cells (**Figure 6.5**) also shows distinct processes from those found in L1-ORF1p interactors in FANCA cells (**Figure 6.4**). Importantly, this variation in results also highlight the importance of the cellular milieu in MS experiments, demonstrating how different cell lines can generate different impressions. YTHDF2, a protein of interest identified in FA cells, is also absent here, indicating that it could be FA specific interactor of the L1-RNP. Thus, I decided to continue investigating YTHDF2 with regard to L1 retrotransposition in FA cells.





**Figure 6.5: GO term analysis of functional pathways enriched in L1-ORF1p interactors identified in NHEJ mutant cells.** A GO term analysis of enriched pathways for L1-ORF1p interactors was performed using the online tool ShinyGO. *Cricetulus griseus* (Chinese hamster) was used as the reference species.

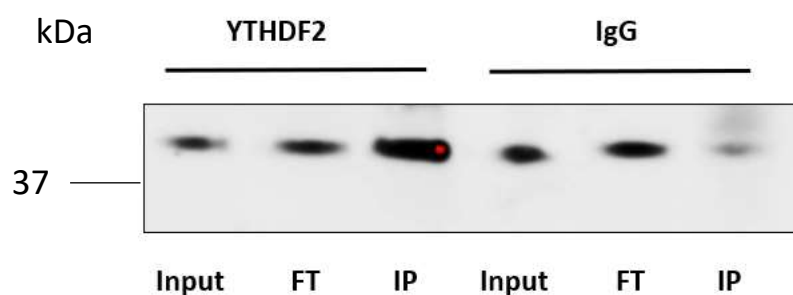
## 6.5 The role of YTHDF2 in the regulation of L1 retrotransposition in FA deficient cells.

Following the MS analyses summarised earlier in this chapter, I decided to further investigate YTH N6-Methyladenosine RNA Binding Protein 2 (YTHDF2) as a protein of potential interest. I speculate that YTHDF2 could be associated with the L1 deregulation phenotype reported in FANCA mutant cells. Indeed, the interaction between L1-ORF1p (i.e., L1-RNPs) and YTHDF2 was significantly enriched specifically in FANCA deficient cells and not in the parental cell line or in other DNA repair deficient cell lines (NHEJ mutant cells).

YTHDF2 works together with YTHDF1 and YTHDF3 to destabilise m6A containing RNAs through de-adenylation (Shi *et al.* 2017). Though the three proteins have been observed to work together to degrade the same transcripts, YTHDF2 is believed to be the primarily regulatory protein. In fact, de-adenylation by YTHDF2 is one of two main pathways of mRNA destabilisation (Wang and Lu 2021). In some cases, YTHDF2 is also believed to perform endoribonucleolytic cleavage of RNAs via a complex it forms with HRSP12 and ribonuclease (RNase) P/mitochondrial RNA-processing (MRP; endoribonuclease) complex (Park *et al.* 2019). Notably, YTHDF2 has also been found to be deregulated in many cancers, including hepatocellular carcinoma (HCC), pancreatic cancer, and acute myeloid leukaemia (Dai *et al.* 2021).

## 6.6 Testing for an interaction between YTHDF2 and ORF1p

Following the detection of YTHDF2 as a L1-ORF1p interactor by mass spectrometry, specifically in FANCA deficient cells, I next wanted to validate their interaction using co-immunoprecipitation. Briefly, following the methodology/approach laid out in **Figure 6.1**, I transfected HEK293T cells with plasmid pAD2TE1-101/L1.3, which contains an active human L1 element in which L1-ORF1p has been tagged with a T7 epitope tag. Next, using an antibody to YTHDF2 or a control antibody of the same isotype, I performed a pull down with magnetic protein G beads. Remarkably, western-blotting using the anti-T7 antibody confirmed interaction of L1-ORF1p with YTHDF2 (**Figure 6.10**, left side, IP lane); as expected, L1-ORF1p was not pull down when using the isotypic Ig G control, although we found some weak interaction with either Ig G or the magnetic protein G beads used in the IP (**Figure 10**, right side, IP lane). Whether the interaction between L1-RNPs and YTHDF2 involve the L1 RNA and/or L1-ORF1p remains to be determined, but these results corroborate the MS findings previously described in this chapter. It is worth mentioning that the interaction between L1-ORF1p and YTHDF2 was found only in FANCA deficient cells, although we could corroborate such interaction using HEK293T. To the best of my knowledge, this is the first known finding of an interaction between YTHDF2 and L1-ORF1p.



**Figure 6.6 Co-immunoprecipitation of YTHDF2 and L1-ORF1p.** A co-immunoprecipitation was performed in HEK293T cells transfected with plasmid pAD2TE1-101/L1.3, using an antibody to YTHDF2 or isotypic IgG as a control. After IP/washes, samples were resolved by SDS-PAGE, transferred and blotted for L1-ORF1p using the anti T7 antibody. The Input (inp), flow through (FT) and Immunoprecipitated (IP) samples are shown from left to right for each antibody. This experiment was performed three times

## 6.7 Investigating LINE-1 retrotransposition in YTHDF2 KO cells

After confirming the interaction between YTHDF2 and L1-ORF1p, I decided to test the role of this protein in regulating L1 retrotransposition. Interestingly, the m6A modification has also recently been found to regulate the stability of DNA: RNA hybrids (Abakir *et al.* 2020). Abakir and colleagues demonstrated, though immunostaining with an anti-m6A antibody, that this mark was present on the RNA components of RNA: DNA hybrids in human pluripotent stem cells (hPSCs); in fact, they were able to identify YTHDF2 on the majority of RNA: DNA hybrids found in these cells. Notably, some of these hybrids correspond with R-loop structures, which accumulate to a high degree in FA deficient cells, and it is therefore of interest that YTHDF2 was observed to interact directly with these hybrids through immunostaining experiments. Consistently, depletion of YTHDF2 leads to an accumulation of R loops in hPSCs. In light of these data, identifying YTHDF2 as an L1-ORF1p interactor uniquely in FANCA deficient cells is an interesting finding. Thus, and to further investigate the role of YTHDF2 and R-loops in regulating retroelement activity, I decided to test LINE-1 retrotransposition in YTHDF2 KO cells, a generous gift of Prof. Alexander Ruzov (Nottingham University, UK).

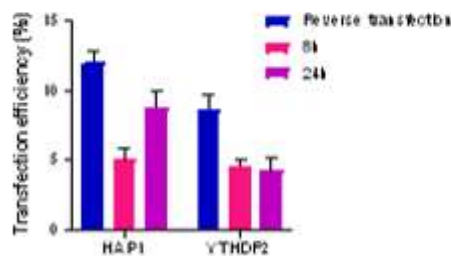
In the recent YTHDF2 study described above, Abakir and colleagues generated a KO cell line using CRISPR/Cas9 genome editing tools and HAP-1 cells (Abakir *et al.* 2020). Briefly, HAP1 cells were derived from a patient suffering chronic myelogenous leukaemia (CML) and are widely used in biological studies, in particular to generate KO models, because these cells have a near-haploid genome (except for chromosomes 8 and 15 (Kotecki *et al.* 1999)). Thus, I used the parental HAP1 line and a derivative

KO subline to investigate the role of YTHDF2 on L1 retrotransposition (Abakir *et al.* 2020). Because L1 retrotransposition in HAP1 cells has not been reported, I first established experimental conditions that would allow me to investigate L1 retrotransposition in these cells and in the YTHDF2 KO subline. Notably, I was able to recapitulate human L1 retrotransposition in parental HAP1 cells and in the YTHDF2 KO subline (**Figure 6.7** and see below). However, preliminary experiments in HAP1 and YTHDF2 KO cells revealed some difficulty in generating sufficient transfection efficiency of plasmids that would allow me to robustly measure L1 retrotransposition. While the parental line generated transfection efficiencies ranging from 12-15% and CFU ranging from 40-60 at a seeding density of 40,000 cells, the number of colonies in the YTHDF2 KO cell line were much lower (transfection efficiency initially ranged from 1-3% and the CFU from a wild type L1 element was 4-5, **Figure 6.7**). In order to optimise the transfection of these cells, I therefore tried a number of different conditions of transfection, including: i) testing three different lipid-based transfection reagents (FuGene 6, FuGene HD and Lipofectamine 2000); ii) different concentrations of these reagents and plated cells; and, iii) different methods of transfection (reverse transfection vs transfecting 16-18 hours after plating). The main results from the optimization experiments are shown in **Figure 6.7** where I was able to increase the transfection efficiency of parental HAP1 and YTHDF2 KO cells to approximately 10%, using reverse transfection and Lipofectamine (**Figure 6.7**).

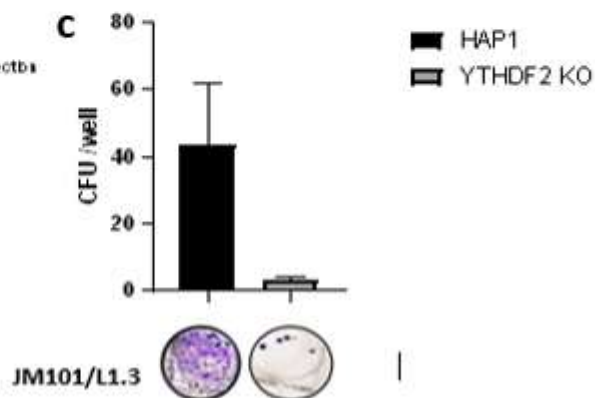
**a**

	Concentration (ul/well)	Avg Transfection efficiency (%)	
		HAP1	YTHDF2
Lipofectamine	2.5	10	7.0
	5	14	9.5
FuGene 6	3	5.1	1.7
	6	8.1	3.2
FuGene HD	3	7.9	2.8
	6	9.1	3.2

**b**

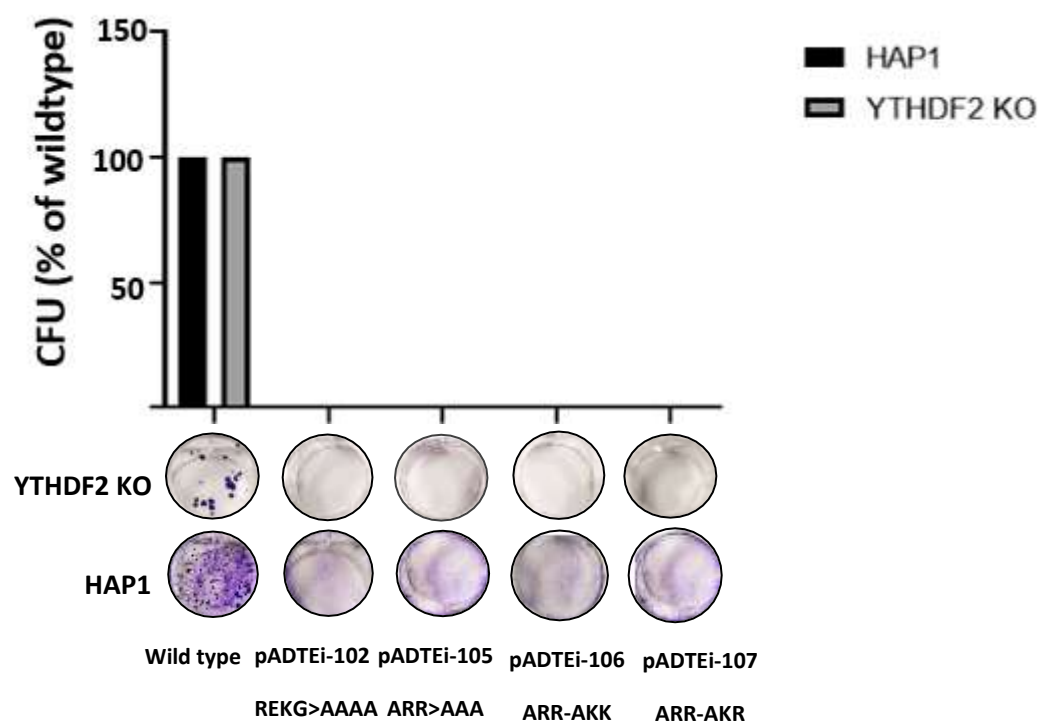


**c**



**Figure 6.7. Optimisation of transfection efficiency in parental HAP 1 and YTHDF2 KO cells. (A)** Initial experiments showed that the retrotransposition frequency of L1 in YTHDF2 KO cells was substantially less than in the parental line (HAP1). **(B)** Transfection efficiency in HAP1 and YTHDF2 KO cells after transfection with either Lipofectamine, FuGene 6 or FuGene HD at two different concentrations. **(C)** Transfection efficiency in HAP1 and YTHDF2 KO cells using Lipofectamine 2000 and reverse transfection, conventional transfection 8h after plating or after 8h or conventional transfection 24h after plating.

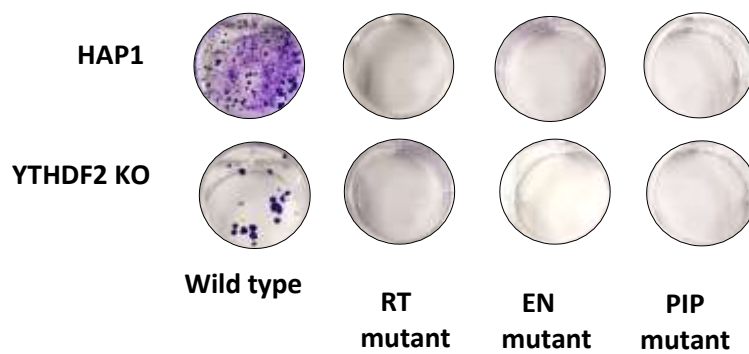
After optimising transfection to around 9-10%, I next tested the series of L1-ORF1p mutants (i.e., pAD2TE1 series vectors) in both lines, parental HAP1 and YTHDF2 KO cells. While WT-L1 retrotransposition was readily detected in parental HAP1 cells, and at a much lower level in YTHDF2 KO cells (**Figure 6.8**), no retrotransposition activity was found for any of the L1-ORF1p mutants tested (**Figure 6.8**). An important limitation in these experiments is the sensitivity of L1 retrotransposition assays in YTHDF2 KO cells, with very low rates of retrotransposition even for the WT-L1, prohibiting me from exploring whether some of the L1-ORF1p mutants might retrotranspose at a low level in YTHDF2 KO cells (**Figure 6.8**). In the future, further optimisation of transfection and of the L1 assay in YTHDF2 KO cells would increase the chance of detecting rare retrotransposition events.



**Figure 6.8: Retrotransposition of L1-ORF1p mutants in HAP1 and YTHDF2 KO cells.** Shown are representative retrotransposition results of four L1-ORF1p mutant elements and of a WT-L1 in HAP1 cells (top) and YTHDF2 KO cells (bottom). The graph shows the quantification of retrotransposition assays. The number of colony forming units (CFU) per well was quantified and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three independent times.

Although the sensitivity of the L1 assay might not be high enough in HAP1 and YTHDF2 KO cells, I next tested L1-ORF1p mutants in assay. As described before, I used pAD2TE1 vectors containing inactivating mutations in the EN, RT and PIP domains of L1-ORF1p, together with the WT-L1. Again, no retrotransposition was detected for any of the L1-ORF1p mutants tested, possibly due to the limited sensitivity of the assay in HAP1 and YTHDF2 KO cells (**Figure 6.7**).





**Figure 6.9: Retrotransposition of L1-ORF1p mutants in HAP1 and YTHDF2 KO cells.** Shown are representative retrotransposition results of three L1-ORF1p mutant elements and of a WT-L1 in HAP1 cells (top) and YTHDF2 KO cells (bottom). The graph shows the quantification of retrotransposition assays. The number of colony forming units (CFU) per well was quantified for each element as a percentage of WT L1 activity and an average was taken over three technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three independent times.

In sum, while I was able to confirm the interaction between L1-ORF1p and YTHDF2 in cultured cells, the retrotransposition rates observed in HAP1 and YTHDF2 KO cells are too low to allow comparing the frequency of mobilization of several L1 mutant elements.

## 6.8 R-loops, FA deficient cells and deregulated L1 retrotransposition

In recent years, the FA phenotype has increasingly become associated with the accumulation of R loops (Okamoto *et al.* 2019). R loops are a structure formed from a RNA: DNA hybrid and a ssDNA, in genomic DNA. These structures form naturally during mitochondrial replication and during transcription, when a strand of RNA invades a DNA duplex, displacing one of the strands (reviewed in García-Muse and Aguilera 2016). Unlike the RNA: DNA duplexes which form during transcription, R-loops can reach lengths of up to 500 bp, disrupting chromatin organisation and generating genetic instability (Duquette *et al.* 2004). The FA pathway is believed to be one of several mechanisms evolved to process R-loops effectively, reducing the risk of DNA damage (Garcia-Rubio *et al.* 2015). Consistently, deleterious R-loops often form at fragile sites of large genes (Barlow *et al.* 2013) and the FA pathway is believed to play a key role in resolving these. Notably, a study by Garcia-Rubio *et al.* (2015) found that R-loops accumulate in cells depleted of FANCA or FANCD2. This finding was reported both in primary patient cell lines and in murine cells. Furthermore, DNA damage was found to be mediated by R-loops in FANCD2 depleted HeLa cells.

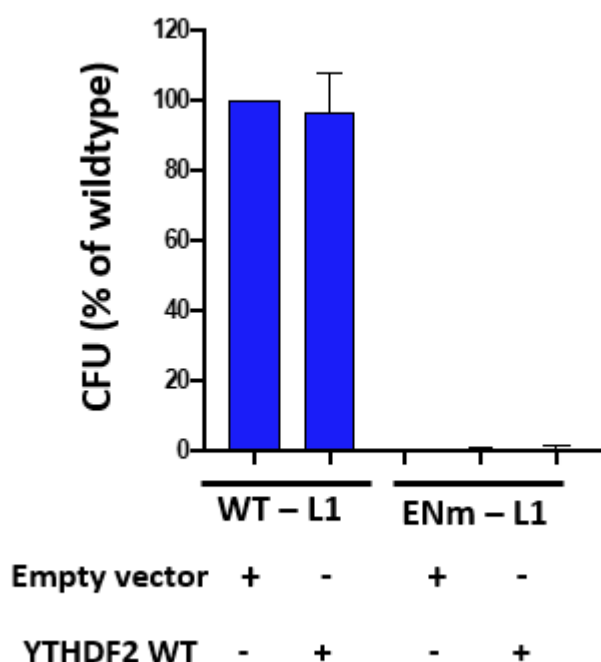
The relationship between R-loops and LINE-1 retrotransposition is currently unclear. On one hand, LINE-1 elements have been found to be R-loop “hotspots”, indicating that R-loops could be involved with LINE-1 transcriptional repression (Yan *et al.* 2020). Additionally, given the reported accumulation of R-loops in FA cells, it is possible that LINE-1 elements could exploit this type of lesion to initiate integration, in the same way as SSBs/DSBs are presumably used by EN-mutant L1s to integrate in NHEJ mutant cells (i.e., ENi-retrotransposition). In fact, if LINE-1 elements were able to exploit R-loops as integration substrates, or to functionally interact with these structures *in vivo*, this might explain the L1 deregulation phenotype reported in some FA deficient cells. Thus, a functional interaction between L1-RNPs and R-loops might explain why some FA cells support increased LINE-1 retrotransposition and potentially why FANCA cells support L1-ORF1p-independent, EN-independent and PIP-independent retrotransposition. I therefore wanted to investigate a functional interaction between R-loops and LINE-1 retrotransposition.

Notably, I was able to confirm an interaction between L1-ORF1p and YTHDF2 in cultured cells; however, due to technical limitations, I was unable to use YTHDF2 KO cells to further support a functional interaction between YTHDF2, R-loops and L1 retroelements, at least using HAP1 cells. Furthermore, it is worth noting that the interaction between L1-ORF1p and YTHDF2 was only found in FANCA deficient cells, suggesting that the impact of YTHDF2 and R-loops might only occur in FA mutant cells. Because of these limitations, I decided to use a reciprocal strategy to explore a functional interaction between YTHDF2, R-loops and L1 retroelements. Briefly, as described by Abakir and colleagues, YTHDF2 KO cells accumulate R-loops in genomic DNA, leading to genomic instability (Abakir *et al.* 2020). Thus, I reasoned that overexpression of YTHDF2 would reduce R-loop levels,



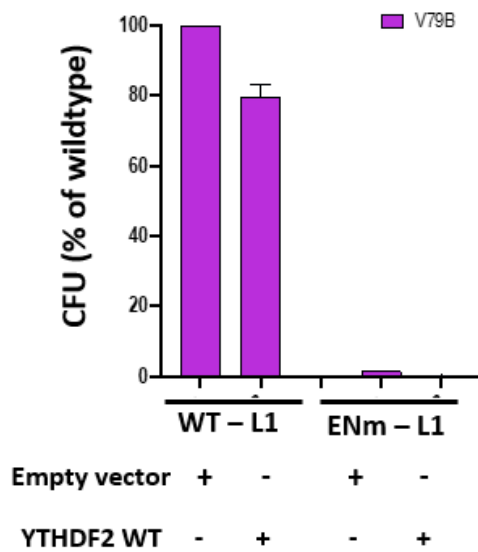
specifically in FANCA deficient cells, allowing me to explore whether reducing R-loops could rescue the L1 phenotype reported in these cells (i.e., L1-ORF1p-independent, EN-independent and PIP-independent retrotransposition).

To test whether YTHDF2 overexpression interfere with L1 retrotransposition, I used an engineered L1 retrotransposition assay that is based on the activation of the blasticidin S deaminase gene after a round of retrotransposition (Morrish *et al.* 2002) (i.e., L1s tagged with the *mblastl* retrotransposition indicator cassette; pJJ101 vector series). Consistently, using *mblastl*-tagged L1s, the number of blasticidin (blast) resistant foci can be used to quantify retrotransposition in cultured cells. Remarkably, because of the kinetic of blast selection, our lab previously demonstrated that the L1 *mblastl*-based assay is ideal to score retrotransposition in the first 72h after transfection (Tristan Ramos 2020). Indeed, because overexpression of proteins from exogenous plasmids typically peaks 48-72h after transfection, I reasoned that *mblastl*-tagged L1 vectors are the best and more robust tool to explore the effect of YTHDF2 overexpression on L1 retrotransposition. Additionally, among the L1 mutant elements tested in FA cells in this Thesis, I decided to use the EN-mutant L1 to explore the impact of YTHDF2 overexpression on L1 deregulation, simply because the EN-mutant L1 retrotransposed at the highest level in FANCA deficient cells. Initially, I decided to use HeLa cells to explore the impact of YTHDF2 overexpression on L1 retrotransposition (**Figure 6.10**); however, I found similar levels of WT-L1 retrotransposition in HeLa cells co-transfected with a plasmid overexpressing HA-tagged YTHDF2 (N-terminus) or with the empty vector (pCMV-HA, **Figure 6.10**). Consistent with its mutant nature, no significant retrotransposition was found with the EN-mutant L1 in any of the conditions tested (**Figure 6.10**).



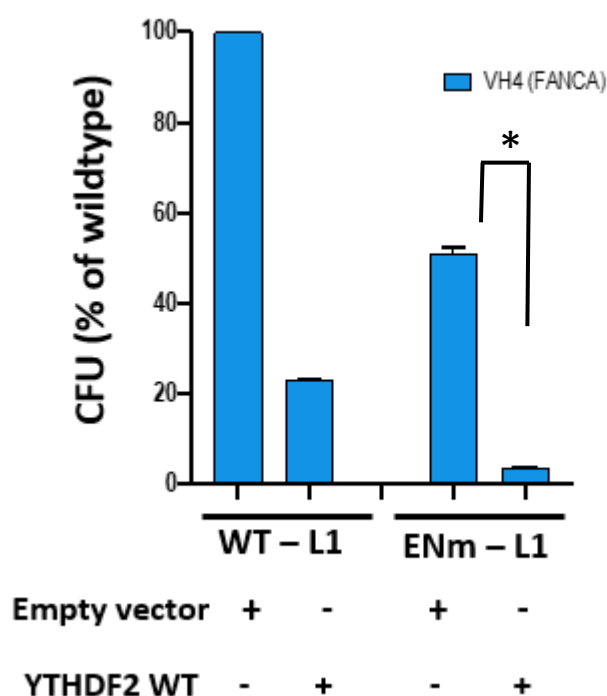
**Figure 6.10: Overexpression of YTHDF2 and L1 Retrotransposition in HeLa cells.** Shown are representative retrotransposition results in HeLa cells after transfecting a WT-L1 (left side) or of an EN-mutant L1 (D205A, right side) with a plasmid overexpressing HA-tagged YTHDF2, or with the empty vector as a negative control (pCMV-HA, indicated in the top). The graph shows the quantification of retrotransposition assays. The number of colony forming units (CFU) per well was quantified and an average was taken over four technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three independent times.

However, because the interaction of L1-RNPs and YTHDF2 was only found in FANCA deficient cells, I decided to test the effect of YTHDF2 overexpression in FANCA deficient (VH4) and parental cells (V79B), using the experimental approach used in HeLa cells. Consistent with the results observed in HeLa cells, I found the same trend in parental V79B hamster cells (**Figure 6.11**); in fact, I found no significant differences in the rate of WT-L1 retrotransposition when parental cells were co-transfected with HA-tagged YTHDF2 or with the empty vector control (**Figure 6.11**). Further consistent with data reported in HeLa cells, the EN-mutant L1 retrotransposed at background levels in all conditions tested (**Figure 6.11**).



**Figure 6.11: Overexpression of YTHDF2 and L1 Retrotransposition in V79B cells.** Shown are representative retrotransposition results in V79B cells after transfecting a WT-L1 (left side) or of an EN-mutant L1 (D205A, right side) with a plasmid overexpressing HA-tagged YTHDF2, or with the empty vector as a negative control (pCMV-HA, indicated in the top). The graph shows the quantification of retrotransposition assays. The number of colony forming units (CFU) per well was quantified and an average was taken over four technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three independent times.

Finally, I tested the effect of YTHDF2 overexpression on WT-L1 and EN-mutant L1 retrotransposition in FANCA deficient cells (**Figure 6.12**). Remarkably, at difference with HeLa and V79B cells, I found that YTHDF2 overexpression in V79B cells lead to significant lower retrotransposition levels of a WT-L1 and of an EN-mutant L1 (D205A, (**Figure 6.12**)). Indeed, YTHDF2 overexpression in FANCA deficient cells reduced WT-L1 and EN-mutant L1 retrotransposition 5 and 10-fold, respectively, when compared with empty vector co-transfected cells. Altogether, the above data suggest that WT-L1 and ENi-L1 retrotransposition is significantly reduced when cellular R-loop levels are reduced in FANCA deficient cells. Furthermore, this data is consistent with the retrotransposition data observed in YTHDF2 KO HAP1 cells, despite any technical limitation. In fact, I speculate that the regulatory role of YTHDF2 on L1 retrotransposition is specific for FANCA deficient cells, further suggesting that genomic R-loops act as a “dominant integration substrate” for L1-RNPs during retrotransposition, explaining why mutant L1s can retrotranspose in FANCA cells.

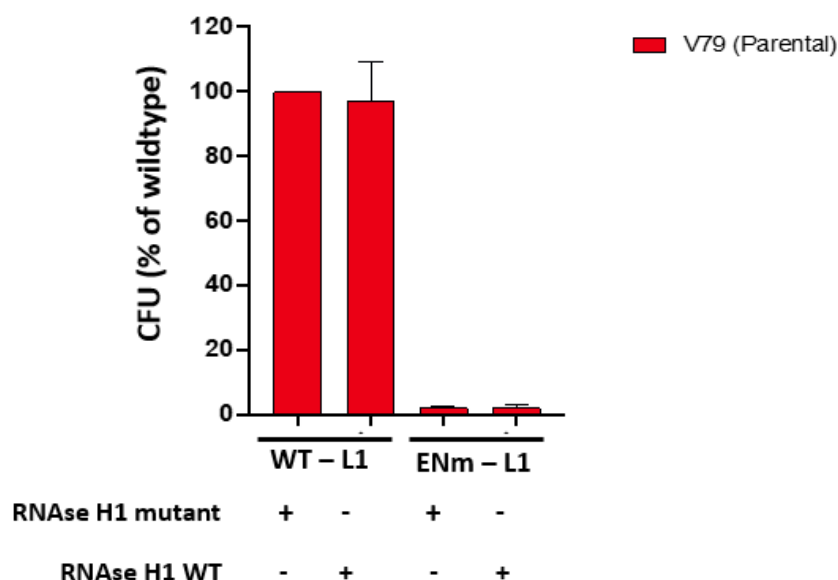


**Figure 6.12: Overexpression of YTHDF2 and L1 Retrotransposition in VH4 cells.** Shown are representative retrotransposition results in FANCA deficient cells (VH4) after transfecting a WT-L1 (left side) or of an EN-mutant L1 (D205A, right side) with a plasmid overexpressing HA-tagged YTHDF2, or with the empty vector as a negative control (pCMV-HA, indicated in the top). The graph shows the quantification of retrotransposition assays. The number of colony forming units (CFU) per well was quantified and an average was taken over four technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed three independent times.

Collectively, the above data suggest that R-loops could act as “dominant integration sites” in FANCA deficient cells. Alternatively, YTHDF2 could bind ORF1 in FANCA cells and sequester it into non-retrotransposition-competent complexes. To narrow down the potential explanations, I next explored whether reducing R-loops using an alternative strategy could also rescue the L1-deregulation phenotype of FANCA cells. To do that, I exploited *mblastl*-tagged L1s and the co-transfection/retrotransposition assay optimised above, to explore the effect of overexpressing RNase H1. In fact, cellular RNase H1 is a protein involved in the resolution of R-loops, and therefore depletion of RNase H1 activity results in increased R-loops (Okamoto *et al.* 2019; Garcia-Rubio *et al.* 2019). In these assays, I used vectors expressing the cDNA of the nuclear isoform of human RNaseH1 (hRNH1 in **Figure 6.13**) or a catalytic dead RnaseH1 allelic mutant (D119N missense mutation in the hRNaseH1 catalytic site), used as an internal negative control (hRNH1mutant in **Figure 6.13**). Additionally, I overexpress *E. coli* RNase H1 (*rnhA* gene) in cultured cells to reduce the levels of cellular R-loops, as reported (Okamoto *et al.* 2019; Garcia-Rubio *et al.* 2019). As above, I first tested the impact of RNase

H1 overexpression on L1 retrotransposition using human HeLa cells; notably, I didn't observe any impact on WT-L1 retrotransposition upon overexpressing nuclear human RNase H1. Unexpectedly, I found that overexpression of *E. coli* RNase H1 significantly reduced WT-L1 retrotransposition >10 fold when compared with cells co-transfected with an allelic inactive RNase H1 mutant (D70N missense mutation in the *rnhA* catalytic site). However, additional controls revealed that overexpression of *E. coli* RNase H1 (*rnhA* gene) significantly reduced the number of blast-resistant foci in HeLa cells co-transfected with the pcDNA6.1 vector, which constitutively express the blast deaminase gene and is routinely used to control toxicity/clonability issues in drug selection-based assays. Notably, no toxicity/changes in clonability were found in HeLa cells co-transfected with nuclear human RNaseH1, or with the inactive allelic control. Thus, we excluded vectors overexpressing *E. coli* RNase H1 in further experimentation, as the induced cellular toxicity difficult quantifying retrotransposition.

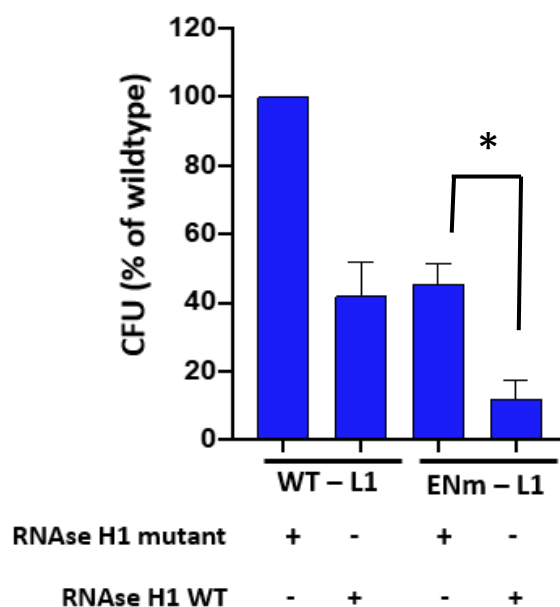
Next, I tested the effect of overexpressing human nuclear RNase H1 in parental V79B cells; notably, I observed the same trend detected in human HeLa cells, where WT-L1 retrotransposition levels were not affected by the overexpression of human RNase H1 (or the catalytic dead allele, see **Figure 6.13**). As expected, no significant retrotransposition was found for the EN-mutant L1 in any of the conditions tested (**Figure 6.13**).



**Figure 6.13: Overexpression of human nuclear RNaseH1 and L1 Retrotransposition in V79B cells.** Shown are representative retrotransposition results in parental cells (V79B) after transfecting a WT-L1 or of an EN-mutant L1 (D205A) with a plasmid overexpressing RnaseH1 (hR1), or a catalytic dead allelic vector as a

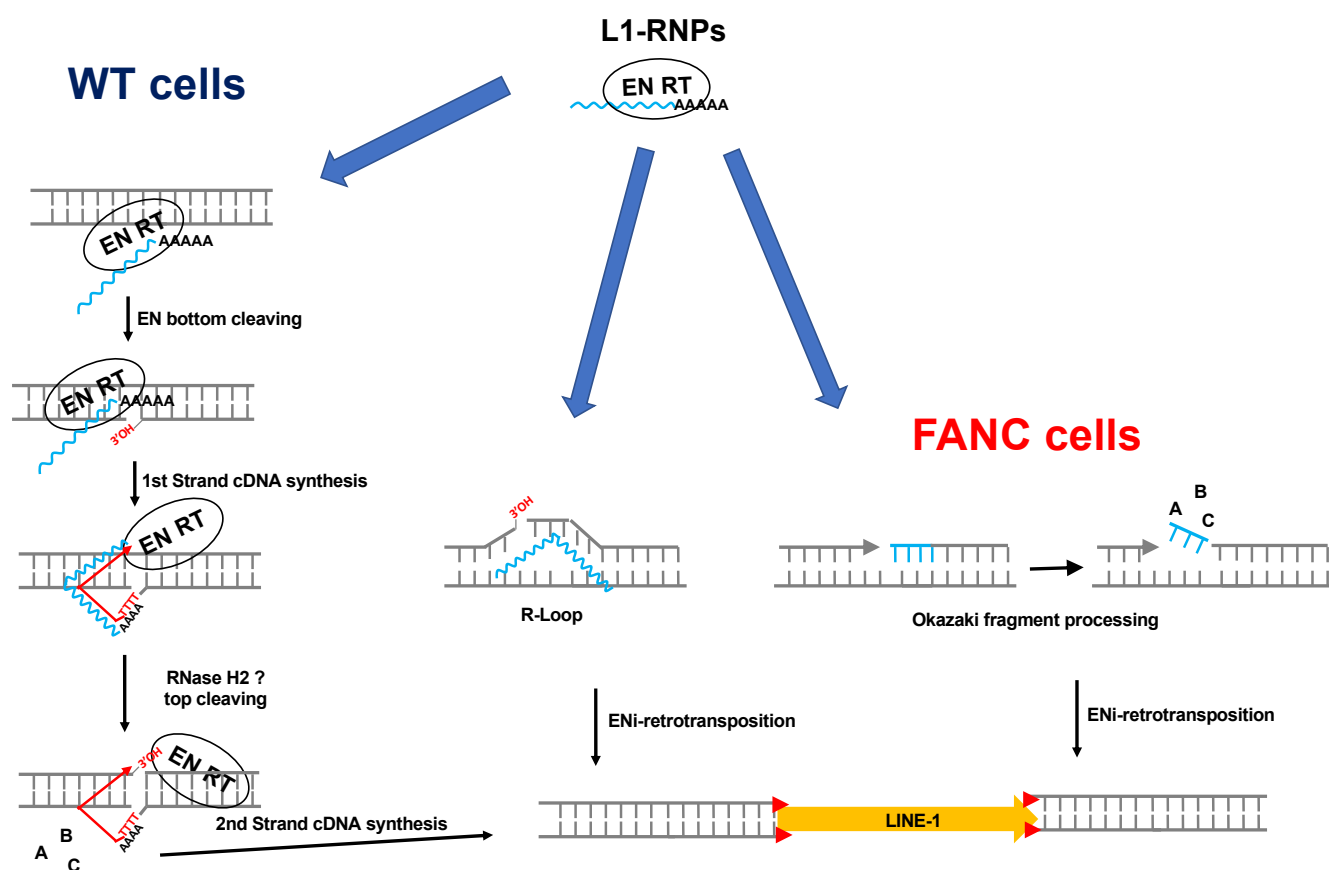
negative control (hR1mutant). The graph shows the quantification of retrotransposition assays. The number of colony forming units (CFU) per well was quantified and an average was taken over four technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed four independent times.

Remarkably, I found that overexpression of human nuclear RNase H1 significantly reduced the WT-L1 and ENi-L1 retrotransposition in FANCA deficient cells, at similar levels as those detected in YTHDF2 overexpression experiments (**Figure 6.14**).



**Figure 6.14: Overexpression of human nuclear RNaseH1 and L1 Retrotransposition in VH4 cells.** Shown are representative retrotransposition results in FANCA deficient cells (VH4) after transfecting a WT-L1 or of an EN-mutant L1 (D205A) with a plasmid overexpressing RNaseH1 (hR1), or a catalytic dead allelic vector as a negative control (hR1mutant). The graph shows the quantification of retrotransposition assays. The number of colony forming units (CFU) per well was quantified and an average was taken over four technical replicates. These values have been adjusted to control for differences in transfection efficiency and this experiment was performed four independent times.

In fact, I found that retrotransposition of the WT- and of the EN-mutant L1 was reduced 3 and 5-fold, respectively, when compared with cells transfected with the catalytic dead allelic construct (**Figure 6.14**). In sum, transfecting VH4 cells with a plasmid overexpressing RNase H1 led to decreased retrotransposition, indicating that L1 could somehow utilise unresolved R loops or replication complexes during Okazaki fragment processing, in order to mobilise (**Figure 6.15**).



**Figure 6.15: Model of L1 Retrotransposition in WT and FA deficient cells.** See text for additional details.

## 6.9 Discussion

In this chapter, I have investigated new aspects of retrotransposition in FANCA deficient cells. An MS analysis investigating the interactors of L1-ORF1p revealed several new proteins of interest which have not been previously identified in similar proteomic studies. I focused subsequent investigation on the n6 N6-Methyladenosine regulator YTHDF2, due to the connection between R-loops and the FA phenotype. Using co-immunoprecipitation, I successfully validated this interaction. However, further work will be required to test how dynamic the interaction between YTHDF2 and L1-ORF1p might be,

and whether L1 RNAs, or others, are also involved in this interaction. Similarly, because the functional interaction between L1-ORF1p and YTHDF2 might be relevant only in a FA background, further research is needed to unveil how YTHDF2 could regulate biological processes in a cellular milieu-dependent manner. Because the cellular background seems to be very relevant in the mechanism of YTHDF2 regulation of retrotransposition, I investigated the impact of YTHDF2 and R-loops in FANCA cells, using parental cells as a control. Indeed, doing that turned out to be quite important, as I found that the connection of L1 with R-loops and YTHDF2 only occur in FANCA deficient cells, as least for the regulation of ENi-retrotransposition.

Further investigation into the relationship between R-loops and L1 retrotransposition provides preliminary evidence to suggest that L1 elements could exploit these structures for mobilisation. An interplay between L1 elements and these structures could explain in part why L1 elements have different dynamics in some cells with a FA phenotype, which have accumulations of these lesions. That said, a recent paper showed that L1 elements could exploit stalled replication forks (Mita *et al.* 2020) which are also found at an increased degree in FA cells. Thus, there are likely several genetic features in these cells which could result in deregulated retrotransposition. However, preliminary experiments by colleagues at my lab revealed that stalled replication forks do not represent major insertion sites for L1, at least when arrest is induced with aphidicolin (unpublished data). Thus, I propose that investigating the relationship between replication forks and L1 would be an interesting area of research, complementary to this work.

From a mechanistic angle, several pieces of evidence in my Thesis strongly suggest that **L1 retrotransposition in FA deficient cells might predominantly occur at R-loops/Okazaki replication complexes** (see model in **Figure 6.15**). First, overexpression of YTHDF2, a reader of m6A modified RNAs enriched in R-loops that promote RNA degradation, can significantly reduce WT-L1 and ENi-L1 retrotransposition specifically in FANCA cells, an effect not observed in parental cells. Second, overexpression of human nuclear RNaseH1, which naturally resolve R-loops by degrading the RNA of hybrids, reduce WT-L1 and ENi-L1 retrotransposition specifically in FANCA cells, an effect that again was not observed in parental cells. Third, and unexpectedly, WT-L1 and ENi-L1 insertions are associated with the same atypical hallmarks of retrotransposition. In fact, the characterization of >100 *de novo* retrotransposition events in FA cells using WT- and EN-mutant L1s revealed that insertions are associated with atypical retrotransposition hallmarks (i.e., presence of long TSDs, deletions at the insertion site, 5' capture of palindromic DNA sequences, etc)(S. Morell unpublished). Unexpectedly, these atypical hallmarks were found for WT-L1 and ENi-L1 retrotransposition events, further



suggesting that DNA lesions accumulated in FA cells, that is R-loops, are dominant to dictate their mode/site of integration (see model in **Figure 6.15**)(S. Morell unpublished).

The first steps of the TPRT process led to the formation of an RNA: cDNA hybrid covalently attached to genomic DNA; a structure similar to an R-loop (see model in **Figure 6.15**). Similarly, after the completion of 1st cDNA synthesis, it is thought that RNaseH2 degrades the L1 RNA from the hybrid (Benitez-Guijarro *et al.* 2018), to allow TPRT to progress (i.e., second strand cDNA synthesis)(this structure functionally resembles the processing of Okazaki fragments during DNA replication (see model in **Figure 6.15**). With these considerations, I speculate that in FA deficient cells L1 retrotransposition intermediates (L1-RNPs) might start reverse transcription from exposed 3'OH in R-loops/replication processes, somehow resembling steps that occur during top strand cleaving and second strand cDNA synthesis during canonical TPRT retrotransposition (see model in **Figure 6.15**). In other words, during TPRT in parental/WT cells, the initiation of L1 reverse transcription requires cleaving of the bottom strand of genomic DNA; however, I speculate that the initiation of reverse transcription in FA cells mimics second strand cDNA synthesis, explaining why mutant L1 elements can retrotranspose at high levels in FA deficient cells, especially mutants in enzymatic activities required before second strand cDNA synthesis. Consistent with this hypothesis, the EN domain of L1-ORF1p is not strictly required to cleave top strand of gDNA to initiate second strand cDNA synthesis, as revealed by EN-mutant L1s being able to retrotranspose in NHEJ and FA cells. Further consistent with this hypothesis, the enzymatic activities associated with all L1 mutants active in FANCA cells [the main role of L1-ORF1p is to form a functional L1-RNP, the main role of the EN domain of L1-ORF1p is to cleave gDNA to initiate TPRT, and the main role of the PIP domain is to access gDNA/facilitate initiation of 1<sup>st</sup> strand cDNA synthesis], are all required/involved prior to the initiation of second strand cDNA synthesis (see model in **Figure 6.15**).

From an evolutionary angle, and consistent with the presumed later origin of the EN domain in L1-ORF1p (Malik *et al.* 1999), the mechanism of L1 integration observed in FA cells resembles an ancient mechanism of integration used prior to the acquisition of the EN domain, targeting replication complexes and R-loops structures (see model in **Figure 6.15**).



## Chapter 6



# Discussion



## 7.1 Summary of the findings in this thesis

The relationship between L1 retrotransposons and DNA repair factors in the human genome is currently an area of active research. There is increasing evidence that several DNA repair factors can serve as direct and indirect regulators of L1 retrotransposition, though the precise nature and context of this regulation has yet to be described. Furthermore, several studies have demonstrated that L1 elements harbouring different mutations can mobilise in cellular backgrounds deficient in DNA repair proteins. In cells deficient in NHEJ proteins, Morrish *et al.* (2007) demonstrated that L1 elements can mobilise independently of the endonuclease domain, raising further questions about the mechanism by which DNA repair proteins could potentially restrict L1 retrotransposition.

In this thesis, I tested a battery of L1 mutants in order to identify whether they were capable of mobilisation in different cellular contexts. These included a series of ORF1 mutants which had been previously characterised in HeLa cells. These mutants are of interest as they were observed either to reduce or abrogate retrotransposition in HeLa cells (Kulpa and Moran 2005). While the exact mechanism for this is not known, it is believed to be through inhibition of the association of ORF1p with the L1 transcript, which is necessary for formation of the ribonucleoprotein particle. A key finding of this thesis is that cells deficient in FANCA protein unexpectedly support retrotransposition of severe ORF1p mutants. Importantly, this effect was rescued in a FANCA complemented cell line. Subsequent testing of ORF1p only elements in FANCA revealed that despite the mobilisation of severe ORF1p mutants, some functions of ORF1 were still required. As soon as the ORF1p domain was removed entirely, retrotransposition of the element was completely abrogated.

I tested these mutants in a range of cell lines, including lines deficient in FANCC, NHEJ proteins, FANCI and FANCD. However, the retrotransposition of ORF1p mutants that I observed was thus far unique to FANCA deficient cells. I did observe, however, that both FANCC and FANCA deficiency appeared to increase the total number of retrotransposition events in comparison to the parental cell lines. This suggests that both FANCC and FANCA could exhibit the capacity to restrict L1 retrotransposition

I also tested a series of ORF1p mutants in these cell lines. This included a PIP mutant, an endonuclease mutant and an RT mutant. In line with data produced by Morrish *et al.* (2007), elements with an endonuclease mutation showed the capacity to mobilise in a number of cell lines deficient in DNA repair. Every FANCC cell line tested with these mutants showed the capacity to support endonuclease independent retrotransposition, as did a NHEJ deficient cell line. Conversely, the wild type cell lines tested always exhibited negligible endonuclease independent retrotransposition. This finding expands

upon previous work in the field, by demonstrating that endonuclease independent retrotransposition is not unique to the cell lines it was originally discovered in.

The PIP mutant was identified in 2013 by Taylor *et al.* The group determined that this discrete sequence, located between the EN and the RT domain, was required to mediate the interaction between PCNA and ORF1p. Mutations in the PIP domain abrogated retrotransposition to negligible levels, suggesting that the PCNA/ORF1p interaction is required for retrotransposition. Using cultured cell retrotransposition assays, I tested the mobilisation of a PIP mutant element across a range of FANCC cell lines. Interestingly, I observed that FANCA mutant cells supported retrotransposition of the PIP mutant at 16% of wild type levels. This unusual retrotransposition was rescued by complementation with FANCA protein. Activity of a PIP mutant element has not been previously reported, and was not observed in any of the other cell lines I tested, indicating that the capacity to support PIP mutant retrotransposition could be unique to FANCA cells.

Having built up a body of data around the capacity of different L1 elements to mobilise in different cell lines, and which L1 domains are required in different cellular contexts, I wanted to further investigate the relationship between the FANC pathway and the L1 element. Several FANC proteins have previously been implicated as regulators of L1 retrotransposition, and the unusual patterns of retrotransposition I observe when these proteins are knocked out indicates a possibility that their absence could alter the mechanism of mobilisation. Because the absence of FANCA in particular enabled retrotransposition of ORF1p mutant, PIP mutant and endonuclease mutant elements, I performed a pull down in a FANCA deficient cells and a parental line using ORF1p. Using this line of enquiry, I hoped I could determine whether a different selection of ORF1p interactors would be present in the mutant cell lines. This pull down revealed several ORF1p interactors which have not been previously identified in L1 screens, including YTHDF2, a protein which recognises and binds N6-methyladenosine (m6A)-containing RNAs, regulating their stability (Wang and Lu 2021). Validation of this finding by co-immunoprecipitation with T7 tagged ORF1p provided further evidence for this interaction in vitro. I also provide evidence for an in vitro interaction between ORF1p and FANCA, and ORF1p and FANCI.

Finally, I investigated the possibility that increased R-loops could increase the frequency of ENi retrotransposition. Overexpression of human nuclease RNaseH1, which suppresses R-loops, reduced both WT-LINE-1 and ENi-LINE-1 retrotransposition in FANCA deficient cells.

## 7.2 The role of ORF1p and the ribonucleoprotein particle in retrotransposition

Though it is known that ORF1p is required for canonical retrotransposition, its precise role in TPRT remains to be elucidated. ORF1p is believed to function as a nucleic acid chaperone and is also required for the formation of the ribonucleoprotein particle (Kulpa and Moran 2005; Khazina *et al.* 2011). The RNP itself is a key intermediate and has been found to be required for retrotransposition (Kulpa and Moran 2005). The fact that severe ORF1p mutants, which have previously been demonstrated to prevent formation of the RNP, can retrotranspose in FANCA cells, may indicate a different mechanistic role for the RNP in this cell line. Interestingly, my experiments demonstrate that there is substantial difference in the retrotransposition capacity of an element severely mutated in ORF1p, and an element which is missing the ORF1p altogether. Indeed, transfection of an ORF1p only element resulted in negligible retrotransposition levels in both FANCA and parental cell lines. It should be noted that cell lines expressing ORF2 only cassettes have been reported to exhibit higher levels of apoptosis as a result of toxicity, presumably induced by the enzymatic activity of the domain (Gasior *et al.* 2006). Therefore, it is possible that this toxicity also reduced the observable retrotransposition events.

The fact that FANCA cells tolerate severe ORF1p mutants which cannot form an RNP, but not removal of the ORF1 domain entirely, could indicate ORF1p provides important functional role in TPRT outside of formation of the ribonucleoprotein particle. This would support research by Kulpa and Moran (2005) who previously characterised these mutants, and observed that while the correct formation of an RNP was required for efficient retrotransposition, it was not sufficient.

Kulpa and Moran observed the behaviour of these mutants using an epitope tagged ORF1p. In contrast to previous approaches, which tended to use *in vitro* expression systems or cell lines expressing high levels of ORF1 transcript, this epitope tagged approach ensured that the RNPs being characterised were bone fide retrotransposition intermediates. As previously discussed, the Class I ORF1p mutants prevented the detection of ORF1p in the sedimentation fraction containing the RNP. A study on the mouse ORF1p (Martin *et al.* 2005) found that a mutation analogous to the human class I 261AA mutation (mouse RR297-298AA) reduces the affinity of ORF1p to bind and fold RNA by 25%. Therefore, this mutant could be altering the stability of the RNP through affecting the association of ORF1p with its RNA transcript.

Interestingly, analysis of the PCNA/ORF1p interactome by Mita *et al.* (2020) discussed in further detail below) suggests that ORF1p is exported from the nucleus during S phase, and that retrotransposition is primarily mediated by RNPs which are depleted of the protein. ORF1p may not be required for the

first steps of TPRT, including the creation of an incision in the DNA and the reverse transcription of a new L1 transcript. This scenario could potentially explain why PCNA, as well as other proteins known to interact with ORF1p in the nucleus, were not detected in the ORF1p pull down I performed in FANCA and parental cells.

Together, these findings may indicate that the FANC background influences whether a canonical RNP is required for retrotransposition, but not the downstream mechanism potentially impacted by the class II mutants. How the absence of FANCA could enable retrotransposition of elements without a canonical or stable RNP remains unclear, though it is conceivable that if FANCA acts as a repressor of L1 retrotransposition, its absence may enable retrotransposition in less optimal conditions. As the precise functional effect of the Cass I mutations is not known, it remains possible that they could prevent a post-translational modification of ORF1 which could alter how they interact with host factors such as FANCA.

A limitation of my experiments is a reliance on cell-based assays, predominantly performed in CHO cell lines. It cannot be ruled out that the ORF1p mutants behave differently in hamster cells than the HeLa cells used by Kulpa and Moran (2005), or interact differently with the host DDR pathways. The ORF1p mutants I tested are variations on highly conserved amino acid motifs (Kulpa and Moran 2005), and mechanistic binding back of ORF1p to its own transcript to form the RNP during TPRT is thought to be largely conserved between rodents and humans (Kolosha and Martin 1997). Thus, I would expect the mutants to have a similar effect mechanistically in my cell lines as that previously reported. However, further experiments to consolidate the behaviour of the ORF1p mutants in hamster cells would address this caveat. In addition, the DDR in rodents can differ to that of humans. For example, Priestly *et al.* (1998) observed that NHEJ activity is reduced by 50% in hamster cells. The reduced concentration of DDR host factors at DNA lesions may enable L1 elements to compete more effectively to exploit various DNA lesions as sites for retrotransposition. Further experiments in human FANCA cells and *in vivo* experiments would help to address these issues.



### 7.3 The role of the PIP box, and the ORF1p/PCNA interaction

A further key finding in this thesis was a mobilisation of L1 elements with a mutant PIP domain against a FANCA deficient background. The PIP box has been observed to mediate the interaction between PCNA and ORF1p (Taylor *et al.* 2013) and was observed to be necessary for retrotransposition in HeLa and HEK293T cells. PCNA is a sliding DNA clamp and an essential accessory for DNA replication (Reviewed Bruck and O'Donnell 2001). Acting primarily at the replication fork, the toroidal shaped protein behaves as a scaffold, encircling the DNA and recruiting and retaining various factors required for repair and replication of DNA (Krishna *et al.* 1994). The enzyme is a required processivity factor for DNA polymerase. Though PCNA is a known interactor of ORF1p (Taylor *et al.* 2013), its mechanistic role in retrotransposition has yet to be fully elucidated.

Mita *et al.* (2020) sought to better understand the dynamics of the PCNA-ORF2 interaction by isolating the complex and characterising the PCNA-ORF1p interactome through mass spectrometry. This experiment confirmed that the nuclear L1 complex is predominantly depleted of ORF1p and associated with replication fork proteins such as MCM3, MCM5 and MCM6, providing further evidence that PCNA is a high specificity interactor for ORF1p. This could explain why, in ORF1p pull down I performed in FANCA cells, PCNA was not detected as an interactor. It also implies that FANCA or a downstream factor is likely to be interacting with L1 at different stages in its lifecycle.

A series of biochemical and functional analyses by Mita *et al.* led to the hypothesis that ORF1p is recruited to chromatin during S phase and localises to sites of DNA replication with PCNA and replication fork proteins such as TOP1, RPA1, and PARP1. Interestingly, Taylor *et al.* 2013 demonstrated that PCNA does not interact with ORF1p mutated in its endonuclease or reverse transcriptase domains, leading to the hypothesis that ORF1p recruits PCNA after the initial steps of TPRT have taken place. Mita *et al.* 2020, therefore propose a model in which L1 RNPS localise to replication forks during S phase, some of these begin the process of retrotransposition. PCNA, which is at hand in the replication fork, could potentially be recruited and utilised to repair new L1 cDNA/genomic DNA junctions.

In the series of experiments, I performed in this thesis, retrotransposition of the PIP box mutant was unique to cells deficient in FANCA. This indicates that the phenomenon could be related to an activity of the protein itself, rather than a general FA phenotype in these cells. While, conceivably, the increase in stalled replication forks characteristic of FA cells could aid retrotransposition, it does not explain why the retrotransposition of PIP mutants is not observed in FANCC cells. One hypothesis could be

that during canonical TPRT, PCNA helps to recruit L1 to stalled replication forks. In FANCA deficient cells, where stalled forks accumulate, an increased number of elements would be able to retrotranspose without this recruitment.

The possibility that elements could be exploiting stalled replication forks was tested by Mita *et al.* by measuring retrotransposition after treatment with Aphidicolin, a chemical which leads to the generation of stalled forks. The group observed that L1 insertions increase as APH treatment causes cells to accumulate in S phase.

## 7.4 FANC proteins as regulators of L1 retrotransposition

A number of previous studies have identified different FANC proteins as potential regulators of L1 retrotransposition (Liu *et al.* 2018; Mita *et al.* 2020). Tristan-Ramos *et al.* (2020) directly observed that cells in which FANCA or FANCC is knocked down supported higher levels of L1 retrotransposition. In both cases, complementation with the respective protein rescued the observed effect.

My data corroborates both of these findings. I observe higher levels of wild type L1 retrotransposition in cells deficient in FANCA than in a parental cell line, and was able to rescue this effect with FANCA complemented cells. Similarly, I observe higher levels of retrotransposition by wild type elements in FANCD efficient cells in comparison to a parental cell line. Interestingly, FANCA has also demonstrated an affinity for single-stranded RNAs (Yuan *et al.* 2012), which could provide one potential means by which the protein could mediate suppression of L1 activity.

Several genome-wide studies searching for potential L1 interactors have been conducted (Taylor *et al.* 2013; Dai *et al.* 2014; Taylor *et al.* 2016; Taylor *et al.* 2018; Mita *et al.* 2018). FANCC and FANCA have been identified as potential interactors by two of these studies (Liu *et al.* 2018) and (Mita *et al.* 2020). Mita *et al.* (2020) also performed secondary validation on a number of factors that they identified to determine if an effect on retrotransposition could be seen. Therefore, together with the evidence presented in this thesis, there seems to be good support for these proteins as direct or indirect repressors of retrotransposition. The rescuing of this effect with complemented cells demonstrates that this finding is not because of secondary DNA damage caused by the FANC phenotype.

As well as FANCA and C numerous other FANC proteins have been identified as potential repressors of retrotransposition, and some as activators. **Table 5.1** summarises these findings in more detail. The proteins identified feature in different stages in the FANC pathway, some are part of the core complex, and some downstream. FANCD2, which is mono-ubiquitinated during the resolution of ICLS by the

FANC pathway (D'Andrea and Grompe 2003; Meetei *et al.* 2004; Walden and Deans 2014), has itself also been identified as a potential L1 interactor. It therefore seems that the FANC pathway could be interacting with L1 retrotransposition at multiple stages. It is not yet clear whether the core proteins could be interacting with/repressing L1 as one complex, or as individuals. As many of the FANC proteins have roles outside of their part of the FANC pathway (Sumpter *et al.* 2017), it is also unclear if this forms part of a secondary function or a result of the pathway as a whole.

## 7.5 YTHDF2 as a potential negative regulator of retrotransposition

One of the primary candidates I identified as a potential interactor of ORF1p was YTHDF2. Of the proteins that were significantly enriched in the sample in comparison to the negative control, YTHDF2 had one of the higher fold changes.

There are several reasons why this protein is interesting with regard to this research. YTHDF2 works to destabilise RNA containing the m6A modification, through recruitment of the CCR4-NOT deadenylase complex (Du *et al.* 2016; Wang and Lu 2021). Methylation at the M6A is one of the most abundant RNA modifications and has also been found to destabilise RNA: DNA hybrids (Reviewed in He *et al.* 2021). Furthermore, YTHDF2 has been found to interact with R-loop containing loci in vitro, and knocking out YTHDF2 results in an increase in the number of R-loops (Okamoto *et al.* 2019). I have already discussed the possibility that L1 uses R-loops to mobilise in **Chapter 6: Figure 6.15**.

Interestingly, the m6A 'writer' METTL3 has been observed to increase retrotransposition, while ALKBH5, the "eraser", appears to remove it (Hwang *et al.* 2021). The group hypothesise that L1 "hijacks" this modification system to facilitate retrotransposition, which would corroborate my finding that YTHDF2 could serve as a negative regulator of retrotransposition. An additional study (Xiong *et al.* 2021) identified a substantial number of L1 elements containing the modification through a technique called MINT-seq, similarly noting that the modification facilitates retrotransposition. Hwang *et al.* (2021) demonstrate that the modification serves as a docking site for Eukaryotic Initiation factor (EIF), which was identified as an L1-ORF1p interactor in my proteomic screens in both FANCA mutant (VH4) and parental (V79) cells. It can therefore also be conceived that regulation of L1 by YTHDF2 is related to the modification, and that the protein interacts with and degrades L1 elements which contain this marker. It remains unclear why YTHDF2 would reduce L1 retrotransposition primarily in FA cells in this instance, unless L1 obtains more m6A in a FA context.

## 7.6 A potential model: R-loops as a site for insertion.

Recently, there has been substantial interest in R loops, a structure that is made up of a RNA: DNA hybrid and a displaced single strand of DNA. R loops represent a source of genetic instability in mammalian cells, and can also contribute to a number of important biological processes. To prevent the formation of these structures, our cells have evolved a number of protective mechanisms. One protein involved in the prevention of R-loops is BRCA1, or FANCD1 of the FANCD pathway. In fact, Garcia-Rubio have shown that the FANCA pathway removes R-loops and suggests that protecting cells from R-loops might be one major function of the pathway.

Previous studies have demonstrated that the L1 is capable of exploiting a variety of different lesions as sites for retrotransposition. ENi retrotransposition has been observed at unrepaired double stranded breaks and telomeres. It is possible that the L1 element could be exploiting the accumulated R-loops in FANCC cells as sites for retrotransposition. Indeed, my work demonstrates that overexpression of RNAseH1, which cleaves RNA: DNA hybrids, leads to the reduction of wild type and ENi retrotransposition.

It cannot be ruled out that overexpression of RNAseH1 reduces L1 retrotransposition via another mechanism. In addition, though this model might explain some of the increased retrotransposition we see in FANCC cells, it is still unclear why I observe different results depending on which FANCD protein is deleted. This may be due to differing contributions of the proteins to the overall pathway, or it might indicate that certain proteins, FANCA for example, play a more direct regulatory role.

While there are limited studies investigating the relationship between RNAseH1 and L1, Benitez-guijarro *et al.* (2018) observed that RNAseH2 promotes L1 retrotransposition. RNAseH2 also degrades DNA: RNA hybrids, and catalyses the majority of RNAse activity in the nucleus. Notably, RNAseH2 is mutated in Aicardi-Goutières syndrome (AGS), an autosomal recessive inflammatory encephalopathy related to errors in DNA metabolism. Cristini *et al.* (2022) demonstrate that RNAseH2 aids the resolution of transcriptional R-loops, and Benitez-Guijarro showed that overexpression of RNAseH2 increased retrotransposition in HeLa and U2OS cells. Thus, the evidence suggests that the suppression of L1 activity by RNAseH1 is specific to the FANCC context.

## 7.7 Future work and outstanding questions

This thesis contributes to the current understanding of the relationship between the FANC pathway and L1 retrotransposition. However, there are still outstanding questions.

In this thesis I have demonstrated that FANCA mutant cells support retrotransposition of L1 elements mutated in a number of domains. While we can speculate, the mechanistic implications of this finding remain to be elucidated. There are several lines of enquiry that could provide further insight. Having established that FANCA interacts with ORF1p *in vitro*, it would be interesting to investigate whether mutating certain domains of the protein affects this interaction. This would help to deduce whether FANCA is binding ORF1p through particular motifs, such as one of the conserved amino acid motifs in the carboxyl terminus (REKG235-238, RR261-262 and YPAKLS282-287) (Kulpa and Moran 2005). It would also be interesting to determine whether binding of FANCA affects the formation of the RNP. This could be tested using a similar rationale to that of Kulpa and Moran: cells could be transfected with a construct overexpressing FANCA. Then, using western blotting and differential centrifugation it could be determined whether ORF1p sediments with the RNP (in the 160,000 g cytoplasmic fraction (Kulpa and Moran 2005)).

Furthermore, it would be interesting to determine whether other members of the FANC pathway also interact with ORF1p. Indeed, as shown in **Table 5.1**, a number of FANCF actors have been identified as potential regulators of L1 retrotransposition. However, the mechanistic basis of this regulation is not known, including whether it is mediated by direct interaction with L1. Among the cell lines tested in this thesis, the FANCA mutant was the only one which supported retrotransposition of ORF1p mutants and the PIP mutant. This could be a direct or indirect result of a particular activity of FANCA, for example direct regulation of the L1 element. Alternatively, other proteins in the core FA complex, or in the pathway as a whole, could also interact with L1 directly.

It also would be interesting to perform *in vivo* immunofluorescence experiments to identify whether FANCA co-localises with the RNP during the retrotransposition cycle. This might provide further insight into whether FANCA and ORF1p are interacting *in vivo* and how modifications of the L1 cassette, for example through the ORF1p mutations, impact this interaction.

Another important finding in this thesis was the identification of YTHDF2 as a potential ORF1p interactor. Having better optimised the retrotransposition assay in YTHDF2 cells and their parental line, I would like to better establish whether a range of different ORF1p mutant elements can retrotranspose efficiently in these cell lines.

Indeed, if these mutants are also able to retrotranspose in YTHDF2 mutant cells, it would indicate the possibility of a connection between the protein and the unusual retrotransposition I see in FANCA deficient cells. It could be that YTHDF2 negatively regulates retrotransposition in cooperation with FANCA, and that in the absence of this protein thus enables mobilisation of these elements. Furthermore, it would be interesting to investigate which specific conditions facilitate the interaction between ORF1p and YTHDF2. My data suggests that YTHDF2 could interact more with ORF1p in FANCD deficient cells, however it's not clear what is causing that distinction. Testing whether a higher proportion of L1 elements contain the m6A modification in FA cells would provide further insight, as well as investigating whether YTHDF2 is upregulated in a FA context.

I would also like to further characterise the insertions that were generated by the ORF1p mutant elements. Studies have shown that ENm insertions have some distinct characteristics providing clues about the mechanism they were generated by.

My work also demonstrated that overexpression of human nuclear RNaseH1, which suppresses R-loops, leads to a reduction of both wild type and ENretrotransposition. This led to the development of the model shown in **Figure 6.15** would also like to further investigate whether the overexpression of RNase H1 would also lead to a reduction in the abnormal retrotransposition I observed in FANCC cells during this thesis, for example of ORF1p mutants or the PIP mutant. This might indicate whether the increased retrotransposition I observe is related to the increase in R-loops or another characteristic of the FANCC cells. Furthermore, it would be interesting to repeat these experiments with RNaseH2, as Benitez-Guijarro previously found that overexpression of the protein resulted in an increase in L1 activity. Repeating this experiment in a FANCC context would better illuminate the relationship between the FANCC pathway, L1 retrotransposition and R-loops



## 8.0 Appendices

### 8.1 Antibodies used in this thesis

Antibody	Species		Reference
<b>Anti T7-tag</b>	Goat	1:1000	Abcam ab9138
<b>Anti FANCI</b>	Rabbit	1:2000	Abcam ab245219
<b>Anti-FANCA /FAA</b>	Rabbit	1:1000	Abcam ab201457
<b>Anti YTHDF2</b>	Rabbit	1:200	Abcam ab220163
<b>Vinculin</b>	Mouse	1:2,000	Sigma, V9264
<b>Lamin B1 (WB)</b>	Rabbit	1:20,000	Abcam, ab16048
<b>Goat Anti-Rabbit IgG</b>	Goat	1:20,000	Abcam ab6721
<b>Donkey Anti-Goat IgG</b>	Donkey	1:5000	Abcam ab6885

## References

Abakir, A., Giles, T. C., Cristini, A., Foster, J. M., Dai, N., Stanczak, M., ... & Ruzov, A. (2020). N 6-methyladenosine regulates the stability of RNA: DNA hybrids in human cells. *Nature genetics*, 52(1), 48-55.

Alisch, R. S., Garcia-Perez, J. L., gure, A. R., Gage, F. H., & Moran, J. V. (2006). Unconventional translation of mammalian LINE-1 retrotransposons. *Genes & development*, 20(2), 210-224.

Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M., & Gvozdev, V. A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Current Biology*, 11(13), 1017-1027.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., ... & Tuschl, T. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, 442(7099), 203-207.



- Aravin, A. A., Hannon, G. J., & Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *science*, 318(5851), 761-764.
- Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., ... & Hannon, G. J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular cell*, 31(6), 785-799.
- Ardeljan, D., Wang, X., Oghbaie, M., Taylor, M. S., Husband, D., Deshpande, V., ... & LaCava, J. (2020). LINE-1 ORF1p expression is nearly imperceptible in human cancers. *Mobile dna*, 11, 1-19.
- Athanikar, J. N., Badge, R. M., & Moran, J. V. (2004). A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic acids research*, 32(13), 3846-3855.
- Auerbach, A. D. (1988). A test for Fanconi's anemia. *Blood*, 72(1), 366-367.
- Babushok, D. V., Ostertag, E. M., Courtney, C. E., Choi, J. M., & Kazazian, H. H. (2006). L1 integration in a transgenic mouse model. *Genome research*, 16(2), 240-250.
- Barlow, J. H., Faryabi, R. B., Callén, E., Wong, N., Malhowski, A., Chen, H. T., ... & Nussenzweig, A. (2013). Identification of early replicating fragile sites that contribute to genome instability. *Cell*, 152(3), 620-632.
- Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., ... & Moran, J. V. (2010). LINE-1 retrotransposition activity in human genomes. *Cell*, 141(7), 1159-1170.
- Beck, C. R., Garcia-Perez, J. L., Badge, R. M., & Moran, J. V. (2011). LINE-1 elements in structural variation and disease. *Annual review of genomics and human genetics*, 12, 187-215.
- 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 international*, 6(1), 1-10.
- Benitez-Guijarro, M., Lopez-Ruiz, C., Tarnauskaitė, Ž., Murina, O., Mian Mohammad, M., Williams, T. C., ... & Garcia-Perez, J. L. (2018). RNase H2, mutated in Aicardi-Goutières syndrome, promotes LINE-1 retrotransposition. *The EMBO Journal*, 37(15), e98506.
- Boissinot, S., Chevret, P., & Furano, A. V. (2000). L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Molecular biology and evolution*, 17(6), 915-928.
- Boorstein, R. J., Haldar, J., Poirier, G., & Putnam, D. (1992). DNA base excision repair stimulates poly (ADP-ribose) synthesis. In *ADP-Ribosylation Reactions* (pp. 173-177). Springer New York.
- Boule, J. B., & Zakian, V. A. (2007). The yeast Pif1p DNA helicase preferentially unwinds RNA-DNA substrates. *Nucleic acids research*, 35(17), 5809-5818.
- Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, 431(7004), 96-99.
- Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science*, 294(5551), 2536-2539.
- Bulut-Karslioglu, A., Inti, A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., ... & Jenuwein, T. (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Molecular cell*, 55(2), 277-290.

- Burwinkel, B., & Kilimann, M. W. (1998). Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *Journal of molecular biology*, 277(3), 513-517.
- Bradley, M. O., Bhuyan, B., Francis, M. C., Langenbach, R., Peterson, A., & Huberman, E. (1981). Mutagenesis by chemical agents in V79 Chinese hamster cells: A review and analysis of the literature: A report of the gene-tox program. *Mutation Research/Reviews in Genetic Toxicology*, 87(2), 81-142.
- Briggs, E. M., Mita, P., Sun, X., Ha, S., Vasilyev, N., Leopold, Z. R., ... & Logan, S. K. (2021). Unbiased proteomic mapping of the LINE-1 promoter using CRISPR Cas9. *Mobile DNA*, 12, 1-12.
- Bratthauer, G. L., Cardiff, R. D., & Fanning, T. G. (1994). Expression of LINE-1 retrotransposons in human breast cancer. *Cancer*, 73(9), 2333-2336.
- Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., & Kazazian Jr, H. H. *the National Academy of Sciences*, 100(9), 5280-5285.
- Bruck, I., & O'Donnell, M. (2001). The ring-type polymerase sliding clamp family. *Genome biology*, 2(1), 1-3.
- Callinan, P. A., & Batzer, M. A. (2006). Retrotransposable elements and human disease. *Genome and disease*, 1, 104-115.
- Cantor, S., Drapkin, R., Zhang, F., Lin, Y., Han, J., Pamidi, S., & Livingston, D. M. (2004). The BRCA1-henmutations. *Proceedings of the National Academy of Sciences*, 101(8), 2357-2362.
- Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, Duc J, Jang SM, Turelli P, Trono D (2014) Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev* 28:1397–1409.
- Ceccaldi, R., Rondinelli, B., & D'Andrea, A. D. (2016). Repair pathway choices and consequences at the double-strand break. *Trends in cell biology*, 26(1), 52-64.
- Cerritelli, S. M., & Crouch, R. J. (2009). Ribonuclease H: the enzymes in eukaryotes. *The FEBS journal*, 276(6), 1494-1505.
- Chang, E. Y. C., Tsai, S., Aristizabal, M. J., Wells, J. P., Coulombe, Y., Busatto, F. F., ... & Stirling, P. C. (2019). MRE11-RAD50-NBS1 promotes Fanconi Anemia R-loop suppression at transcription–replication conflicts. *Nature communications*, 10(1), 4265.
- Charles, M., Belcram, H., Just, J., Huneau, C., Viollet, A., Couloux, A., ... & Chalhoub, B. (2008). Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. *Genetics*, 180(2), 1071-1086.
- Chaudhury, I., Sareen, A., Raghunandan, M., & Sobeck, A. (2013). FANCD2 regulates BLM complex functions independently of FANCI to promote replication fork recovery. *Nucleic acids research*, 41(13), 6444-6459.
- Chaung, W., Mi, L. J., & Boorstein, R. J. (1997). The p53 status of Chinese hamster V79 cells frequently used for studies on DNA damage and DNA repair. *Nucleic Acids Research*, 25(5), 992-994.

- Chen, Z., Li, Y., & Krug, R. M. (1999). Influenza A virus NS1 protein targets poly (A)-binding protein II of the cellular 3'-end processing machinery. *The EMBO journal*, 18(8), 2273-2283.
- Chen, J. M., Stenson, P. D., Cooper, D. N., & Férec, C. (2005). A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Human genetics*, 117, 411-427.
- Cherubini, G., Naim, V., Caruso, P., Burla, R., Bogliolo, M., Cundari, E., ... & Rosselli, F. (2011). The FANCD1 pathway is activated by adenovirus infection and promotes viral replication-dependent recombination. *Nucleic acids research*, 39(13), 5459-5473.
- Chow JC, Ciaudo C, Fazzari MJ, Mise N, Servant N, Glass JL et al (2010) LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* 141(6):956–969
- Christian, C. M., Sokolowski, M., deHaro, D., Kines, K. J., & Belancio, V. P. (2017). Involvement of conserved amino acids in the C-terminal region of LINE-1 ORF1p in retrotransposition. *Genetics*, 205(3), 1139-1149.
- Clements, A. P., & Singer, M. F. (1998). The human LINE-1 reverse transcriptase: effect of deletions outside the common reverse transcriptase domain. *Nucleic acids research*, 26(15), 3528-3535.p
- Cordaux, R. (2008). The human genome in the LINE of fire. *Proceedings of the National Academy of Sciences*, 105(49), 19033-19034.
- Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nature reviews genetics*, 10(10), 691-703.
- Cost, G. J., & Boeke, J. D. (1998). Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry*, 37(51), 18081-18093.
- Cost, G. J., Feng, Q., Jacquier, A., & Boeke, J. D. (2002). Human L1 element target-primed reverse transcription in vitro. *The EMBO journal*, 21(21), 5899-5910.
- Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., ... & Gage, F. H. (2009). L1 retrotransposition in human neural progenitor cells. *Nature*, 460(7259), 1127-1131.
- Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Marchetto, M. C., Muotri, A. R., Mu, Y., ... & Gage, F. H. (2011). Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proceedings of the National Academy of Sciences*, 108(51), 20382-20387.
- Coulthard, R., Deans, A. J., Swuec, P., Bowles, M., Costa, A., West, S. C., & McDonald, N. Q. (2013). Architecture and DNA recognition elements of the Fanconi anemia FANCD1-FAAP24 complex. *Structure*, 21(9), 1648-1658.
- Cumming, R. C., Liu, J. M., Youssoufian, H., & Buchwald, M. (1996). Suppression of apoptosis in hematopoietic factor-dependent progenitor cell lines by expression of the FANCD1 gene.
- Dai, L., LaCava, J., Taylor, M. S., & Boeke, J. D. (2014). Expression and detection of LINE-1 ORF-encoded proteins. *Mobile genetic elements*, 4(3), 1034-48.

- Day, D. S., Luquette, L. J., Park, P. J., & Kharchenko, P. V. (2010). Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome biology*, 11, 1-12.
- Daza-Martin, M., Starowicz, K., Jamshad, M., Tye, S., Ronson, G. E., MacKay, H. L., ... & Morris, J. R. (2019). Isomerization of BRCA1–BARD1 promotes replication fork protection. *Nature*, 571(7766), 521-527.
- D'Andrea, A. D., & Grompe, M. (2003). The Fanconi anaemia/BRCA pathway. *Nature Reviews Cancer*, 3(1), 23-34.
- Denli, A. M., Narvaiza, I., Kerman, B. E., Pena, M., Benner, C., Marchetto, M. C., ... & Gage, F. H. (2015). Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell*, 163(3), 583-593.
- Deragon, J. M., Sinnett, D., & Labuda, D. (1990). Reverse transcriptase activity from human embryonal carcinoma cells NTera2D1. *The EMBO Journal*, 9(10), 3363-3368.
- Dmitriev, S. E., Andreev, D. E., Terenin, I. M., Olovnikov, I. A., Prassolov, V. S., Merrick, W. C., & Shatsky, I. N. (2007). Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Molecular and cellular biology*, 27(13), 4685-4697.
- Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F., & Kazazian Jr, H. H. (1991). Isolation of an active human transposable element. *Science*, 254(5039), 1805-1808.
- Doucet, A. J., Wilusz, J. E., Miyoshi, T., Liu, Y., & Moran, J. V. (2015). A 3' poly (A) tract is required for LINE-1 retrotransposition. *Molecular cell*, 60(5), 728-741.
- Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., ... & Wu, L. (2016). YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. *Nature communications*, 7(1), 12626.
- Dunn, C. A., Medstrand, P., & Mager, D. L. (2003). An endogenous retroviral long terminal repeat is the dominant promoter for human  $\beta$ 1, 3-galactosyltransferase 5 in the colon. *Proceedings of the National Academy of Sciences*, 100(22), 12841-12846.
- Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., & Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes & development*, 18(13), 1618-1629.
- Eickbush, T. H. (2002). Repair by retrotransposition. *Nature Genetics*, 31(2), 126-127.
- Esnault, C., Maestre, J., & Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nature genetics*, 24(4), 363-367.
- Ernst, C., Odom, D. T., & Kutter, C. (2017). The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. *Nature communications*, 8(1), 1411.
- Faulkner, G. J., & Billon, V. (2018). L1 retrotransposition in the soma: a field jumping ahead. *Mobile DNA*, 9(1), 1-18.

- Feng, Q., Moran, J. V., Kazazian Jr, H. H., & Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, 87(5), 905-916.
- Flasch, D. A., Macia, Á., Sánchez, L., Ljungman, M., Heras, S. R., García-Pérez, J. L., ... & Moran, J. V. (2019). Genome-wide de novo L1 retrotransposition connects endonuclease activity with replication. *Cell*, 177(4), 837-851.
- Freeman, J. D., Goodchild, N. L., & Mager, D. L. (1994). A modified indicator gene for selection of retrotransposition events in mammalian cells. *Biotechniques*, 17(1), 46-52.
- Friedman, J. R., Fredericks, W. J., Jensen, D. E., Speicher, D. W., Huang, X. P., Neilson, E. G., & Rauscher, F. J. (1996). KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes & development*, 10(16), 2067-2078.
- Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M. S., Timmers, C., Hejna, J., ... & D D'Andrea, A. (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Molecular cell*, 7(2), 249-262.
- García-Muse, T., & Aguilera, A. (2016). Transcription–replication conflicts: how they occur and how they are resolved. *Nature reviews Molecular cell biology*, 17(9), 553-563.
- García-Pérez, E., Parra Soto, J. B., Ovín Ania, M. C., García Sánchez, A., Van Baten, J. M., Krishna, R., ... & Calero, S. (2007). A computational study of CO<sub>2</sub>, N<sub>2</sub>, and CH<sub>4</sub> adsorption in zeolites.
- Garcia-Perez, J. L., Morell, M., Scheys, J. O., Kulpa, D. A., Morell, S., Carter, C. C., ... & Moran, J. V. (2010). Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature*, 466(7307), 769-773.
- García-Rubio, M. L., Perez-Calero, C., Barroso, S. I., Tumini, E., Herrera-Moyano, E., Rosado, I. V., & Aguilera, A. (2015). The Fanconi anemia pathway protects genome integrity from R-loops. *PLoS genetics*, 11(11), e1005674.
- Gilbert, N., Lutz-Prigge, S., & Moran, J. V. (2002). Genomic deletions created upon LINE-1 retrotransposition. *Cell*, 110(3), 315-325.
- Gilbert, N., Lutz, S., Morrish, T. A., & Moran, J. V. (2005). Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Molecular and cellular biology*, 25(17), 7780-7795.
- Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J. L., & Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes & development*, 25(19), 2041-2056.
- García-de Teresa, B., Frias, S., Molina, B., Villarreal, M. T., Rodriguez, A., Carnevale, A., ... & Torres, L. (2019). FANCC Dutch founder mutation in a Mennonite family from Tamaulipas, México. *Molecular Genetics & Genomic Medicine*, 7(6), e710.
- Gasior, S. L., Roy-Engel, A. M., & Deininger, P. L. (2008). ERCC1/XPF limits L1 retrotransposition. *DNA repair*, 7(6), 983-989.
- Gasior, S. L., Wakeman, T. P., Xu, B., & Deininger, P. L. (2006). The human LINE-1 retrotransposon creates DNA double-strand breaks. *Journal of molecular biology*, 357(5), 1383-1393.

- Giaccia, A. J., MacLaren, R. A., Nicholas, D., Nicolaou, D., & Stamato, T. D. (1990). Increased sensitivity to killing by restriction enzymes in the XR-1 DNA double-strand break repair-deficient mutant. *Mutation Research/DNA Repair*, 236(1), 67-76.
- Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I., & Chédin, F. (2013). GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome research*, 23(10), 1590-1600.
- Goodier, J. L., Ostertag, E. M., & Kazazian Jr, H. H. (2000). Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Human molecular genetics*, 9(4), 653-657.
- Goodier, J. L., Zhang, L., Vetter, M. R., & Kazazian Jr, H. H. (2007). LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Molecular and cellular biology*, 27(18), 6469-6483.
- Goodier, J. L., & Kazazian Jr, H. H. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*, 135(1), 23-35.
- Goodier, J. L., Ostertag, E. M., Engleka, K. A., Seleme, M. C., & Kazazian Jr, H. H. (2004). A potential role for the nucleolus in L1 retrotransposition. *Human molecular genetics*, 13(10), 1041-1048.
- Gonçalves, A., Oliveira, J., Coelho, T., Taipa, R., Melo-Pires, M., Sousa, M., & Santos, R. (2017). Exonization of an intronic LINE-1 element causing Becker muscular dystrophy as a novel mutational mechanism in dystrophin gene. *Genes*, 8(10), 253.
- Gotea, V., & Makołowski, W. (2006). Do transposable elements really contribute to proteomes?. *TRENDS in Genetics*, 22(5), 260-267.
- Grimaldi, G., Skowronski, J., & Singer, M. F. (1984). Defining the beginning and end of KpnI family segments. *The EMBO journal*, 3(8), 1753-1759.
- Hancks, D. C., & Kazazian Jr, H. H. (2012). Active human retrotransposons: variation and disease. *Current opinion in genetics & development*, 22(3), 191-203.
- Haneline, L. S., Broxmeyer, H. E., Cooper, S., Hangoc, G., Carreau, M., Buchwald, M., & Clapp, D. W. (1998). Multiple inhibitory cytokines induce deregulated progenitor growth and apoptosis in hematopoietic cells from Fac-/- mice. *Blood, The Journal of the American Society of Hematology*, 91(11), 4092-4098.
- Hashimoto Y, Ray Chaudhuri A, Lopes M, Costanzo V. Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. *Nature structural & molecular biology*. 2010 Nov;17(11):1305-11.
- Hata, K., Okano, M., Lei, H., & Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice.
- He, P. C., & He, C. (2021). m6A RNA methylation: from mechanisms to therapeutic potential. *The EMBO journal*, 40(3), e105977.
- Hefferin, M. L., & Tomkinson, A. E. (2005). Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA repair*, 4(6), 639-648.

- Heras, S. R., Macias, S., Plass, M., Fernandez, N., Cano, D., Eyra, E., ... & Cáceres, J. F. (2013). The Microprocessor controls the activity of mammalian retrotransposons. *Nature structural & molecular biology*, 20(10), 1173-1181
- Higgs, Martin R., John J. Reynolds, Alicja Winczura, Andrew N. Blackford, Valérie Borel, Edward S. Miller, Anastasia Zlatanou et al. "BOD1L is required to suppress deleterious resection of stressed replication forks." *Molecular cell* 59, no. 3 (2015): 462-477.
- Hohjoh, H., & Singer, M. F. (1996). Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *The EMBO journal*, 15(3), 630-639.
- Holmes, S. E., Dombroski, B. A., Krebs, C. M., Boehm, C. D., & Kazazian Jr, H. H. (1994). A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nature genetics*, 7(2), 143-148.
- Horn, A. V., Celic, I., Dong, C., Martirosyan, I., & Han, J. S. (2017). A conserved role for the ESCRT membrane budding complex in LINE retrotransposition. *PLoS Genetics*, 13(6), e1006837.
- Houghtaling, S., Newell, A., Akkari, Y., Taniguchi, T., Olson, S., & Grompe, M. (2005). Fancd2 functions in a double strand break repair pathway that is distinct from non-homologous end joining. *Human molecular genetics*, 14(20), 3027-3033.
- Howlett, N. G., Taniguchi, T., Durkin, S. G., D'Andrea, A. D., & Glover, T. W. (2005). The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. *Human molecular genetics*, 14(5), 693-701.
- Huang, Y., Leung, J. W., Lowery, M., Matsushita, N., Wang, Y., Shen, X., ... & Li, L. (2014). Modularized functions of the Fanconi anemia core complex. *Cell reports*, 7(6), 1849-1857.
- Huang, R., & Zhou, P. K. (2021). DNA damage repair: Historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. *Signal Transduction and Targeted Therapy*, 6(1), 254.
- Hussain, S., Wilson, J. B., Medhurst, A. L., Hejna, J., Witt, E., Ananth, S., ... & Mathew, C. G. (2004). Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. *Human molecular genetics*, 13(12), 1241-1248.
- Hwang, S. Y., Jung, H., Mun, S., Lee, S., Park, K., Baek, S. C., ... & Ahn, K. (2021). L1 retrotransposons exploit RNA m6A modification as an evolutionary driving force. *Nature communications*, 12(1), 880.
- Ikegami, S., TAGUCHI, T., OHASHI, M., OGURO, M., NAGANO, H., & MANO, Y. (1978). Aphidicolin prevents mitotic cell division by interfering with the activity of DNA polymerase- $\alpha$ . *Nature*, 275(5679), 458-460.
- Ivancevic, A. M., Kortschak, R. D., Bertozzi, T., & Adelson, D. L. (2016). LINEs between species: evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. *Genome biology and evolution*, 8(11), 3301-3322.
- Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D (2014) An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516:242–245.

- Januszyk, K., Li, P. W. L., Villareal, V., Branciforte, D., Wu, H., Xie, Y., ... & Clubb, R. T. (2007). Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1. *Journal of Biological Chemistry*, 282(34), 24893-24904.
- Jones, N. J. (1994). Genetic analysis of mitomycin C-hypersensitive Chinese hamster cell mutants. *Mutagenesis*, 9(5), 477-482.
- Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proceedings of the National Academy of Sciences*, 94(5), 1872-1877.
- Kaer, K., & Speek, M. (2013). Retroelements in human disease. *Gene*, 518(2), 231-241.
- Kais, Z., Rondinelli, B., Holmes, A., O'Leary, C., Kozono, D., D'Andrea, A. D., & Ceccaldi, R. (2016). FANCD2 maintains fork stability in BRCA1/2-deficient tumors and promotes alternative end-joining DNA repair. *Cell reports*, 15(11), 2488-2499.
- Kaneda, A., Tsukamoto, T., Takamura-Enya, T., Watanabe, N., Kaminishi, M., Sugimura, T., ... & Ushijima, T. (2004). Frequent hypomethylation in multiple promoter CpG islands is associated with global hypomethylation, but not with frequent promoter hypermethylation. *Cancer science*, 95(1), 58-64.
- Kano, H., Godoy, I., Courtney, C., Vetter, M. R., Gerton, G. L., Ostertag, E. M., & Kazazian, H. H. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes & development*, 23(11), 1303-1312.
- Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., ... & Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Human molecular genetics*, 16(19), 2272-2280.
- Kazazian Jr, H. H., 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), 164-166.
- Kemp, M. G., & Sancar, A. (2012). DNA excision repair: where do all the dimers go?. *Cell Cycle*, 11(16), 2997-3002.
- Khazina, E., & Weichenrieder, O. (2009). Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. *Proceedings of the National Academy of Sciences*, 106(3), 731-736.
- Khazina, E., Truffault, V., Büttner, R., Schmidt, S., Coles, M., & Weichenrieder, O. (2011). Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nature structural & molecular biology*, 18(9), 1006-1014.
- Kim S, Günesdogan U, Zylicz JJ, Hackett JA, Cougot D, Bao S, Lee C, Dietmann S, Allen GE, Sengupta R, Surani MA (2014) PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Mol Cell* 56:564–579.



- Kim, H., Yang, K., Dejsuphong, D., & D'Andrea, A. D. (2012). Regulation of Rev1 by the Fanconi anemia core complex. *Nature structural & molecular biology*, 19(2), 164-170.
- Kines, K. J., Sokolowski, M., deHaro, D. L., Christian, C. M., & Belancio, V. P. (2014). Potential for genomic instability associated with retrotranspositionally-incompetent L1 loci. *Nucleic acids research*, 42(16), 10488-10502.
- Kirilyuk, A., Tolstonog, G. V., Damert, A., Held, U., Hahn, S., Löwer, R., ... & Schumann, G. G. (2008). Functional endogenous LINE-1 retrotransposons are expressed and mobilized in rat chloroleukemia cells. *Nucleic acids research*, 36(2), 648-665.
- Knipscheer, P., Räschele, M., Smogorzewska, A., Enou, M., Ho, T. V., Schärer, O. D., ... & Walter, J. C. (2009). The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science*, 326(5960), 1698-1701.
- Kolinjivadi, A. M., Crismani, W., & Ngeow, J. (2020). Emerging functions of Fanconi anemia genes in replication fork protection pathways. *Human molecular genetics*, 29(R2), R158-R164.
- Kolosha, V. O., & Martin, S. L. (1997). In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proceedings of the National Academy of Sciences*, 94(19), 10155-10160.
- Kolosha, V. O., & Martin, S. L. (2003). High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). *Journal of Biological Chemistry*, 278(10), 8112-8117.
- Kondo-Iida, E., Kobayashi, K., Watanabe, M., Sasaki, J., Kumagai, T., Koide, H., ... & Toda, T. (1999). Novel mutations and genotype-phenotype relationships in 107 families with Fukuyama-type congenital muscular dystrophy (FCMD). *Human Molecular Genetics*, 8(12), 2303-2309.
- Kopera, H. C., Larson, P. A., Moldovan, J. B., Richardson, S. R., Liu, Y., & Moran, J. V. (2016). LINE-1 cultured cell retrotransposition assay. *Transposons and Retrotransposons: Methods and Protocols*, 139-156.
- Kotecki, M., Reddy, P. S., & Cochran, B. H. (1999). Isolation and characterization of a near-haploid human cell line. *Experimental cell research*, 252(2), 273-280.
- Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M., & Kuriyan, J. (1994). Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*, 79(7), 1233-1243.
- Kubo, S., Seleme, M. D. C., Soifer, H. S., Perez, J. L. G., Moran, J. V., Kazazian Jr, H. H., & Kasahara, N. (2006). L1 retrotransposition in nondividing and primary human somatic cells. *Proceedings of the National Academy of Sciences*, 103(21), 8036-8041.
- Kulpa, D. A., & Moran, J. V. (2005). Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Human molecular genetics*, 14(21), 3237-3248.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, 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 JP, 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 JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, 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 DR, DoucetteStamm L, Rubenfield M, Weinstock K, Lee HM, 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 RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Retrotransposons and the Mammalian Germline 24 Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowski J, International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

Lanikova, L., Kucerovala, J., Indrak, K., Divoka, M., Issa, J. P., Papayannopoulou, T., ... & Divoky, V. (2013).  $\beta$ -Thalassemia Due to Intronic LINE-1 Insertion in the  $\beta$ -Globin Gene (HBB): Molecular Mechanisms Underlying Reduced Transcript Levels of the  $\beta$ -Globin L 1 Allele. *Human mutation*, 34(10), 1361-1365.

Lahr, R. M., Fonseca, B. D., Ciotti, G. E., Al-Ashtal, H. A., Jia, J. J., Niklaus, M. R., ... & Berman, A. J. (2017). La-related protein 1 (LARP1) binds the mRNA cap, blocking eIF4F assembly on TOP mRNAs. *Elife*, 6, e24146.

Leibold, D. M., Swergold, G. D., Singer, M. F., Thayer, R. E., Dombroski, B. A., & Fanning, T. G. (1990). Translation of LINE-1 DNA elements in vitro and in human cells. *Proceedings of the National Academy of Sciences*, 87(18), 6990-6994.

Levin, H. L., & Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics*, 12(9), 615-627.

Levrano, O., Diotti, R., Pujara, K., Batish, S. D., Hanenberg, H., & Auerbach, A. D. (2005). Spectrum of sequence variations in the FANCA gene: an International Fanconi Anemia Registry (IFAR) study. *Human mutation*, 25(2), 142-149.

- Li, F., Sarangi, P., Feng, H., Moreau, L., Nguyen, H., Clairmont, C., & D'Andrea, A. D. (2021). REV7/FANCV Binds to CHAMP1 and Promotes Homologous Recombination Repair. *bioRxiv*, 2021-10.
- Li, Z., Otevrel, T., Gao, Y., Cheng, H. L., Seed, B., Stamato, T. D., ... & Alt, F. W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V (D) J recombination. *Cell*, 83(7), 1079-1089.
- Liang, C. C., Zhan, B., Yoshikawa, Y., Haas, W., Gygi, S. P., & Cohn, M. A. (2015). UHRF1 is a sensor for DNA interstrand crosslinks and recruits FANCD2 to initiate the Fanconi anemia pathway. *Cell reports*, 10(12), 1947-1956.
- Litman, R., Peng, M., Jin, Z., Zhang, F., Zhang, J., Powell, S., ... & Cantor, S. B. (2005). BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCI. *Cancer cell*, 8(3), 255-265.
- Liu, N., Lee, C. H., Swigut, T., Grow, E., Gu, B., Bassik, M. C., & Wysocka, J. (2018). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature*, 553(7687), 228-232.
- Lloyd, R. S. (2005). Investigations of pyrimidine dimer glycosylases—a paradigm for DNA base excision repair enzymology. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 577(1-2), 77-91.
- Loeb, D. D., Padgett, R. W., Hardies, S. C., Shehee, W. R., Comer, M. B., Edgell, M. H., & Hutchison 3rd, C. A. (1986). The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Molecular and Cellular Biology*, 6(1), 168-182.
- Luan, D. D., Korman, M. H., Jakubczak, J. L., & Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, 72(4), 595-605.
- Lutz, S. M., Vincent, B. J., Kazazian Jr, H. H., Batzer, M. A., & Moran, J. V. (2003). Allelic heterogeneity in LINE-1 retrotransposition activity. *The American Journal of Human Genetics*, 73(6), 1431-1437.
- Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, 19(2), 81-92.
- Machida, Y. J., Machida, Y., Chen, Y., Gurtan, A. M., Kupfer, G. M., D'Andrea, A. D., & Dutta, A. (2006). UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation. *Molecular cell*, 23(4), 589-596.
- Macias, S., Plass, M., Stajuda, A., Michlewski, G., Eyra, E., & Cáceres, J. F. (2012). DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nature structural & molecular biology*, 19(8), 760-766.
- Malik, H. S., Burke, W. D., & Eickbush, T. H. (1999). The age and evolution of non-LTR retrotransposable elements. *Molecular biology and evolution*, 16(6), 793-805.
- María Martín-Núñez, G., Rubio-Martín, E., Cabrera-Mulero, R., Rojo-Martínez, G., Oliveira, G., Valdés, S., ... & Morcillo, S. (2014). Type 2 diabetes mellitus in relation to global LINE-1 DNA methylation in peripheral blood: a cohort study. *Epigenetics*, 9(10), 1322-1328.

- Martin, S. L. (1991). Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Molecular and cellular biology*, 11(9), 4804-4807.
- Martin, S. L., Branciforte, D., Keller, D., & Bain, D. L. (2003). Trimeric structure for an essential protein in L1 retrotransposition. *Proceedings of the National Academy of Sciences*, 100(24), 13815-13820.
- Martin, S. L., Cruceanu, M., Branciforte, D., Li, P. W. L., Kwok, S. C., Hodges, R. S., & Williams, M. C. (2005). LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *Journal of molecular biology*, 348(3), 549-561.
- Martin, S. L. (2006). The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *Journal of Biomedicine and Biotechnology*, 2006.
- Mathias, S. L., Scott, A. F., Kazazian Jr, H. H., Boeke, J. D., & Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. *Science*, 254(5039), 1808-1810.
- McMillan, J. P., & Singer, M. F. (1993). Translation of the human LINE-1 element, L1Hs. *Proceedings of the National Academy of Sciences*, 90(24), 11533-11537
- McNamara, C., Zinkernagel, A. S., Macheboeuf, P., Cunningham, M. W., Nizet, V., & Ghosh, P. (2008). Coiled-coil irregularities and instabilities in group A Streptococcus M1 are required for virulence. *Science*, 319(5868), 1405-1408.
- Meetei, A. R., De Winter, J. P., Medhurst, A. L., Wallisch, M., Waisfisz, Q., Van de Vrugt, H. J., ... & Wang, W. (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nature genetics*, 35(2), 165-170.
- Meetei, A. R., Levitus, M., Xue, Y., Medhurst, A. L., Zwaan, M., Ling, C., ... & Joenje, H. (2004). X-linked inheritance of Fanconi anemia complementation group B. *Nature genetics*, 36(11), 1219-1224.
- Meischi, C., de Boer, M., Åhlin, A., & Roos, D. (2000). A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. *European Journal of Human Genetics*, 8(9), 697-703.
- Messerschmidt, D. M., Knowles, B. B., & Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes & development*, 28(8), 812-828.
- Michl, J., Zimmer, J., & Tarsounas, M. (2016). Interplay between Fanconi anemia and homologous recombination pathways in genome integrity. *The EMBO journal*, 35(9), 909-923.
- Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., ... & Nakamura, Y. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer research*, 52(3), 643-645.
- Minion, L. E., Dolinsky, J. S., Chase, D. M., Dunlop, C. L., Chao, E. C., & Monk, B. J. (2015). Hereditary predisposition to ovarian cancer, looking beyond BRCA1/BRCA2. *Gynecologic oncology*, 137(1), 86-92.

- Mita, P., Sun, X., Fenyő, D., Kahler, D. J., Li, D., Agmon, N., ... & Boeke, J. D. (2020). BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nature structural & molecular biology*, 27(2), 179-191.
- Mita, P., Wudzinska, A., Sun, X., Andrade, J., Nayak, S., Kahler, D. J., ... & Boeke, J. D. (2018). LINE-1 protein localization and functional dynamics during the cell cycle. *Elife*, 7, e30058.
- Moldovan, J. B., & Moran, J. V. (2015). The zinc-finger antiviral protein ZAP inhibits LINE and Alu retrotransposition. *PLoS genetics*, 11(5), e1005121.
- Monot, C., Kuciak, M., Viollet, S., Mir, A. A., Gabus, C., Darlix, J. L., & Cristofari, G. (2013). The specificity and flexibility of l1 reverse transcription priming at imperfect T-tracts. *PLoS genetics*, 9(5), e1003499.
- Moran, J. V., DeBerardinis, R. J., & Kazazian Jr, H. H. (1999). Exon shuffling by L1 retrotransposition. *Science*, 283(5407), 1530-1534.
- Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D., & Kazazian Jr, H. H. (1996). High frequency retrotransposition in cultured mammalian cells. *Cell*, 87(5), 917-927.
- Morell Hita, S. (2015). *Papel del elemento Line-1 en la patología humana anemia de Fanconi*. Universidad de Granada.
- Moreno, O. M., Paredes, A. C., Suarez-Obando, F., & Rojas, A. (2021). An update on Fanconi anemia: Clinical, cytogenetic and molecular approaches. *Biomedical reports*, 15(3), 1-10.
- Morgan, N. V., Tipping, A. J., Joenje, H., & Mathew, C. G. (1999). High frequency of large intragenic deletions in the Fanconi anemia group A gene. *The American Journal of Human Genetics*, 65(5), 1330-1341.
- Morisada, N., Rendtorff, N. D., Nozu, K., Morishita, T., Miyakawa, T., Matsumoto, T., ... & Kusuvara, K. (2010). Branchio-oto-renal syndrome caused by partial EYA1 deletion due to LINE-1 insertion. *Pediatric Nephrology*, 25, 1343-1348.
- Moro, F., Ottaggio, L., Bonatti, S., Simili, M., Miele, M., Bozzo, S., & Abbondandolo, J. J. (1995). p53 expression in normal versus transformed mammalian cells. *Carcinogenesis*, 16(10), 2435-2440.
- Morrish, T. A., Gilbert, N., Myers, J. S., Vincent, B. J., Stamato, T. D., Taccioli, G. E., ... & Moran, J. V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature genetics*, 31(2), 159-165.
- 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), 208-212.
- 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), 259-263.
- 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), 903-910.

- Narita, N., Nishio, H., Kitoh, 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. *The Journal of clinical investigation*, 91(5), 1862-1867.
- Neveling, K., Endt, D., Hoehn, H., & Schindler, D. (2009). Genotype–phenotype correlations in Fanconi anemia. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 668(1-2), 73-91.
- Newkirk, S. J., Lee, S., Grandi, F. C., Gaysinskaya, V., Rosser, J. M., Vanden Berg, N., ... & An, W. (2017). Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proceedings of the National Academy of Sciences*, 114(28), E5635-E5644.
- Nolan, M., Knudson, K., Holz, M. K., & Chaudhury, I. (2021). Fanconi anemia and mTOR pathways functionally interact during stalled replication fork recovery. *FEBS letters*, 595(5), 595-603.
- Oakes, C. C., La Salle, S., Smiraglia, D. J., Robaire, B., & Trasler, J. M. (2007). Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells. *Developmental biology*, 307(2), 368-379.
- Okamoto, Y., Hejna, J., & Takata, M. (2019). Regulation of R-loops and genome instability in Fanconi anemia. *The journal of biochemistry*, 165(6), 465-470.
- Orvedahl, A., MacPherson, S., Sumpter Jr, R., Tallóczy, Z., Zou, Z., & Levine, B. (2010). Autophagy protects against Sindbis virus infection of the central nervous system. *Cell host & microbe*, 7(2), 115-127.
- Orvedahl, A., Sumpter Jr, R., Xiao, G., Ng, A., Zou, Z., Tang, Y., ... & Levine, B. (2011). Image-based genome-wide siRNA screen identifies selective autophagy factors. *Nature*, 480(7375), 113-117.
- Ostertag, E. M., Goodier, J. L., Zhang, Y., & Kazazian Jr, H. H. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. *The American Journal of Human Genetics*, 73(6), 1444-1451.
- Overkamp, W. J. I., Rooimans, M. A., Neuteboom, I., Telleman, P., Arwert, F., & Zdzienicka, M. Z. (1993). Genetic diversity of mitomycin C-hypersensitive Chinese hamster cell mutants: a new complementation group with chromosomal instability. *Somatic cell and molecular genetics*, 19, 431-437.
- Pace, J. K., & Feschotte, C. (2007). The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome research*, 17(4), 422-432.
- Pang, Q., Fagerlie, S., Christianson, T. A., Keeble, W., Faulkner, G., Diaz, J., ... & Bagby, G. C. (2000). The Fanconi anemia protein FANCC binds to and facilitates the activation of STAT1 by gamma interferon and hematopoietic growth factors. *Molecular and cellular biology*, 20(13), 4724-4735.
- Pang, Q., Christianson, T. A., Keeble, W., Koretsky, T., & Bagby, G. C. (2002). The anti-apoptotic function of Hsp70 in the interferon-inducible double-stranded RNA-dependent protein kinase-mediated death signaling pathway requires the Fanconi anemia protein, FANCC. *Journal of Biological Chemistry*, 277(51), 49638-49643.

- Park, O. H., Ha, H., Lee, Y., Boo, S. H., Kwon, D. H., Song, H. K., & Kim, Y. K. (2019). Endoribonucleolytic cleavage of m6A-containing RNAs by RNase P/MRP complex. *Molecular cell*, 74(3), 494-507.
- Peng, M., Cong, K., Panzarino, N. J., Nayak, S., Calvo, J., Deng, B., ... & Cantor, S. B. (2018). Opposing roles of FANCI and HLTIF protect forks and restrain replication during stress. *Cell reports*, 24(12), 3251-3261.
- Pergola, F., Zdzienicka, M. Z., & Lieber, M. R. (1993). V (D) J recombination in mammalian cell mutants defective in DNA double-strand break repair. *Molecular and cellular biology*, 13(6), 3464-3471.
- Pezic, D., Manakov, S. A., Sachidanandam, R., & Aravin, A. A. (2014). piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes & development*, 28(13), 1410-1428.
- Pickeral, O. K., Makiłowski, W., Boguski, M. S., & Boeke, J. D. (2000). Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome research*, 10(4), 411-415.
- Pierce, A. J., Stark, J. M., Araujo, F. D., Moynahan, M. E., Berwick, M., & Jasin, M. (2001). Double-strand breaks and tumorigenesis. *Trends in cell biology*, 11, S52-S59.
- Piskareva, O., Ernst, C., Higgins, N., & Schmatchenko, V. (2013). The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding. *FEBS open bio*, 3, 433-437.
- Priestley, A., Beamish, H. J., Singleton, B. K., Blunt, T., Jeggo, P. A., Gell, D., ... & Schalkwyk, L. C. (1998). Molecular and biochemical characterisation of DNA-dependent protein kinase-defective rodent mutant irs-20. *Nucleic acids research*, 26(8), 1965-1973.
- Qian, Y., Mancini-DiNardo, D., Judkins, T., Cox, H. C., Daniels, C., Holladay, J., ... & Roa, B. B. (2015). Identification of retrotransposon insertion mutations in hereditary cancer. In *65th annual meeting of the American Society of Human Genetics*.
- Rappsilber, J., Ishihama, Y., & Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical chemistry*, 75(3), 663-670.
- Renaud, E., Barascu, A., & Rosselli, F. (2016). Impaired TIP60-mediated H4K16 acetylation accounts for the aberrant chromatin accumulation of 53BP1 and RAP80 in Fanconi anemia pathway-deficient cells. *Nucleic Acids Research*, 44(2), 648-656.
- Renaud, E., & Rosselli, F. (2013). FANCI pathway promotes UV-induced stalled replication forks recovery by acting both upstream and downstream Polη and Rev1. *PLoS One*, 8(1), e53693.
- Renaudin, X., & Rosselli, F. (2020). The FANCI/BRCA pathway releases replication blockades by eliminating DNA interstrand cross-links. *Genes*, 11(5), 585.
- Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M. S., ... & Burns, K. H. (2014). Long interspersed element-1 protein expression is a hallmark of many human cancers. *The American journal of pathology*, 184(5), 1280-1286.

- Roques, C., Coulombe, Y., Delannoy, M., Vignard, J., Grossi, S., Brodeur, I., ... & Masson, J. Y. (2009). MRE11–RAD50–NBS1 is a critical regulator of FANCD2 stability and function during DNA double-strand break repair. *The EMBO journal*, 28(16), 2400-2413.
- Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., ... & Trono, D. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, 463(7278), 237-240.
- Rutherford, T., Clegg, J. B., Higgs, D. R., Jones, R. W., Thompson, J., & Weatherall, D. J. (1981). Embryonic erythroid differentiation in the human leukemic cell line K562. *Proceedings of the National Academy of Sciences*, 78(1), 348-352.
- Sassaman, D. M., Dombroski, B. A., Moran, J. V., Kimberland, M. L., Naas, T. P., DeBerardinis, R. J., ... & Kazazian Jr, H. H. (1997). Many human L1 elements are capable of retrotransposition. *Nature genetics*, 16(1), 37-43.
- Schindler, D., & Hoehn, H. (1988). Fanconi anemia mutation causes cellular susceptibility to ambient oxygen. *American journal of human genetics*, 43(4), 429.
- Schlacher, K., Wu, H., & Jasin, M. (2012). A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer cell*, 22(1), 106-116.
- Schwab, R. A., Nieminuszczy, J., Shah, F., Langton, J., Martinez, D. L., Liang, C. C., ... & Niedzwiedz, W. (2015). The Fanconi anemia pathway maintains genome stability by coordinating replication and transcription. *Molecular cell*, 60(3), 351-361.
- Sen, S. K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P. A., ... & Batzer, M. A. (2006). Human genomic deletions mediated by recombination between Alu elements. *The American Journal of Human Genetics*, 79(1), 41-53.
- 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 Research*, 35(11), 3741-3751.
- Servant, G., Strevi, V. A., Derbes, R. S., Wijetunge, M. I., Neeland, M., White, T. B., ... & Deininger, P. L. (2017). The nucleotide excision repair pathway limits L1 retrotransposition. *Genetics*, 205(1), 139-153.
- Sijen, T., & Plasterk, R. H. (2003). Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature*, 426(6964), 310-314.
- Smogorzewska, A., Matsuoka, S., Vinciguerra, P., McDonald III, E. R., Hurov, K. E., Luo, J., ... & Elledge, S. J. (2007). Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell*, 129(2), 289-301.
- Solomon, P. J., Margaret, P., Rajendran, R., Ramalingam, R., Menezes, G. A., Shirley, A. S., ... & Seo, S. H. (2015). A case report and literature review of Fanconi Anemia (FA) diagnosed by genetic testing. *Italian journal of pediatrics*, 41, 1-8.
- Solyom, S., Ewing, A. D., Hancks, D. C., Takeshima, Y., Awano, H., Matsuo, M., & Kazazian Jr, H. H. (2012). Pathogenic orphan transduction created by a nonreference LINE-1 retrotransposon. *Human mutation*, 33(2), 369-371.



- Shi, H., Wang, X., Lu, Z., Zhao, B. S., Ma, H., Hsu, P. J., ... & He, C. (2017). YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell research*, 27(3), 315-328.
- Stamato, T. D., Weinstein, R., Giaccia, A., & Mackenzie, L. (1983). Isolation of cell cycle-dependent gamma ray-sensitive Chinese hamster ovary cell. *Somatic Cell Genetics*, 9, 165-173.
- Stojic, L., Brun, R., & Jiricny, J. (2004). Mismatch repair and DNA damage signalling. *DNA repair*, 3(8-9), 1091-1101.
- Streva, V. A., Jordan, V. E., Linker, S., Hedges, D. J., Batzer, M. A., & Deininger, P. L. (2015). Sequencing, identification and mapping of primed L1 elements (SIMPLE) reveals significant variation in full length L1 elements between individuals. *BMC genomics*, 16(1), 1-12.
- Sumpter Jr, R., & Levine, B. (2017). Emerging functions of the Fanconi anemia pathway at a glance. *Journal of cell science*, 130(16), 2657-2662.
- Sutton, R. B., Fasshauer, D., Jahn, R., & Brunger, A. T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature*, 395(6700), 347-353.
- Suzuki, J., Yamaguchi, K., Kajikawa, M., Ichiyanagi, K., Adachi, N., Koyama, H., ... & Okada, N. (2009). Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS genetics*, 5(4), e1000461.
- Swergold, G. D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. *Molecular and cellular biology*, 10(12), 6718-6729.
- Swuec, P., Renault, L., Borg, A., Shah, F., Murphy, V. J., van Twest, S., ... & Costa, A. (2017). The FA core complex contains a homo-dimeric catalytic module for the symmetric mono-ubiquitination of FANCI-FANCD2. *Cell reports*, 18(3), 611-623.
- Symer, D. E., Connelly, C., Szak, S. T., Caputo, E. M., Cost, G. J., Parmigiani, G., & Boeke, J. D. (2002). Human I1 retrotransposition is associated with genetic instability in vivo. *Cell*, 110(3), 327-338.
- 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 biology*, 3(10), 1-18.
- Taniguchi, T., Garcia-Higuera, I., Xu, B., Andreassen, P. R., Gregory, R. C., Kim, S. T., ... & D'Andrea, A. D. (2002). Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. *Cell*, 109(4), 459-472.
- Taylor, M. S., LaCava, J., Mita, P., Molloy, K. R., Huang, C. R. L., Li, D., ... & Dai, L. (2013). Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell*, 155(5), 1034-1048.
- Taylor, M. S., LaCava, J., Dai, L., Mita, P., Burns, K. H., Rout, M. P., & Boeke, J. D. (2016). Characterization of L1-ribonucleoprotein particles. *Transposons and Retrotransposons: Methods and Protocols*, 311-338.
- Taylor, M. S., Altukhov, I., Molloy, K. R., Mita, P., Jiang, H., Adney, E. M., ... & LaCava, J. (2018). Dissection of affinity captured LINE-1 macromolecular complexes. *elife*, 7, e30094.

- Tchénio, T., Casella, J. F., & Heidmann, T. (2000). Members of the SRY family regulate the human LINE retrotransposons. *Nucleic acids research*, 28(2), 411-415.
- Telleman, P., Overkamp, W. J. I., & Zdzienicka, M. Z. (1996). Spectrum of spontaneously occurring mutations in the HPRT gene of the Chinese hamster V79 cell mutant V-H4, which is homologous to Fanconi anemia group A. *Mutagenesis*, 11(2), 155-159.
- Teng, S. C., Kim, B., & Gabriel, A. (1996). Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature*, 383(6601), 641-644.
- Teugels, E., De Brakeleer, S., Goelen, G., Lissens, W., Sermijn, E., & De Grève, J. (2005). De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes. *Human mutation*, 26(3), 284-284.
- Tonzi, P., Yin, Y., Lee, C. W. T., Rothenberg, E., & Huang, T. T. (2018). Translesion polymerase kappa-dependent DNA synthesis underlies replication fork recovery. *elife*, 7, e41426.
- Tremblay, A., Jasin, M., & Chartrand, P. (2000). A double-strand break in a chromosomal LINE element can be repaired by gene conversion with various endogenous LINE elements in mouse cells. *Molecular and cellular biology*, 20(1), 54-60.
- Trelogan, S. A., & Martin, S. L. (1995). Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proceedings of the National Academy of Sciences*, 92(5), 1520-1524.
- Tristán-Ramos, P., Rubio-Roldan, A., Peris, G., Sánchez, L., Amador-Cubero, S., Viollet, S., ... & Heras, S. R. (2020). The tumor suppressor microRNA let-7 inhibits human LINE-1 retrotransposition. *Nature Communications*, 11(1), 5712.
- Tyanova, S., Temu, T., & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols*, 11(12), 2301-2319.
- Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., & Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, 313(5785), 320-324.
- Van Den Hurk, JAJM, Van De Pol, DJR, Wissinger, B, Van Driel MA, Hoefsloot LH, De Wijs IJ et al (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):268–275
- Van Den Hurk, J. A., Meij, I. C., del Carmen Seleme, M., Kano, H., Nikopoulos, K., Hoefsloot, L. H., ... & Cremers, F. P. (2007). L1 retrotransposition can occur early in human embryonic development. *Human molecular genetics*, 16(13), 1587-1592.
- Voliva, C. F., Martin, S. L., Hutchison III, C. A., & Edgell, M. H. (1984). Dispersal process associated with the L1 family of interspersed repetitive DNA sequences. *Journal of molecular biology*, 178(4), 795-813.
- Vuong, L. M., Pan, S., & Donovan, P. J. (2019). Proteome Profile of Endogenous Retrotransposon-Associated Complexes in Human Embryonic Stem Cells. *Proteomics*, 19(15), 1900169.

- Walden, H., & Deans, A. J. (2014). The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annual review of biophysics*, 43, 257-278.
- 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.
- Wang, L., Norris, E. T., & Jordan, I. K. (2017). Human retrotransposon insertion polymorphisms are associated with health and disease via gene regulatory phenotypes. *Frontiers in microbiology*, 8, 1418.
- Wanichnopparat, W., Suwanwongse, K., Pin-On, P., Apornthewan, C., & Mutirangura, A. (2013). Genes associated with the cis-regulatory functions of intragenic LINE-1 elements. *BMC genomics*, 14(1), 1-9.
- Wang, L. C., Stone, S., Hoatlin, M. E., & Gautier, J. (2008). Fanconi anemia proteins stabilize replication forks. *DNA repair*, 7(12), 1973-1981.
- Wang, J. Y., & Lu, A. Q. (2021). The biological function of m6A reader YTHDF2 and its role in human disease. *Cancer cell international*, 21(1), 109.
- Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., ... & Sasaki, H. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, 453(7194), 539-543.
- Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., ... & Moran, J. V. (2001). Human L1 retrotransposition: cis preference versus trans complementation. *Molecular and cellular biology*, 21(4), 1429-1439.
- Wellinger, R. E., Prado, F., & Aguilera, A. (2006). Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Molecular and cellular biology*, 26(8), 3327-3334.
- Wheelan, S. J., Aizawa, Y., Han, J. S., & Boeke, J. D. (2005). Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome research*, 15(8), 1073-1078.
- Williams, Z., Morozov, P., Mihailovic, A., Lin, C., Puvvula, P. K., Juranek, S., ... & Tuschl, T. (2015). Discovery and characterization of piRNAs in the human fetal ovary. *Cell Reports*, 13(4), 854-863.
- Wu, X., Lu, Y., Ding, Q., You, G., Dai, J., Xi, X., ... & Wang, X. (2014). Characterisation of large F9 deletions in seven unrelated patients with severe haemophilia B. *Thrombosis and haemostasis*, 112(09), 459-465.
- Xiong, F., Wang, R., Lee, J. H., Li, S., Chen, S. F., Liao, Z., ... & Li, W. (2021). RNA m6A modification orchestrates a LINE-1–host interaction that facilitates retrotransposition and contributes to long gene vulnerability. *Cell Research*, 31(8), 861-885.
- Xing, Z., Ma, W. K., & Tran, E. J. (2019). The DDX5/Dbp2 subfamily of DEAD-box RNA helicases. *Wiley Interdisciplinary Reviews: RNA*, 10(2), e1519.

- Xing, J., Wang, H., Belancio, V. P., Cordaux, R., Deininger, P. L., & Batzer, M. A. (2006). Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proceedings of the National Academy of Sciences*, 103(47), 17608-17613.
- Xing, J., Zhang, Y., Han, K., Salem, A. H., Sen, S. K., Huff, C. D., ... & Jorde, L. B. (2009). Mobile elements create structural variation: analysis of a complete human genome. *Genome research*, 19(9), 1516-1526.
- Yamashita, T., Wu, N., Kupfer, G., Corless, C., Joenje, H., Grompe, M., & D'Andrea, A. D. (1996). Clinical variability of Fanconi anemia (type C) results from expression of an amino terminal truncated Fanconi anemia complementation group C polypeptide with partial activity.
- Yan, P., Liu, Z., Song, M., Wu, Z., Xu, W., Li, K., ... & Liu, G. H. (2020). Genome-wide R-loop landscapes during cell differentiation and reprogramming. *Cell reports*, 32(1), 107870.
- Yuan, F., Qian, L., Zhao, X., Liu, J. Y., Song, L., D'Urso, G., ... & Zhang, Y. (2012). Fanconi anemia complementation group A (FANCA) protein has intrinsic affinity for nucleic acids with preference for single-stranded forms. *Journal of Biological Chemistry*, 287(7), 4800-4807.
- Zdzienicka, M. Z., Arwert, F., Neuteboom, I., Rooimans, M., & Simons, J. W. I. M. (1990). The Chinese hamster V79 cell mutant V-H4 is phenotypically like Fanconi anemia cells. *Somatic cell and molecular genetics*, 16, 575-581.
- Zellweger, R., Dalcher, D., Mutreja, K., Berti, M., Schmid, J. A., Herrador, R., ... & Lopes, M. (2015). Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. *Journal of Cell Biology*, 208(5), 563-579.
- Zhao, B., Wu, Q., Ye, A. Y., Guo, J., Zheng, X., Yang, X., ... & Huang, A. Y. (2019). Somatic LINE-1 retrotransposition in cortical neurons and non-brain tissues of Rett patients and healthy individuals. *PLoS genetics*, 15(4), e1008043.
- Zhang, X., Zhang, R., & Yu, J. (2020). New understanding of the relevant role of LINE-1 retrotransposition in human disease and immune modulation. *Frontiers in cell and developmental biology*, 8, 657

