Retrotransposons in chronic pain: focusing on roles in gene regulation and neuroinflammation

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Emma Price MBiolSci

December 2019

Acknowledgements

The biggest thank you goes to my PhD supervisors, Prof John Quinn and Dr Jill Bubb. This PhD has taught me to have confidence in my own ideas and to be resilient. Your support and guidance, both professional and personal, has helped me beyond words and I can't thank you enough. Thanks also go to my supervisor Dr Bernhard Frank for his clinical insight and contribution to the project.

Thank you to everybody who has contributed to my training and provided resources to enable the success of this work including; Zsuzsanna Helyes, Eva Szoke, Maja Payrits, Angela Kecskes, József Kun, Lynn McLaughlin, Anne McArdle, Sarah Roper, staff at the Biomedical Services Unit, Neil Pendleton, Antony Payton, Daniel Carr, Felicity Lumb, Jenny Crowe, Margaret Harnett and Paddy Harrison. Thanks also go to others who made this work possible and kept us happy in the lab, including technical staff, cleaning staff, building managers and of course the staff in Gregg's. Big thanks also go to all my other previous teachers, tutors and mentors, especially Luciane Mello, who played their part in getting me here.

On a personal note, a huge thanks goes to all my colleagues in Quinn lab, past and present, but especially my lab family; Kimberley Billingsley, Jack Marshall, Olympia Gianfrancesco, Ben Middlehurst, Ana Illera, Ashley Hall and Li Li. I appreciate you all so much for the help with experiments, writing up, and emergency trips to the AJ. Jack, you have made me smile every single day, helped me through the hardest times and this PhD wouldn't have been as half as fun if it wasn't for you. Kim, you inspire me all the time and have given me the courage to go for it! And now for my amazing family… To my Nan and Grandad, thanks for letting me stay up late and watch forensic science documentaries. Without that, I wouldn't have become obsessed with DNA. To Gran and Grandad, thanks for your constant encouragement over the years, I hope I've done you proud. And to the rest of my wonderful family… Mum, thanks for being my listener of new theories and number one fan! Dad, for being my absolute rock. Claire, for the drive you have given me to succeed. And to my sibs; Jack, Lara, Megan and Libby – thanks for supporting me always, helping me achieve my goals, and keeping me sane.

Finally, thank you to the Pain Relief Foundation (registered charity number: 1156227) who funded this research.

For my Aunty Ness and my Granny Anne, who were and will forever be the most inspirational people in my life.

Abstract

Pain has evolved as a protective mechanism to signal danger, threat, injury and inflammation to an organism; however, this becomes problematic and detrimental to human physiological and psychological wellbeing when it persists without a noxious cause. Preclinical rodent models are frequently and successfully used to identify the molecular components of pain however novel therapeutics that are developed from these findings often fail to translate into human clinical intervention. Recent RNA-seq studies have identified transcriptomic differences in the dorsal root ganglia (DRG) of humans and mice that likely contribute to this lack of translational research. Transposable elements (TEs) are functional components of genomes that contribute to species specific gene expression patterns; therefore, we explored the role of TEs in the context of pain gene expression. This work focuses on non-LTR retrotransposons as they remain the major active class of autonomous TEs in the human genome.

We first present a proof of principle study using CRISPR that demonstrates the effect of a human specific retrotransposon SINE-VNTR-*Alu* (SVA) insertion as a contributor to human specific gene regulation at the locus encoding pain genes *TRPV1* and *TRPV3.* TEs also supply a vast source of genetic variation to individual genomes. Chronic pain is difficult to manage due to its heterogeneity in underlying cause, severity and response to therapeutics. Modern genetic approaches such as genome-wide association studies and candidate gene studies often do not identify causal variants responsible for these interpersonal differences. Here, we show that the SVA at *TRPV1* and *TRPV3* has regulatory properties and we further demonstrate its polymorphic nature and highlight potential links to pain phenotypes. An improved understanding of mechanisms that contribute to the development and maintenance of chronic pain are required to address the ever-increasing socio-economic burden it creates, which is amplified by the increasing ageing population and a lack of effective therapeutics.

Recent work has identified that expression of the retrotransposon LINE-1 (L1) is a pathogenic contributor to neurodegenerative disease in the central nervous system and an inducer of inflammation in non-pathological ageing. Chronic pain is an age associated disorder and is driven by the onset of neuroinflammation in the peripheral nervous system. To assess whether L1 expression contributes to this mechanism associated with chronic pain, we present another proof of principle experiment in which we demonstrate increased L1 expression in ageing DRG. Furthermore, we explored L1 expression in response to a novel anti-inflammatory drug and found unexpected increases in expression. The data from this study implies a strong case for further investigation and a potential role in neuroinflammation in the peripheral nervous system.

To summarise, the work presented in this thesis describes the first studies to investigate the role of retrotransposons in several contexts that may contribute to the development of chronic pain and modulation of pain phenotype. This provides the basis for a wide range of future studies that can further explore the role of retrotransposons in chronic pain which could aid our understanding of species differences, age associated molecular changes and provide an alternative route to develop novel therapeutics.

Abbreviations

ADRs	adverse drug reaction
AGS	Aicardi-Goutières syndrome
ALS	amyotrophic lateral sclerosis
ASD	autism spectrum disorder
CHO	Chinese hamster ovary
CIP	congenital insensitivity to pain
CNS	central nervous system
CRE	*cis*-regulatory element
DNMT1	DNA methyltransferase
DRG	dorsal root ganglia
ERV	endogenous retrovirus
ETn	early transposons
gDNA	genomic DNA
GWAS	genome wide association study
HERV	human endogenous retrovirus
IAP	intracisternal-A particles
IFN	interferon
KAP1	KRAB-associated protein 1
KRAB	Kruppel-associated box
LCA	last common ancestor
LD	linkage disequilibrium
LINE	long interspersed nuclear element
LTR	long terminal repeat
L1	long interspersed nuclear element 1
M1	metabolite O-desmethyltramadol
MeCP2	methyl CpG binding protein 2
MELT	mobile element locator tool
MIR	mammalian interspersed repeat
MND	motor neuron disease
MS	multiple sclerosis
Myrs	million years

NGF	nerve growth factor
NHEJ	non-homologous end joining
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PNS	peripheral nervous system
RC-seq	retrotransposon capture sequencing
RIP	retrotransposon insertion polymorphism
RNP	ribonucleoprotein particle
SEM	standard error from the mean
SINE	short interspersed nuclear element
siRNA	small interfering RNA
SLE	system lupus erythematosus
SVA	SINE-VNTR-*Alu*
TE	transposable element
TDP-43	transactive response DNA binding protein 43 kDa
TFBS	transcription factor binding site
TG	trigeminal ganglia
TIR	two inverted tandem repeats
TNFα	tumour necrosis factor α
TPRT	target primed reverse transcription
TRPV1	transient receptor potential vanilloid 1
TRPV3	transient receptor potential vanilloid 3
TSD	target site duplication
TSS	transcriptional start site
UTR	untranslated region
UV	ultraviolet
VAS	visual analogue scale
VNTR	variable number tandem repeat
XDP	X-linked dystonia parkinsonism
YFP	yellow fluorescent protein
ZFP	zinc finger protein
ZNF	zinc nuclease finger

Contents page

Chapter 1. Introduction

1.1 General introduction

1.1.1 Thesis overview

The requirement for a better understanding of chronic pain mechanisms and the development of effective and safe therapeutics is rapidly increasing due to an ageing population[1,2], the large socio-economic burden placed on the workplace/healthcare systems and increasing rates of addiction to opioid based medication[3,4]. Genetic studies have focused mainly on protein coding candidate gene studies and genome-wide association studies (GWAS)[5]. This has led to an increase in understanding of pain mechanisms and highlighted hundreds of associated genetic variants, however this has also demonstrated the heterogeneity of this condition[5]. Other attempts to identify novel pain targets have used RNA-seq based methods in the peripheral nervous system, with a focus on the dorsal root ganglia (DRG), which have highlighted species specific gene expression differences[6]. Due to a lack of translational research and heterogeneity of the disease, there has been little success in treating chronic pain in the clinic. Therefore, a better understanding of genetic and epigenetic factors that contribute to these challenges in required.

In this thesis, we address the potential role of transposable elements (TEs), specifically non-LTR retrotransposon subfamilies SVA and L1 in biological contexts relevant to pain. TEs play a fundamental role in shaping and regulating the genomes of individual species and provide a large source of genetic variation in the general population[7,8]. The role of TEs is an area of pain research that has not yet been explored. First, we focused on the role of a human specific retrotransposon insertion, termed SVA, at the genomic locus encoding pain genes *TRPV1* and *TRPV3. TRPV3* is expressed in DRG of humans but not the DRG of mice, highlighting species specific differences in gene regulation at this locus[6,9,10]. TEs, especially those that are lineage specific, can function as drivers of species-specific expression patterns; therefore, the SVA insertion was highlighted as a human specific candidate regulatory domain at the *TRPV1/TRPV3* locus. We functionally characterised this

insertion and demonstrated its regulatory properties; our results suggest it contributes to the species differences in gene expression documented in the literature. We further characterised genetic and epigenetic variation encoded within this element and revealed associations with pain phenotypes.

In contrast to their role in pain, the pathogenic role for TE activity in neurodegenerative disorders that affect the central nervous system has been established in recent years. Increased levels of retrotransposon activity, specifically increased L1 expression, has been demonstrated in ageing and as a cause of age-associated inflammation. An underlying cause of chronic pain is neuroinflammation which can occur in response to injury and trauma. Neuroinflammation results in epigenetic changes contributing to central and peripheral sensitisation, which causes the transition from acute to chronic pain. However, the prevalence of chronic pain is largely associated with increasing age. Neuroinflammation is a hallmark of ageing and therefore may be causative of spontaneous chronic pain without the requirement of nerve or tissue damage. We suspected that L1 expression also plays a fundamental role in age associated neuroinflammation in the DRG which contributes to chronic pain development. This is addressed in the latter part of the thesis.

1.1.2 Thesis structure

This body of work ultimately aims to raise the profile of TEs in the field of chronic pain research. The work undertaken in this thesis has explored several key themes (listed below) of transposable element biology that may address some of the problems currently facing research in chronic pain.

1. **TEs contribute to species specific gene regulation** – chapter 3 focuses on the role of a human specific SVA insertion at the locus encoding pain genes *TRPV1* and *TRPV3* as a candidate regulatory domain to explain species differences in gene regulation.
2. **TEs are polymorphic in the general population** – chapter 4 focuses on characterising polymorphism within the SVA at the *TRPV1* and *TRPV3* locus and explores the potential utility of SVA polymorphism as a genetic variant associated with pain phenotypes.
3. **TEs contribute to tissue specific gene regulation** – chapter 4 also focuses on profiling the epigenetic status of the SVA at *TRPV1* and *TRPV3* between individuals, and between blood and brain in the same individuals to address variability and potential tissue specific function.
4. **TE activity increases during ageing and induces inflammation** – chapter 5 focuses on profiling L1 mRNA in ageing mouse DRG, as a first look into whether L1 associated inflammation present in central nervous system (CNS) neurodegenerative disorders may also be a potential mechanism underlying age associated inflammation in the peripheral nervous system (PNS). L1 expression was further explored in the brains of mice with high fat diet (HFD) induced inflammation to assess if L1 expression could be modulated by diet with ageing. L1 expression was also measured in this HFD model in response to treatment with a novel anti-inflammatory drug ES-62 to assess if L1 could be modulated via interference with inflammatory pathways.

1.2 Retrotransposons: drivers of ageing and neuroinflammation

1.2.1 Overview of transposable elements in the human genome

Transposable elements (TEs), commonly referred to as "jumping genes", are DNA sequences that can transpose (move) to new locations in the genome[11]. Since the time of their discovery in 1956 by Barbara McClintock in maize, DNA sequencing has revealed that an estimated 45% of the human genome is derived from TE sequences[12,13] (Figure 1.1). This, however, is likely to be an underestimation, due to ancient TEs undergoing sequence divergence throughout evolution, making them difficult to identify[14]. It is predicted that the current estimate of TEs in the human genome will increase in coming years with improvements to the reference human genome and bioinformatic TE detection tools[15].

All TEs in the human genome can be classified based on their mechanism of transposition and sequence structure[16] (Figure 1.2.). Class I TEs transpose via a copy and paste mechanism involving an RNA intermediate (Figure 1.3) and are therefore referred to as retrotransposons. One subclass of retrotransposons are the LTR retrotransposons, which have acquired their name due to the presence of long terminal repeats (LTRs) at each end of their sequence[14]. LTR retrotransposons constitute approximately 8% of the human genome (Figure 1.1). LTR retrotransposons and all other transposon subclasses, can be further divided into families. The main family of LTR retrotransposons are human endogenous retroviruses (HERVs) (Figure 1.4B). HERVs have integrated into the human genome throughout evolution caused by repeated infection of germline cells by exogenous retroviruses[17]. HERVs (e.g. HERV-K) contain sequences for *gag* (group-specific antigen), *pol* (polymerase) and *env* (envelope) genes (Figure 1.4B). In modern humans, the majority of HERVs have accumulated mutations (which were acquired throughout evolution) and are now unable to replicate (transpose), thus are referred to as "fossil viruses"[17]. HERVs have recently been implicated in the pathology of autoimmune and neurodegenerative diseases including multiple sclerosis (MS)[17] and motor neuron disease (MND) (commonly referred to in the literature as

Amyotrophic Lateral Sclerosis; ALS)[18]. HERV insertional polymorphisms have been identified in the human genome, suggesting some may still undergo transposition – however the mechanism responsible has not been identified[19]. Class I TEs also consist of non-LTR retrotransposons and include the families; LINE-1, *Alu* and SVA (Figure 1.2.). Together, non-LTR retrotransposons account for approximately one third of the human genome (Figure 1.1). These elements form the basis of the work presented in this thesis therefore are discussed in detail in the next section (1.2.2).

Class II TEs are DNA transposons that transpose via a cut and paste mechanism (Figure 1.3), and include elements such as hAT, MuDR, piggyBAC and Tc1/Mariner elements (Figure 1.2.)[16]. Class II DNA transposons account for only 3% of the human genome (Figure 1.1) as they no longer actively transpose due to an accumulation of mutations acquired throughout evolution that have rendered them inactive, however some examples remain active in other organisms , e.g. the P element in *Drosophila*[20,21]. A hallmark of transposition is the generation of target site duplications (TSDs) (Figure 1.4), which are generated when target site sticky ends (created during transposition) are filled in, generating direct repeats that flank the new TE insertion (Figure 1.5)[16].

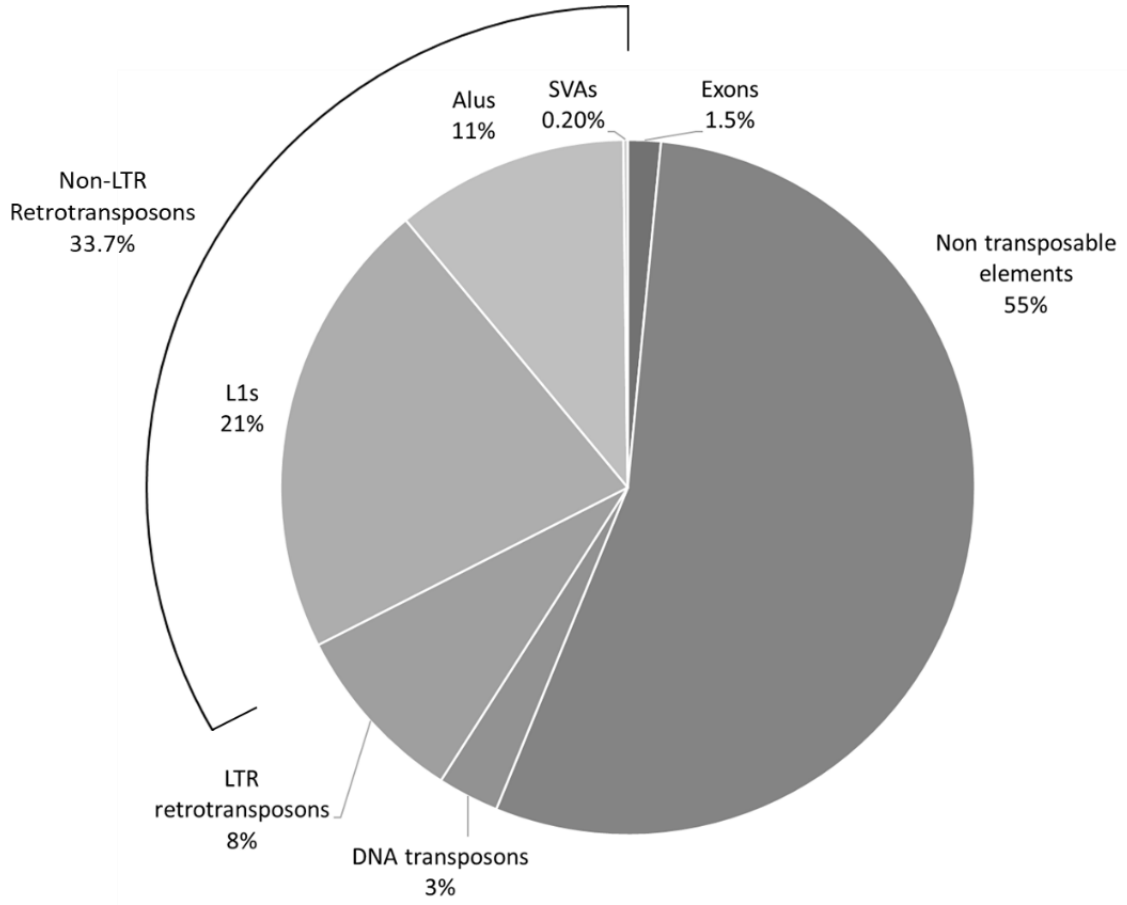


Figure 1.1. Transposable element content in the human genome. Transposable elements constitute approximately 45% of the human genome. Non LTR-retrotransposons constitute 33.7% including L1s (21%), Alus (11%) and SVAs (0.2%). Adapted from Cordeaux & Batzer[14] and Kazazian & Moran[20].

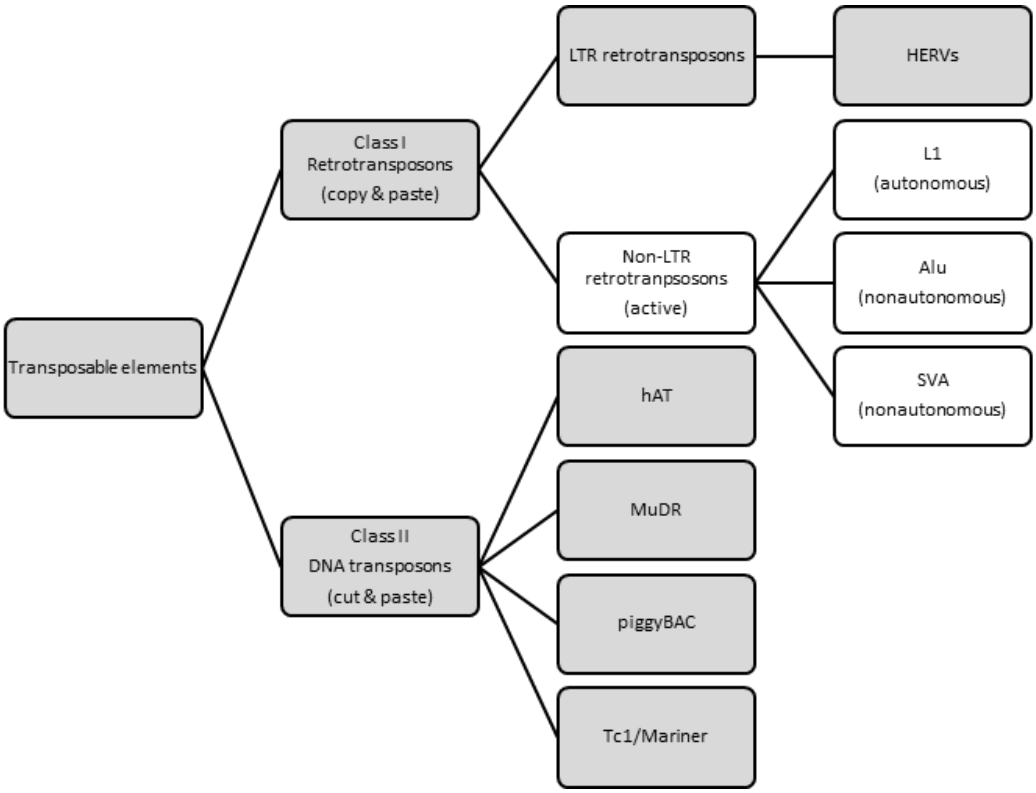


Figure 1.2. Classification of transposable elements in the human genome. Non-LTR retrotransposons including L1, Alu and SVA are still actively mobilising in the human genome via a copy and paste mechanism. L1 mobilises its own RNA in *cis* therefore classed as autonomous. Alu and SVA RNA are mobilised in *trans* via L1 machinery therefore classed as nonautonomous. Adapted from Misiak et al[16].

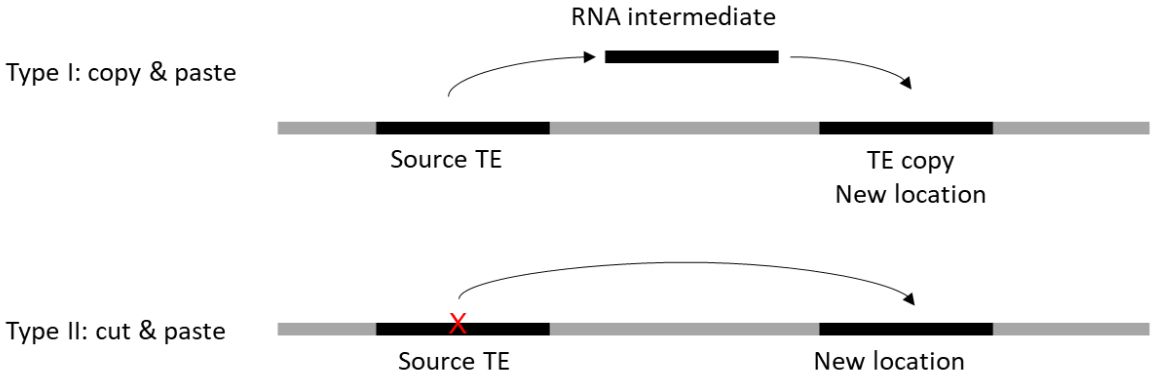


Figure 1.3. Schematic of type I and type II transposition mechanisms. (Above) Type I copy and paste mechanism. The source TE is transcribed into an RNA intermediate, which is then reverse transcribed into cDNA by a reverse transcriptase encoded by the TE sequence. The cDNA copy is then integrated back into the genome in a new location. (Below) Type II cut and paste mechanism. The source TE is excised from the genome and reintegrated into a new location.

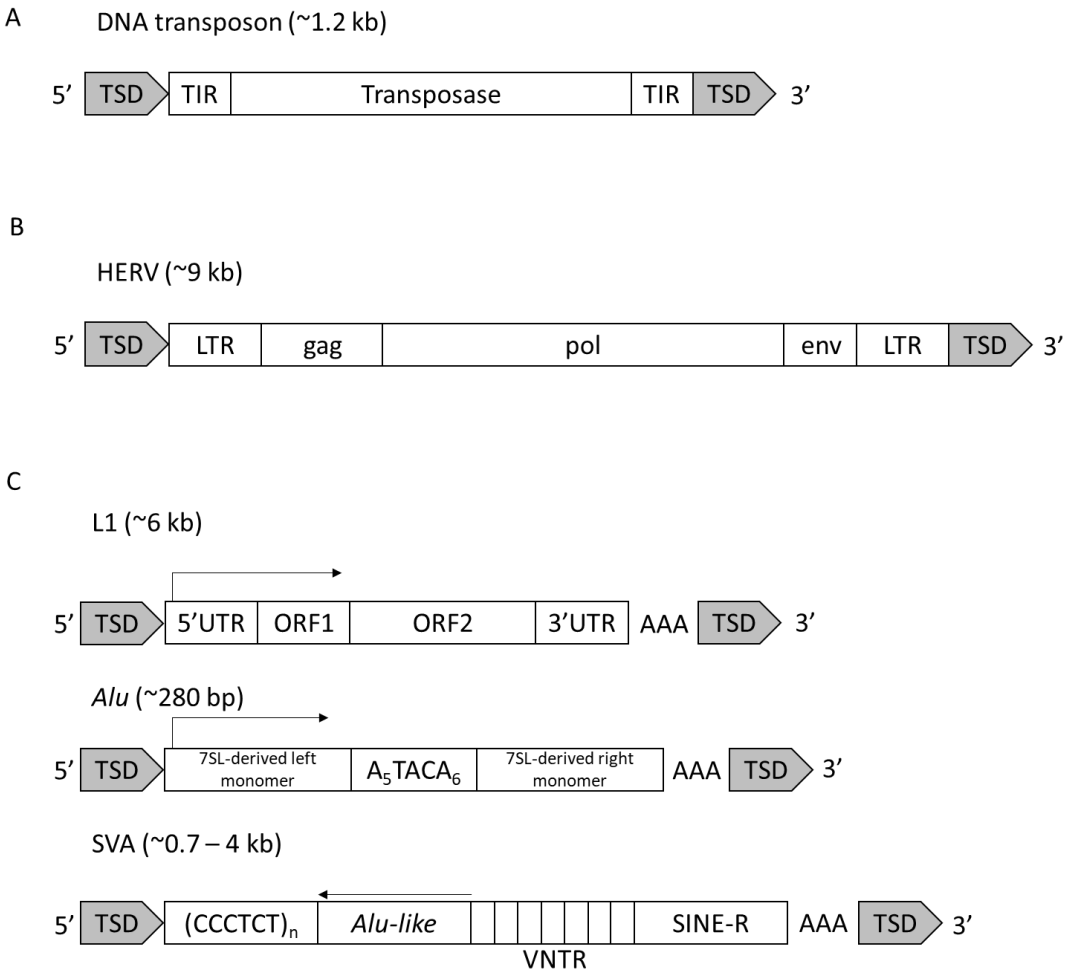


Figure 1.4. Structure of transposable elements in the human genome. (A) Example of a DNA transposon Hsmar-1 (a mariner like element) containing the sequence encoding a transposase, which is flanked by two inverted tandem repeats (TIR) and target site duplications (TSD) following transposition. (B) Example of an LTR retrotransposon; HERV, encoding *gag*, *pol* and *env*. This is flanked by LTRs and TSD. (C) Examples of non-LTR retrotransposons; L1, *Alu* and SVA. L1 contains a 5’ untranslated region (UTR) which contains an internal promoter. It encodes ORF1 and ORF2 followed by a 3’UTR and a poly A tail. This is flanked by TSD. The Alu contains two monomers derived from the 7SL RNA gene, separated by an AT-rich sequence. This contains a poly A tail at the 3’ end and is flanked by TSD. SVA contains a CT-rich domain at the 5’ end followed by an alu-like region, a variable number tandem repeat (VNTR) domain, a SINE-R region and a poly A tail at the 3’ end. This is flanked by TSD. Elements are not drawn to scale.

1.2.2 Active transposable elements: non-LTR retrotransposons

The large proportion of TEs in the human genome (approximately 45%) is mainly attributed to non-LTR retrotransposons (33.7%) (Figure 1.1)[14]. The non-LTR retrotransposon families; L1, *Alu* and SVA have proliferated and accumulated throughout the past 80 million years of primate evolution[14]. These elements are considered to be the only TEs capable of transposition in the human genome.

1.2.2.1 L1

Long interspersed nuclear element 1 (LINE-1, often further abbreviated to L1) accounts for approximately 21% of the human genome (Figure 1.1)[14]. A full length L1 element is approximately 6 kb in length and consists of a 5' untranslated region (UTR) which contains an internal promoter, two open reading frames named ORF1 and ORF2, a 3'UTR and a poly adenylation signal, followed by a poly A tail (Figure 1.4C)[22]. ORF1 encodes a 40kDa protein with RNA binding and nucleic acid chaperone activity. ORF2 encodes a 150kDa protein with endonuclease (EN) and reverse transcription (RT) activity. These proteins function as the machinery which transposes the mRNA molecule from which they were translated. This is referred to as *cis*-mobilisation.

L1 mobilisation (Figure 1.5A) occurs via a process called target-primed reverse transcription (TPRT) (Figure 1.5B)[22]. In brief, L1 is transcribed by RNA polymerase II from the promoter in the 5'UTR. The mRNA is exported into the cytoplasm were *ORF1* and *ORF2* are then translated by ribosomes. ORF1p, ORF2p and the L1 mRNA then form a ribonucleoprotein (RNP) complex (Figure 1.5A). The L1 RNP is then transported back into the nucleus where ORF2p (EN) creates a nick on a DNA strand containing the sequence 3'-AA/TTTT-5' (Figure 1.5B). This exposes a 3' hydroxyl group which enables the ORF2p (RT) to initiate TPRT from the associated L1 mRNA and reverse transcribe it into the site (Figure 1.5B). DNA synthesis on the complimentary strand completes the integration process

however the mechanisms behind this are at present unclear (Figure 1.5B). New genomic insertions that are mobilised by L1 machinery and TPRT are flanked by short genomic sequences called target site duplications (TSDs), which are carried over from the original site of the source element (Figure 1.4C).

L1 are referred to as "autonomous" as they encode the machinery necessary to retrotranspose, which is a feature unique to this retrotransposon family. There are over 500,000 copies of L1 in the genome, however the majority are neither full length or able to mobilise further, due to accumulation of mutations acquired throughout evolution, or truncations and structural rearrangements that can occur during the mobilisation process. It is estimated that there are approximately 80 – 100 full length L1 copies that are capable of facilitating mobilisation in the human genome. These are referred to as "hot L1s"[23].

1.2.2.2 Alu

Alu retrotransposons are primate specific short interspersed nuclear elements (SINEs) that account for 11% of the human genome[24] (Figure 1.1). *Alu* elements are approximately 280 bp in length and consist of two dimers derived from the 7SL RNA gene, which are separated by a short adenine rich region and contain a poly A region at the 3' end[22] (Figure 1.4C). Within the first monomer is an RNA polymerase III promoter.

Whilst they do not contribute as much DNA sequence in terms of base pairs as L1, they are the most abundant TE in the human genome by copy number (estimated >1 million copies)[24]. These elements do not encode the functional proteins required for mobilisation, instead they depend upon association with L1 encoded ORF2 in *trans* for their reverse transcription and reintegration into the genome, therefore are referred to as "nonautonomous" elements[24].

1.2.2.3 SVA

SINE-VNTR-*Alu* (SVA) retrotransposons are another primate specific non-LTR retrotransposon family (like *Alu*) (Figure 1.2)[25]. These elements account for only 0.2% of the human genome and consist of relatively few copies (estimated 3000), when compared to *Alu*. SVAs are derived from multiple composite sequences (Figure 1.4C). At the 5' end there is typically a $(CCCTCT)_n$ hexamer repeat region, an *Alu*-like region comprised of two Alu fragments in the antisense orientation, a central variable number tandem repeat (VNTR, typically 30 – 50 bp in length) with a high CG content, a SINE-R region derived from a HERV-K10 element and poly A tail at the 3' end[26] (Figure 1.4C).

SVAs do not contain an internal promoter, unlike L1 and *Alu,* however they are transcribed by RNA pol II, likely as a result from nearby promoters in flanking regions and transcriptional readthrough[25]. An average SVA is approximately 2 kb in length but their actual length can range from 700 bp to 4 kb. The majority of SVAs are full length (63%) but they can be subject to truncations, internal rearrangements and inversions, resulting from the mobilisation process. Similar to *Alu*, SVA do not encode the machinery required for its transposition, and are hypothesised to retrotranspose via L1 machinery, therefore are also referred to as "nonautonomous"[26].

Interrogation of primate whole genome sequencing data and qPCR validation using primate genomic DNA has identified SVA copies in hominoid genomes including; human, chimpanzee, gorilla and orangutan[25]. No copies have been identified in old world primate genomes e.g. rhesus macaque. As a result, SVAs are classed as hominoid-specific retrotransposons and are thus considered to be the youngest subfamily of TEs currently identified in the human genome[25]. SVAs can be further assigned different subfamilies, based on evolutionary age which is calculated by sequence divergence in the *Alu*-like region[25]. There are currently six recognised SVA subfamilies in the human genome that are designated by evolutionary age as follows; SVA A (13.56 million years; Myrs), SVA B (11.56 Myrs), SVA C

(10.88 Myrs), SVA D (9.55 Myrs), SVA E, (3.46 Myrs), SVA F (3.18 Myrs)[25]. There is also an SVA F_1 subfamily, distinguished by the inclusion of the *MAST2* exon 1, which account for 3% of all SVAs[26]. SVA subfamilies A, B and C are conserved with other primates including chimpanzee, gorilla, and orangutan. Some SVA D are conserved with chimpanzee only, and some SVA D are human specific due to continued mobilisation following divergence with the last common ancestor shared with chimpanzee. SVA E and F are human specific as these subfamilies emerged after this divergence event[15,25]. SVA D is the most frequent SVA subfamily in the human genome comprising 42.7% of all SVAs, followed by SVA B (15%), SVA F (9.5%), SVA C (7.3%), SVA A (5.8%) and SVA E (4.4%). Approximately 15% of SVAs are difficult to designate due to structural rearrangements previously discussed[25].

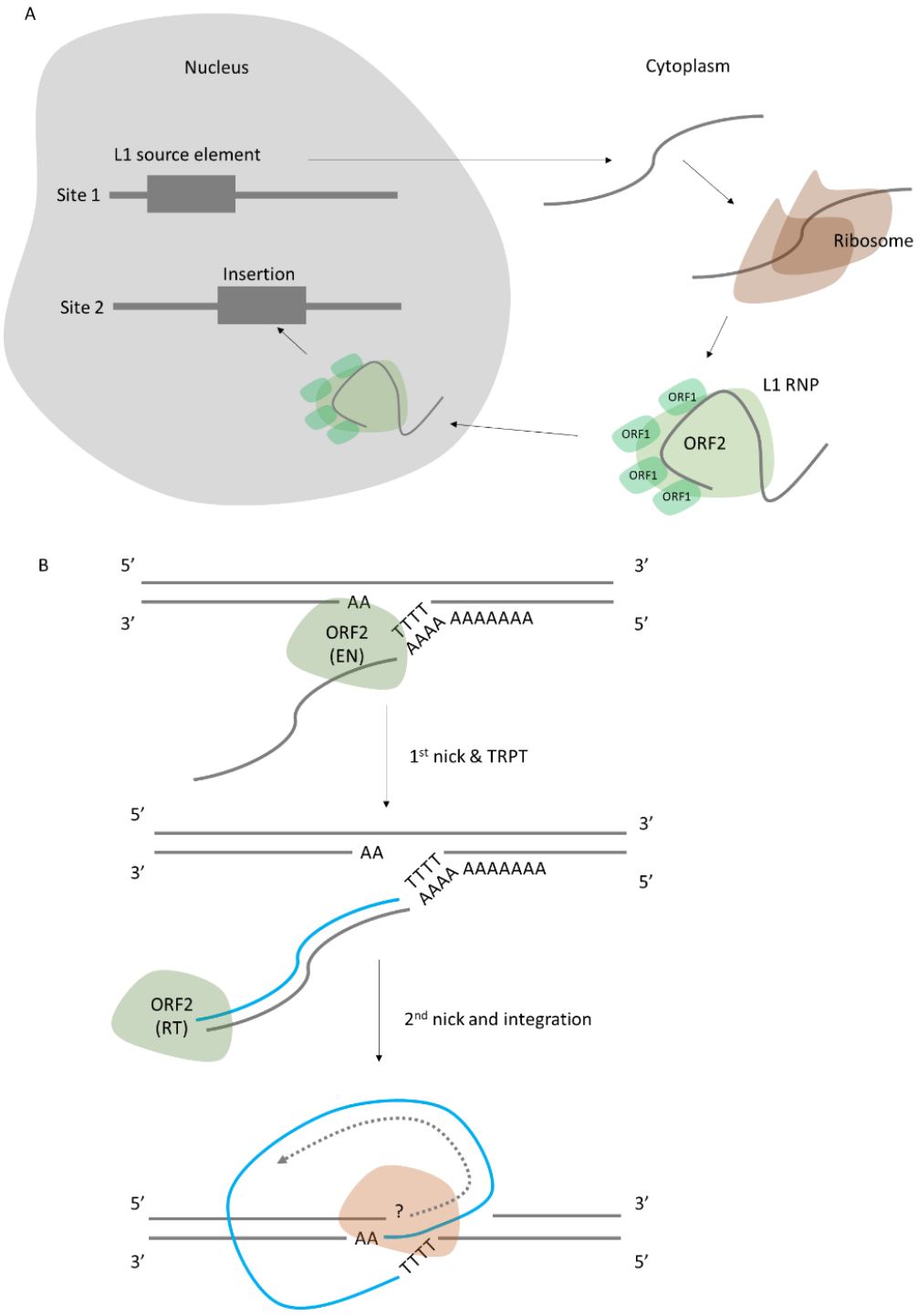


Figure 1.5. Overview of L1 mediated mobilisation and target primed reverse transcription. (A) An overview of mobilisation. The L1 source element is transcribed in the nucleus. This mRNA then moves into the cytoplasm where it is translated by ribosomes, producing ORF1 and ORF2 proteins. ORF1 and ORF2 associate in *cis* with the mRNA from which is was transcribed, forming a ribonucleoprotein complex (RNP). The RNP shuttles back into the nucleus and creates a new L1 copy in a new site using target primed reverse transcription (TPRT). ORF1 and ORF2 can also associate in *trans* with mRNA from Alu and SVA, facilitating their mobilisation too. (B) An overview of TPRT; ORF2 endonuclease activity creates a nick on the bottom strand at the AT-rich recognition sequence, exposing a 3' hydroxyl residue. The poly A tail of the mRNA associates with the poly T at the recognition sequence. ORF2 reverse transcriptase activity then creates a cDNA extending from the poly T, which is used to prime the process. Finally, a second nick is created on the upper strand, resulting in a double strand break (DSB) and the newly synthesised molecule is integrated into the genome, however the mechanisms underlying this part of the process are currently unknown.

1.2.3 A comparison of transposable elements in the mouse genome

In 2002, Waterston *et al.* published the first draft of the mouse genome and compared it against the human genome[8,27]. From this analysis, it was observed that the mouse genome (2.5 Gb in size) is 14% smaller than the human genome (2.9 Gb in size). When comparing genomes at a structural level, it was found that 90% could be aligned into blocks of synteny, which highlighted conservation of gene order since humans and mice diverged from the last common ancestor (LCA) approximately 65 Mya – 75 Mya[27]. When comparing genomes at the nucleotide level, only 40% of human DNA sequences could be aligned, highlighting sequence divergence between the two species[827].

Waterston *et al.* further compared the content of transposable elements in both human and mouse genomes. TEs constitute approximately 45% of the human genome, yet this analysis estimated that TEs constitute approximately 37.5% of the mouse genome[27]. It was predicted that ancestral TEs are more difficult to identify due to base substitutions in the DNA sequence. One hypothesis was that mice have a greater nucleotide substitution rate compared to humans, perhaps due to increased generation time, and is therefore a possible explanation regarding the current difference in identifiable TE content across human and mouse genomes. In contrast to this disparity in ancestral TE sequences, 32.4% (812 Mb) of the mouse genome and 24.4% (695 Mb) of the human genome were composed of lineage specific TE sequences (those that emerged following divergence from the LCA) [827]. This suggests that transposition is more prolific in the mouse genome than in the human genome.

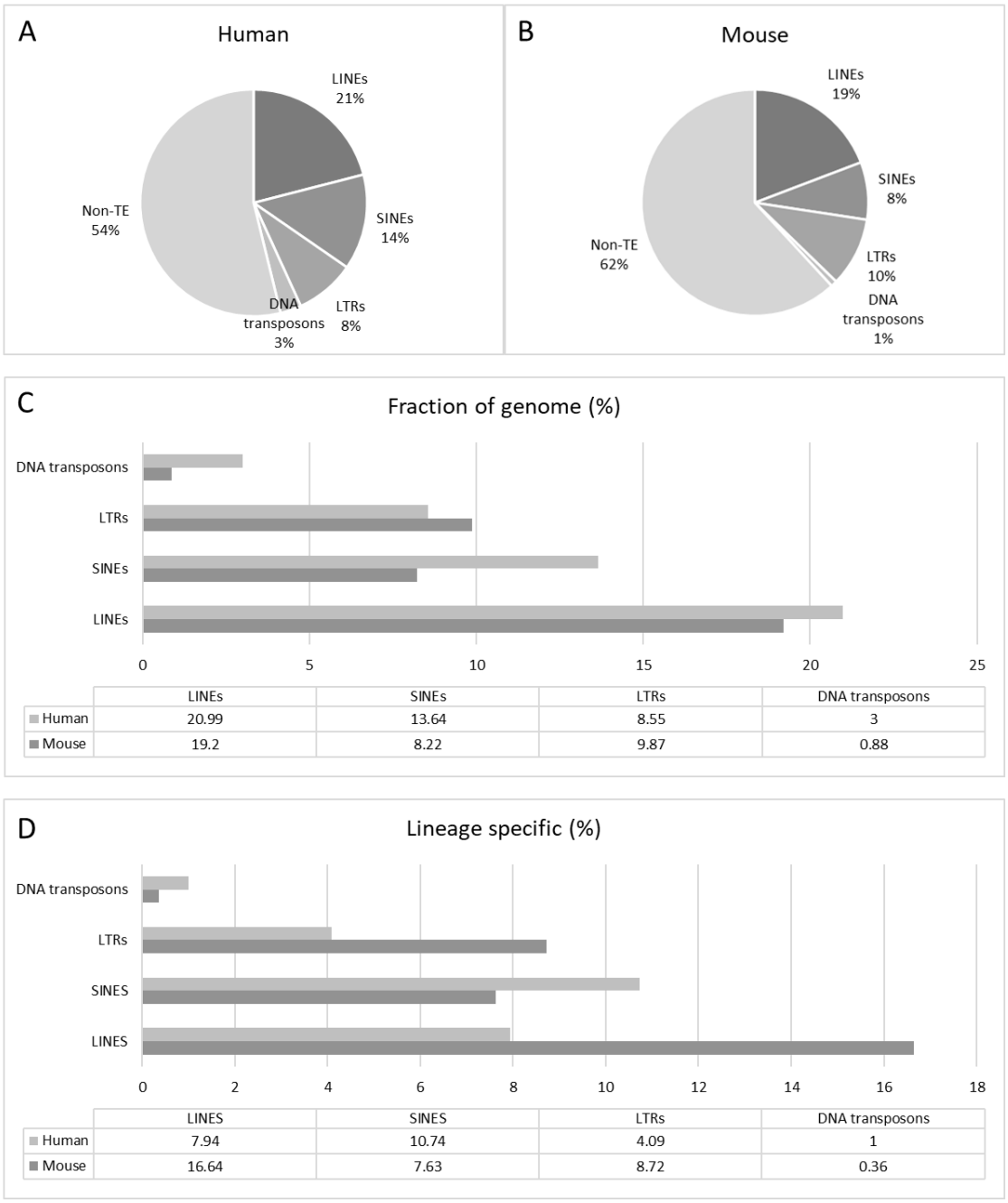
TEs within the mouse genome follow the same classification (DNA transposons, LTRs, and non-LTR SINEs and LINEs) as shown in Figure 1.2., however different subfamilies do occur within each species. A comparison of TE content (based simplistically on classification) in human and mouse genomes is shown in Figure 1.6C. There are some similarities between TE content and activity between human and mice however there are also some differences. LINEs constitute the largest class of TE in both species (Figure 1.6C) but are more active in

the mouse[16,] (Figure 1.6D). There are an estimated 3000 copies of L1 in the mouse genome that are transposition competent, compared to less than 100 "hot L1s" in human[28]. Humans have a greater proportion of the genome attributed to SINEs when compared to the mouse (Figure 1.6C) however there are actually more copies of SINEs in the mouse genome than in the human genome[16] - due to the larger size of SINE elements in the human genome compared to SINEs in the mouse genome (based on sequence length in terms of base pairs). The only active SINE family in the human genome is the primate specific *Alu*, however mice have four including B1, B2, ID and B4. B1 is similar to *Alu* as it is derived from the same 7SL RNA gene[27].

Unlike non-LTR retrotransposons, LTR retrotransposons in human and mouse genomes have a highly dissimilar evolutionary history[16,]. LTR retrotransposons in the mouse genome including intracisternal-A particles (IAP), early transposons (ETn) and ERVs (e.g. murine leukemia virus, MuRRS, MuRVY and VL30) are still active and have contributed to a near two-fold increase in lineage specific repeats (Figure 1.6D), unlike HERVs in the human genome which are considered inactive[16,]. The overall contributions of each TE subfamily to the mouse and human genomes are of relatively equal proportions, highlighting similarities in genome TE composition between these two mammalian species.

Non-LTR retrotransposon insertion sites across mammalian genomes follow a common trend[27,29]. Analysis of placental mammalian genomes (including human and mouse) by Buckley *et al.* identified an insertion bias of lineage specific non-LTR retrotransposons in euchromatic regions and proposed a model that described different effects of purifying selection on different sized elements that was shared across species[29]. For example, LINEs are larger elements and therefore carry a larger mutational burden upon insertion, compared to shorter elements such as SINEs and are consequently more frequently purged from the

genome. This results in an accumulation of SINEs in GC rich, gene dense, euchromatic regions and accumulation of LINEs in GC poor, gene sparse, heterochromatic regions[29].



	LINEs	SINEs	LTRs	DNA transposons
Human	20.99	13.64	8.55	3
Mouse	19.2	8.22	9.87	0.88

	LINES	SINES	LTRs	DNA transposons
Human	7.94	10.74	4.09	1
Mouse	16.64	7.63	8.72	0.36

Figure 1.6. Comparison of TE content in human and mouse genomes. Overview of TE content in mouse genome and human genome. TEs classified based on Class I LTRs, SINEs and LINEs, or Class II DNA transposons. (A) TE content in human genome shown as percentage of DNA sequence. (B) TE content in mouse genome shown as percentage of DNA sequence. (C) Contribution of TEs relative to the entire genome sequence. (D) Contribution of TE classifications to lineage specific repeats. Adapted from Waterston *et al.*[27]

1.2.4 Impact of retrotransposons on gene regulation

Non-LTR retrotransposon insertions can result in changes to host gene regulation through a variety of mechanisms (Figure 1.7). One mechanism is the insertion of a TE that can introduce a novel transcriptional start site (TSS) from the internal promoter encoded within the TE (Figure 1.7A). An example of this is demonstrated in the context of the gene *NAIP*, in which a novel transcript is expressed from an upstream *Alu*, resulting in a novel protein being translated[30].

Another mechanism is via the introduction of novel transcription factor binding sites (TFBS), which are encoded within the TE sequence(Figure 1.7B). Therefore, TEs can function as *cis*-regulatory elements (CREs) and can up or downregulate gene expression dependent upon the proteins bound. This principle was first introduced as the "gene-battery" model by Britten and Davidson in 1969[31]. This was revisited in a recent analysis of the human genome - in which Sundaram *et al.* identified >2 million TFBS overlapping with retrotransposons that resided within 5 kb of a gene promoter[32]. TFBS within TEs accounted for 17% of all TFBS identified within promoter neighbourhoods, highlighting that TEs provide a large source of TFBS in the genome[32].

Whilst insertion of TEs can be tolerated by the genomic neighbourhood and exapted as regulatory domains, they can also be deleterious if inserted into an intronic or exonic sequence, as they can introduce loss of function mutations, premature stop codons (Figure 1.7C) and induce alternative splicing (Figure 1.7D). This is evident in the case of Rotor syndrome, an autosomal recessive disorder which is characterised by the accumulation of bilirubin in the blood[33]. It is a relatively benign yet rare disease, in which affected individuals carry a homozygous L1 insertion in intron 5 of *SLCO1B3*. This insertion causes aberrant mRNA splicing and loss of protein expression in the liver, which is causative of the disease[33].

The variable content of TEs in the genomes of different species drives species-specific gene expression patterns. One study that compared human and mouse genomes, identified using a panel of 26 orthologous transcription factors that 20% of their TFBS were in TEs yet less than 2% were conserved between species[7]. This indicated that TEs supply a large source of species specific TFBS[7], which may contribute to the rewiring of gene regulatory networks. This principle was further validated by Chuong *et al.* who first showed that inducible gene expression changes in response to the proinflammatory cytokine interferon-γ (IFN-γ) differed between humans and mice[34]. This group then performed CRIPSR in HeLa cells, to delete ERVs in the human genome that were suspected of being co-opted as CREs for genes that were found to be differentially expressed between both species in response to IFN-γ. They found that inducible expression of target genes was severely affected in the ERV deleted HeLa cell lines, supporting the role of lineage specific TEs as CREs that influence species specific gene networks. Utilising cancer cell lines *in vitro*, like HeLa, provides an inexpensive route to gain insight into certain biological mechanisms, but their use does carry some limitations[35]. Cancer cell lines possess unstable genomic and transcriptomic profiles compared to the original tissue from which they were derived, and also non-cancer tissues *in vivo*[36]. As a consequence, results obtained from *in vitro* studies utilising cancer cell lines may not necessarily reflect the biological processes that would occur in normal tissues. Thus, results obtained from cancer cell lines should be interpreted cautiously.

Ultimately, any heritable mutation event will be subject to natural selection to determine its fate. TE insertions that confer an adaptive regulatory function will be selected for and more likely become fixed in the population[34,37]. Conversely, TE insertions that confer deleterious regulatory activity will likely be selected against and more likely be purged[38]. In a third scenario, it is also possible that a TE insertion may not confer an immediate adaptive or maladaptive effect, i.e. be a neutral insertion, however with time, they may be latently co-opted via acquisition of sequence mutation that confers an adaptive function. Alternatively,

such sequences can acquire mutations that may lead to deleterious activity, resulting in them being purged from the genome or undergo sequence decay and lose TE sequence identity[38].

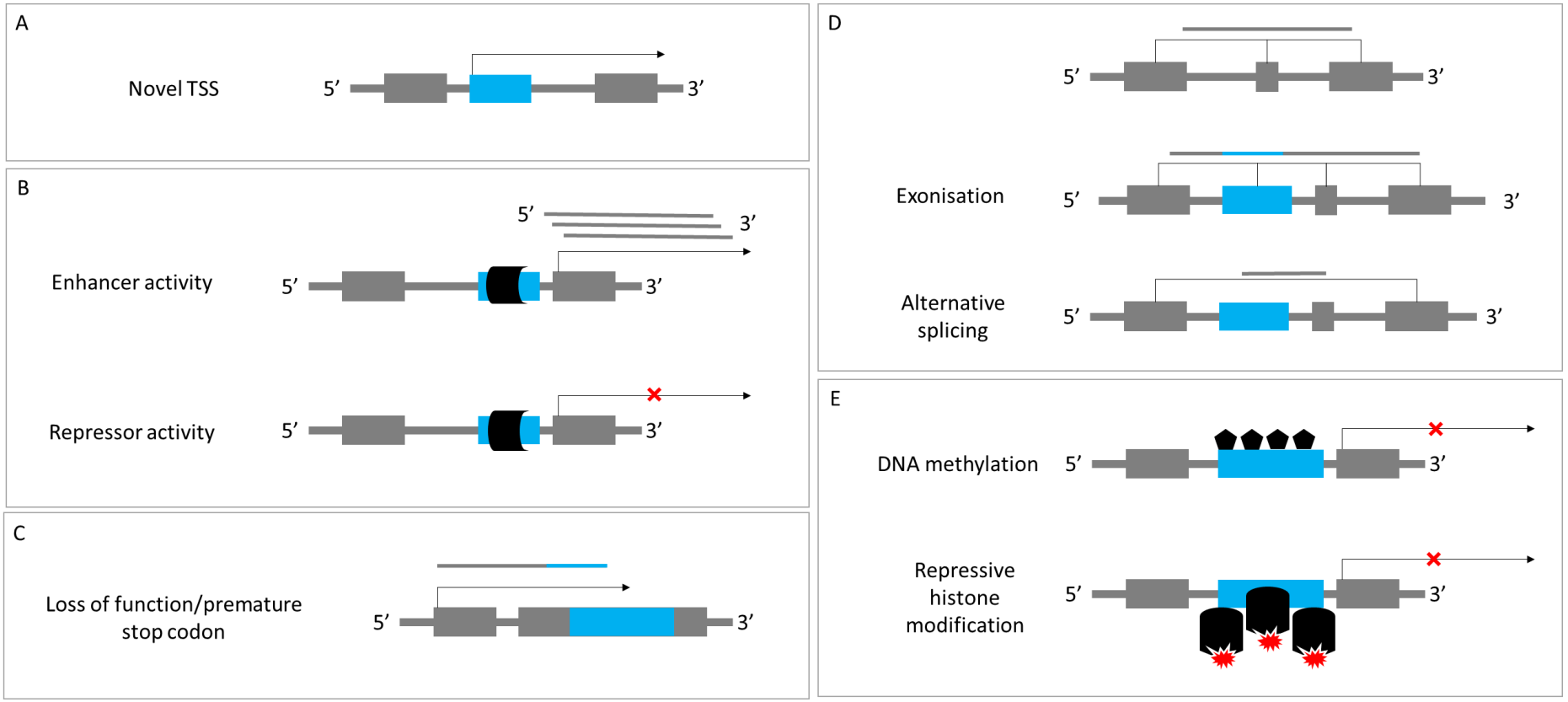


Figure 1.7. Effects of transposable element insertions on *cis*-gene regulation. (A) Insertion of a TE upstream of a gene can introduce a novel transcriptional start site (TSS) and generate new transcripts. (B) TEs can function as CREs by introducing novel transcription factor binding sites that may consequently upregulate or downregulate gene expression. (C) Insertion of TEs into genic sequences can introduce premature stop codons or introduce loss of function mutations. (D) TEs in genic sequences can also affect splicing or undergo exonisation, resulting in novel transcripts. (E) TEs can also be epigenetically modified which can affect the nearby genes (DNA methylation) or genetic neighbourhood (histone modifications/chromatin changes).

1.2.5 Epigenetics modulation of retrotransposon activity and function

For TEs to be maintained in the genome throughout evolution, equilibrium must be struck between expression (and consequent mobilisation) and repression (to protect genome integrity). Control of TEs is mediated through epigenetic modifications such as DNA methylation and histone modifications[38] (Figure 1.7D). This balancing act has resulted in TEs (in mammalian genomes) being locked in an evolutionary arms race to evade epigenetic silencing mechanisms that suppress their activity.

One such mechanism involves Kruppel-associated box (KRAB) zinc finger proteins (ZFPs) which bind to TEs and recruit KRAB-associated protein 1 (KAP1), also known as TRIM28. TRIM28 induces repressive histone modifications and further recruits DNA methyltransferases to methylate and silence the DNA[39]. Accumulation of sequence mutations within supressed TEs enable them to escape suppression and continue to mobilise. KRAB-ZFPs then rapidly evolve to suppress the new wave of TEs that escaped silencing[38]. This is referred to in evolutionary biology as the Red Queen Hypothesis, which describes the nature of antagonistic co-evolution. This hypothesis typically describes adaptation of an organism in response to a selective pressure (e.g. predator versus prey) however this theory can also be applied to TEs and silencing mechanisms which co-evolve within a genomic environment[38]. Other post transcriptional mechanisms such as Piwi-interacting RNA silencing also exist to suppress TEs[40].

The epigenetic status of TEs can influence nearby gene expression (Figure 1.7D). This has been demonstrated in a mouse neural progenitor cell model, in which deletion of TRIM28 resulted in loss of TE silencing (evident by an induction of ERV expression), and correlated with upregulation (average 3-fold) of genes with nearby ERVs (<50kb)[41]. In addition, epigenetic modulation of TEs plays a role in tissue specific gene regulatory networks. Tissues have distinct epigenetic signatures which can affect the ability of TEs to bind transcription factors[7]. One study using human tissues identified hypomethylated TEs in some tissues but

not in others[42]. Hypomethylated TEs were also enriched with histone marks H3K4Me1, indicative of enhancer activity. This was then correlated with increased expression of proximal genes functionally relevant to the tissue type[42] – highlighting TE function and downstream gene regulation may be modulated in a tissue specific manner.

TE activity can also be modulated in response to stimulus. For example, loss of DNA methyltransferase (*DNMT1*) and enrichment of active chromatin histone marks (e.g. H3K4Me3/H3K9ac) at the L1 promoter has been demonstrated in response to Benzo(a)pyrene (a carcinogen found in tobacco smoke) in HeLa cells[43]. A recent study also demonstrated that a low level of maternal care received by mouse pups in the first few weeks of life, reduced the methylation status of the L1 promoter, increased *L1* mRNA expression and increased L1 genomic copy number in the hippocampus[44]. One theory proposes that the increase in TE activity in response to stress introduces random genetic variation which can in turn undergo natural selection[45]. These examples highlight the influence of external physical and behavioural stimuli on epigenetic modulation of TEs across mammalian genomes, which ultimately carry important consequences for gene regulation and genome integrity.

1.2.6 Retrotransposons and genetic variation

Retrotransposons are a large source of polymorphism within the human genome (Figure 1.8C). Retrotransposon insertion polymorphisms (RIPs) refer to specific elements that can be present or absent in individual genomes (i.e. not present in all human genomes). All non-LTR retrotransposons have the capacity to mobilise therefore form the basis of RIPs within the human genome. These RIPs can occur in the germline, with an estimated *de novo* insertion rate of 1 in every 150 live human births for L1s, 1 in 20 for *Alus* and 1 in 1000 for SVAs[46]. It is currently estimated that each person carries 180 L1, 1283 *Alu* and 56 SVA RIPs[8].

Due to the impact of TEs on gene regulation, RIPs also contribute to the phenotypic variation observed between individuals. For example, individuals carrying an SVA RIP in the second intron of the *B4GALT1* gene, have reduced *B4GALT1* mRNA expression in B-cells, highlighting the genetic variation supplied by TEs as important contributors of gene regulatory phenotypes[47]. In addition there are at least 142 documented examples of RIPs being responsible for disease, including haemophilia, X-linked dystonia parkinsonism (XDP) and neurofibromatosis type 1 – all classed as examples of insertional mutagenesis occurring in the germline[48].

RIPs can also occur in somatic cells during embryogenesis[49]. This generates somatic mosaicism within cell populations, or even entire tissues (dependent upon developmental stage), of an individual organism that is not heritable. At present, insertion rates are contested, ranging between 0.04 – 80 insertions per neuron[49,50]. In the last decade, there has been a focus on characterising L1 somatic insertion events in neurons, where they are hypothesised to contribute to neuronal plasticity by reshaping the genetic circuitry responsible for normal neuronal function[49-51]. One early study performed qPCR analysis on whole tissues and identified an increase in L1 genomic copy number in the adult human brain (compared to heart and liver)[50]. The greatest increase was observed in the hippocampus. Another study, using RC-seq (retrotransposon capture sequencing), profiled TE somatic

insertions in the hippocampus and caudate nucleus of three individuals[49]. This method identified thousands of germline L1, SVA and Alu insertions (found in brain and blood, but not the reference genome) and identified thousands of putative somatic insertions (found in brain, but not blood or the reference genome). Whilst these studies indicated that *de novo* L1 insertions are a source of somatic mosaicism in the brain, they relied on using bulk tissue, therefore could not provide any insight into whether these events occurred during or post-development.

More recently, single cell sequencing has provided more insight into somatic insertions in individual neurones. A study by Evrony *et al.* sequenced 50 single neurons taken from cerebral cortex and caudate nucleus from 3 individuals (300 nuclei in total), versus 'bulk DNA' taken from whole tissues (cortex, caudate, cerebellum, heart, liver, lung)[51]. They found that 18% of single neurons taken from the brain contained a unique somatic insertion (i.e. insertion present in a single neuron and not in others), indicating that L1 mobilisation in the brain does occur post-development. The same research effort has not yet been carried out in non-neuronal cells, or in brain regions where highest levels of genomic L1 copy number have previously been reported (i.e. hippocampus).

The majority of the literature focusing on TE polymorphism is dominated by RIPs but it is not the only manner in which TEs can be polymorphic. Retrotransposon sequence length polymorphisms also occur within TEs in the genome, due to their repetitive DNA content which is prone to replication slippage[52]. Several examples of sequence length polymorphisms have been documented in SVAs[53,54]. In addition to sequence variation supplied by TEs, they can also be subject to different epigenetic modifications, contributing epialleles to the genome which may affect regulated genes differently. An example of this is an LTR IAP retrotransposon insertional mutation upstream of the *Agouti* locus in the mouse[55]. This insertional mutation is subject to differential methylation, which drives aberrant gene

expression and determines variable coat colour in individuals that are genetically identical at the DNA sequence level. Genetic variation supplied by TEs is often overlooked due to difficulties mapping RIPs and inner sequence polymorphisms from short read sequencing, however improvements in long-read sequencing technologies and bioinformatic RIP detection tools such as the Mobile Element Locator Tool (MELT) are progressing this field forward[56]. At present, TE polymorphism (both RIPs and sequence length) represents a large and relatively untapped source of genetic variation (Figure 1.8C)

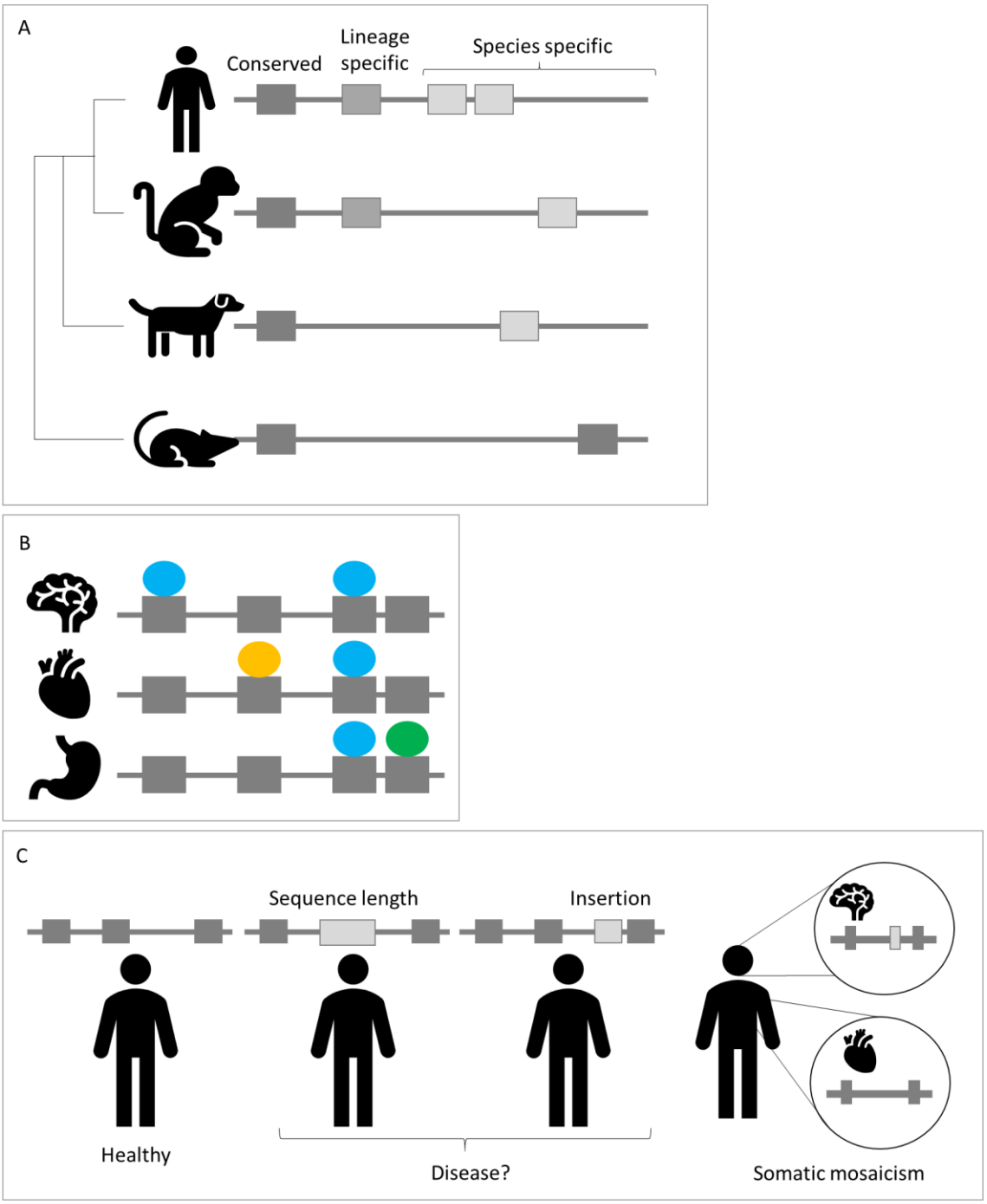


Figure 1.8. Contribution of TEs to phenotype. (A) Species specific TE insertions contribute to the rewiring of gene regulatory networks. (B) TEs also provide transcription factor binding sites that function in response to different transcription factors and epigenetic landscapes present in different cell types, driving the expression of genes in a tissue specific manner. (C) TEs contribute to genetic variation by being polymorphic in sequence length and generating retrotransposon insertion polymorphisms, which in some instances may affect gene expression and potentially contribute to disease. Retrotransposon insertion polymorphisms can also occur during embryogenesis and neurogenesis which contributes to somatic mosaicism.

1.2.7 LINE-1 activity in ageing, disease and inflammation

Cellular ageing can be characterised by an accumulation of genomic instability[57]. The role of TEs in the ageing process was first proposed by Driver & McKechnie in 1992, who demonstrated an association between somatic transposition of P elements in *Drosophila* and reduced lifespan[58]. TEs are enriched in heterochromatic domains, yet heterochromatin decondenses during ageing, which has been shown to lead to an increase in TE activity[59]. This leads to increased somatic mutagenic insertions and DNA damage, causing genomic instability associated with ageing[59].

TE activity also contributes to disease mechanisms that are not simply attributed to single TE mutagenic insertions. Increased levels of L1 ORF2p endonuclease causes DNA damage, which has been demonstrated in HeLa cells that showed increased numbers of γ-H2AX foci (markers of DNA double strand breaks) in response to increased levels of ORF2p[60,61]. Therefore it is not surprising that there are numerous reports implicating the loss of L1 hypomethylation and increased L1 activity in conditions characterised by genomic instability like cancer[62-71] and ataxia telangiectasia[72].

Loss of L1 methylation and increased L1 activity is also implicated in multiple neurological disorders including; Rett syndrome[73], autism[74], schizophrenia[13,75], Alzheimer's[76] and major depressive disorder[77]. The accumulation of DNA damage in the nervous system is a major problem with increasing age due to a lack of DNA-replication-dependent repair pathways (like homologous recombination) in post-mitotic neurons[78], which will be discussed in more detail later.

However, increased TE activity has also been characterised as normal part of chronological ageing. Several studies using model organisms like *Drosophila melanogaster, Caenorhabditis elegans* and *Saccharomyces cerevisiae* have demonstrated that the mechanisms which suppress TE activity become inefficient in ageing, resulting in increased TE activity[79-83]. In

2013, De Cecco *et al.* demonstrated in mouse skeletal muscle and liver, and in a mouse cellular model of senescence, that L1 mRNA expression increased with age[59,84]. Further study by this group in 2018 demonstrated that the increase in L1 mRNA results in cytoplasmic accumulation of L1 cDNA[85,86]. The cDNA is detected by cGAS, which stimulates STING signalling and the induction of type-I interferon (IFN), ultimately causing an inflammatory response[85,86]. This process is thought to resemble exogenous nucleic acids resulting from viral infection and has been shown to be at play in human autoimmune conditions like systemic lupus erythematosus[87,88]. L1 activity certainly plays a role in multiple disease contexts however the work by De Cecco *et al.* also highlighted the role of L1 activity in non-pathological ageing –hypothesised to contribute to sterile inflammation associated with age related pathologies – which has been referred to as "inflammageing".

Neurological diseases involving neuroinflammation are increasingly being characterised by aberrant TE activity and induction of cytokines that mediate inflammatory pathways (Table 1.1). Most of the research regarding TEs and inflammation in disease is focused on neurodegenerative conditions of the central nervous system, for which age is a primary risk factor. No efforts to address this mechanism have yet been published in the peripheral nervous system or relating to conditions like chronic pain.

To summarise, from the evidence presented here, there are two models by which increased transcription of L1 contributes to age-associated cellular dysfunction (Figure 1.9):

1. Genomic instability: Mobilisation of L1 resulting in ORF2p endonuclease mediated DNA damage and mutagenic insertion
2. Sterile inflammation: Accumulation of cDNA in cytoplasm, triggering an inflammatory response via cGAS/STING signalling and type I IFN induction.

Table 1.1. Diseases characterised by retrotransposon activity and inflammation. Adapted from Saleh *et al.*[89]

Disease	Retrotransposon	Associated gene	Elevated Cytokines	CNS vs systemic
Multiple sclerosis (MS)	HERV-W	n/a	IFNγ, IL-6, TNF-α	CNS
Aicardi-Goutieres Syndrome (AGS)	L1	TREX1, RNaseH2	TNF-α, IL-15, IFN-α	CNS
Rett syndrome	L1	MeCP2	IL-6, IL-8	CNS
Amyotrophic Lateral Sclerosis (ALS)	HERV-K, L1	TDP-43	TNF-α, IL-6, IL-8, IL-1β	CNS
System Lupus Erythematosus (SLE)	HERV-E	Sgp3	IL-15, IL-10, IFN α/β, IL-6	Systemic
Aging-related pathologies	L1	SIRT6	IFN	Systemic
Autism Spectrum Disorder (ASD)	L1	n/a	IFNγ, IL-1β, IL-6	CNS

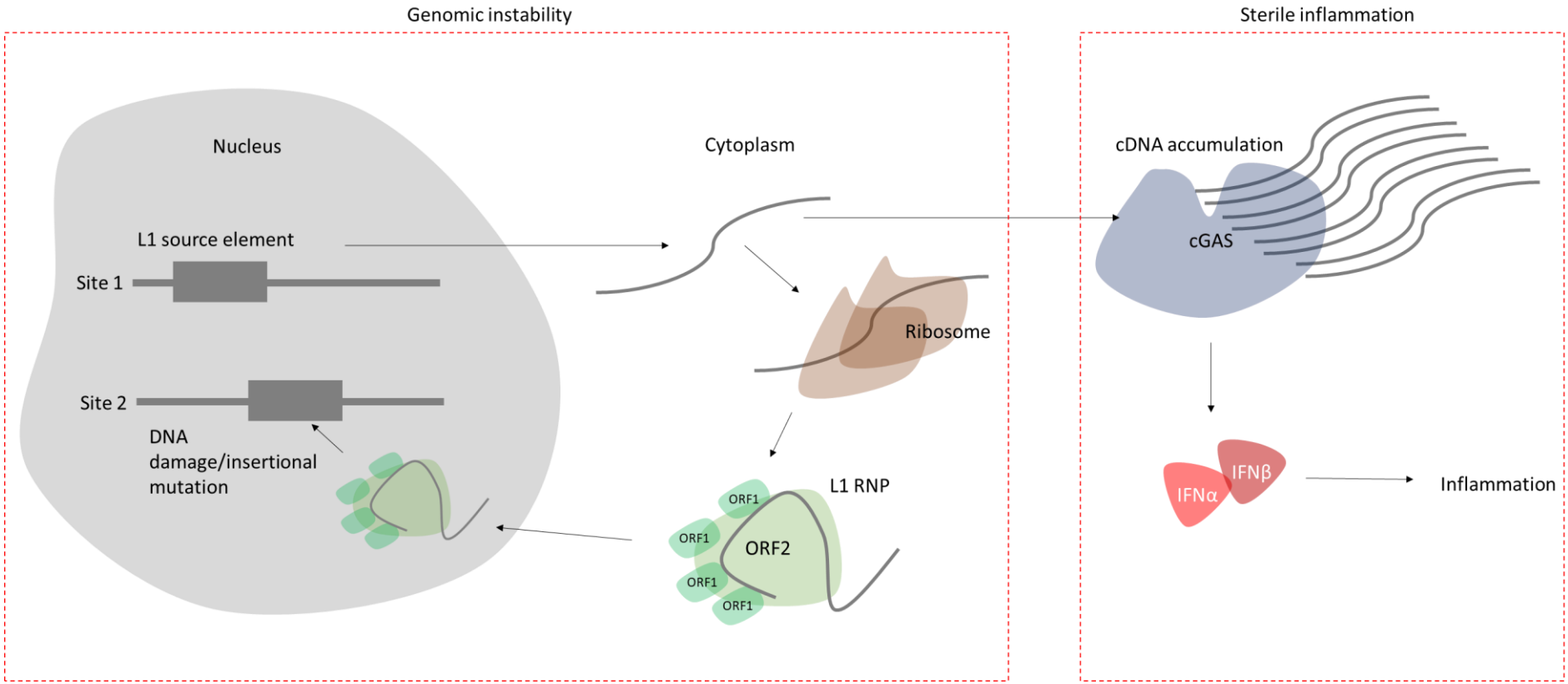


Figure 1.9. Model of L1 activation and consequences. Transcription of L1 and translation of the reverse transcriptase and endonuclease in the cytoplasm generates a ribonucleoprotein particle (RNP) containing the reverse transcribed L1 cDNA. This is then transported to the nucleus, where reintegration of the cDNA occurs through target-primed reverse transcription, potentially resulting in DNA damage. Accumulation of L1 cDNA in the cytoplasm following L1 activation is detected by cGAS, which initiates cGAS/STING signalling. This induces a type-I interferon response and results in inflammation. Adapted from Simon *et al*[86]**.**

1.2.8 Associations between LINE-1 retrotransposons, TDP-43 and neurodegeneration

TARDBP encodes transactive response DNA binding protein 43 kDa (TDP-43), a ubiquitously expressed protein with roles in DNA/RNA binding, gene regulation and mRNA regulation[90]. Several studies have highlighted associations between TDP-43 and regulation of TEs. Analysis of RNA-protein interactions and expression datasets in mouse, rat and human revealed TDP-43 binds TEs including L1[91]. In addition, transgenic mouse models that overexpress TDP-43 and knockdown studies have both shown an increase in L1 transcripts[91]. This indicates that the proper regulation of TDP-43 is crucial for regulation of TEs, however the mechanism responsible for dysregulation in both contexts are not completely understood.

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease, characterised by death of motor neurons and cytoplasmic inclusions[90]. In ALS, lack of nuclear TDP-43 and cytoplasmic aggregates are present in over 97% of cases therefore a lot of current research explores the relationship between TDP-43 and L1 within this disease context[90]. Overexpression of hTDP-43 in a *D. melanogaster* model of ALS revealed impairment of small interfering RNA (siRNA) silencing mechanisms against TEs, resulting in an increase of L1 transcripts[92]. Study of human brain cells with and without nuclear TDP-43 further revealed loss of nuclear TDP-43 was associated with chromatin decondensation across L1 loci, increased L1 mobilisation and overall transcriptome deregulation[93]. Multiple studies across several transgenic models (*C. elegans, D. melanogaster* and mouse) overexpressing human TDP-43 mRNA documented neurotoxic effects including reduced cell viability, neuron loss and mitochondrial dysfunction[94-102]. *Drosophila* and mouse models of ageing (non-ALS) have both shown that TDP-43 protein and mRNA levels decrease with age[103,104], which is interesting as the previous work discussed by De Cecco *et al.* described an increase of L1 mRNA. It is not yet known if these two trends are biologically linked in ageing or in the peripheral nervous system.

1.3 Challenges in chronic pain research

1.3.1 Chronic pain: a socio-economic burden

Chronic pain is defined as persistent pain that occurs for a minimum of 12 weeks and can manifest from musculoskeletal pain, neuropathic pain, inflammatory pain and visceral pain[5]. Pain has evolved as a mechanism to protect an organism from noxious (harmful) stimuli in order to prevent injury and ensure survival (for e.g. individuals with congenital insensitivity to pain often have injuries that go unnoticed[105]), however this physiological response becomes problematic when it persists in the absence of harmful stimuli[106].

Pain is initiated when receptors that detect noxious stimuli are activated in sensory neurons (nociceptors). The cell bodies of sensory neurons are located in structures called the dorsal root ganglia (DRG) in the body and trigeminal ganglia (TG) in the head[107]. DRG and TG also contain glial cells, immune cells and blood vessels[107]. Nociceptors project into the skin and peripheral tissues to detect thermal, mechanical and inflammatory signals that indicate potential dangers and/or injury. When nociceptors are stimulated, they initiate an action potential along the neuron, which is then relayed to the central nervous system via the DRG[107].

In the UK and USA, it is estimated that chronic pain affects between one-third and one-half of the population[1,2]. The occurrence of chronic pain causes the individual significant physiological and psychological distress, with depression often occurring as a comorbid condition[108]. The cost of healthcare expenses and loss of productivity attributed to pain is estimated to cost the US economy at least $560 billion each year[3]. The prevalence of chronic pain is associated with ageing therefore this burden is predicted to grow with an ever increasing ageing population[1].

Research conducted over the past few decades has revealed many of the neurobiological mechanisms that underly pain, however this condition is still poorly managed

by current therapeutics, which relies heavily on administration of opioids, and often leads to severe side effects and addiction. In 2018, it was estimated that 3% of the population in the US misused opioids[4]. As a result of this, the USA is currently in the midst of an opioid epidemic which has led to a drastic increase in the number of opioid based deaths from the use of illegal substances like heroin and heroin analogues like fentanyl[4]. This has resulted in the need for accelerated research, which will give further insight into mechanisms underlying pain and improve current management, with the National Institutes of Health launching the HEAL initiative – a $945 million research fund to aid this process.

1.3.2 The genetics of chronic pain

Systematic review of twin studies has estimated the heritability of chronic pain to be 16–50%, indicating a strong genetic component[109,110]. Several monogenic pain disorders have been identified, a well-known example being a nonsynonymous mutation in the gene *SCN9A*, which encodes the $Na_v1.7$ channel, and causes congenital insensitivity to pain (CIP)[105]. However, in general, chronic pain is a heterogeneous and complex disease, with variability between phenotypes including pain severity and response to therapeutics[111].

Extensive use of transgenic and knockout mouse models has identified 430 genes associated with pain (at the time of writing). These are published in the Pain Genes Database, curated by the Mogil lab[101]. Profiling genetic variants from hypothesis driven candidate gene studies in human cohorts has identified 846 common variant SNPs (found in >1% of the population) associated with pain (at the time of writing)[112,113]. These results are published in the Human Pain Genetics Database, curated by the Diatchenko lab. These databases demonstrate the large efforts made in pain biology and have been paramount in identifying the molecular components that constitute pain pathways. However, our understanding of pain is far from complete and the treatment of pain remains problematic.

Genome wide association studies (GWAS) enable identification of common SNPs in association with phenotypes using an unbiased, non-hypothesis driven approach. Migraine is the most successful pain condition for which GWAS has been applied; four GWAS and four meta-analyses GWAS have been conducted to date[114-120]. Chronic pain GWAS has not been as successful, with only a few studies currently published that that focus on specific subsets of chronic pain phenotypes including; chronic widespread pain[121], diabetic neuropathic pain[122], multisite chronic pain[123] and fibromyalgia[124]. Chronic pain is a difficult condition to apply the concept of GWAS to, due to its heterogeneity and difficulties in consistent pain reporting. An emerging trend is now beginning to view chronic pain as an umbrella term for a collection of different diseases which are driven by key changes in the peripheral nervous

system[125]. Going forwards, an alternative approach to chronic pain genetics may be needed, as the likelihood of obtaining a homogenous cohort to yield biologically relevant results from GWAS is unlikely.

1.3.3 From bench to bedside: are current pain models appropriate?

The use of rodent models in pain research has been essential in reaching our current level of understanding of pain mechanisms, however one limitation has been the failure of analgesic development in the last two decades. The latest review of therapeutics that have failed at clinical trial phase II revealed that 48% of failure was attributed to poor efficacy, followed by 25% of trials failing due to safety issues[126]. This review included therapeutics across multiple disease areas (e.g. CNS, cancer, cardiovascular) and there is widespread debate regarding the translatability of preclinical research into effective therapeutics across human disease, including pain[127]. Several therapeutics for pain which have failed at phase II over the last two decades are listed in Table 1.2.

A recurrent question in pain research is how representative are rodent models of human pain? In addition to genetic studies, other attempts to identify relevant clinical targets has been through the use of RNA-seq (Table 1.3Tab). This strategy aims to discover aberrantly expressed mRNA in association with disease or specific gene expression profiles in relevant pain tissues, however most studies have been conducted within a single species (i.e human or rodent), therefore species similarities and differences are frequently ignored.

In an attempt to assess the appropriateness of mouse DRG as a model for human pain, Ray *et al.* compared the expression profiles of DRG in human and mice[6]. They described a subset of orthologous genes that were enriched in DRG, of which 63.8% of genes showed conserved expression patterns between the two species ($R>0.68$). This indicated that a large subset of genes specific to DRG function are conserved (many of them being core transcription factors), however there was also a large proportion of genes that showed species specific expression in DRG. Species specific expression patterns in DRG may therefore contribute to some of the difficulties facing chronic pain therapeutic development – which remains to be addressed.

Table 1.2. Pain therapeutics that failed at clinical trial phase II between 2001 - 2020.

Therapeutic target	Trial group	Name	Reason for failure	Year	Reference
NMDA antagonist	Post-surgical pain	Amantadine	Efficacy	2001	[128]
Glycine antagonist	Neuropathic pain	GV196771	Efficacy and safety (adverse reactions)	2002	[129]
Na^+ channel blocker	Neuropathic pain	4030W92	Efficacy	2002	[130]
NMDA antagonist and morphine	Chronic pain	MorphiDex	Efficacy	2005	[131]
TRPV1 antagonist	Post-surgical pain	AMG 517	Safety (hyperthemia)	2008	[132]
COX-2 inhibitor	Neuropathic pain	GW406381	Efficacy	2009	[133]
Cannabinoid agonist	Post-surgical pain	GW842166	Efficacy	2011	[134]
5-HT-3 antagonist	Neuropathic pain	Ondansetron	Efficacy	2011	[135]
Chemokine antagonist	Posttraumatic neuralgia	AZD2423	Efficacy	2013	[136]
Glia modulators	Neuropathic pain	Minocycline	Efficacy	2013	[137]

Table 1.3. RNA-seq studies conducted to identify novel pain targets.

Tissue	Disease	Species	Reference
Knee synovial tissue	Osteoarthritis	Human	[138]
Prefrontal cortex	Peripheral nerve injury	Mouse	[139]
DRG	Neuropathic pain	Human	[140]
DRG	Diabetic neuropathy	Rat	[141]
DRG Peripheral nerve	Inherited sensory neuropathy	Human Rat	[142]
DRG Trigeminal ganglia Brain Olfactory epithelium Liver Skeletal muscle	n/a	Mouse	[143]
DRG Trigeminal ganglia Brain Liver Lung Colon Skeletal muscle	n/a	Human	[144]
DRG Spinal cord Hippocampus Nucleus accumbens Frontal cortex Liver Small intestine Lung Whole blood Skeletal muscle Heart	n/a	Human Mouse	[6]

1.3.4 Hot topic: targeting temperature receptors TRPV1 and TRPV3

TRPV1 encodes transient receptor potential vanilloid 1, a voltage-gated ion channel that is activated by multiple physiological stimuli therefore considered to be polymodal signal mediator. TRPV1 is activated by noxious heat[145] (>42°c), voltage[146], ATP[147], capsaicin[148], endocannabinoids[149] and other compounds. Interestingly, capsaicin is a compound found in chilli peppers, which is why it feels "hot" when we eat them. It is a key player in detection of nociceptive stimuli and pain signalling[150-152]. It is important to note that there are many receptors expressed in sensory neurons, that detect different types of noxious stimuli such as chemical[153] and mechanical[154], yet the work in this thesis focuses exclusively on thermal receptors TRPV1 and TRPV3.

In the human genome, *TRPV1* is encoded at a locus on chromosome 17 (chr17p13.2), adjacent to the gene *TRPV3* which encodes transient receptor potential vanilloid 3. TRPV3 is activated by innocuous temperatures (>32°c)[155], organic compounds including camphor[156], 2-ABP[157] and calcium[158]. The genomic organisation of *TRPV1* and *TRPV3* in the human genome is conserved in the mouse genome in a syntenic block on chromosome 11 (Figure 1.10).

TRPV1 and TRPV3 are key nociceptive receptors[159]. However, they do not act in isolation and function as components of a complex molecular environment that contributes to pain signalling. Peripheral and central sensitisation that leads to chronic pain often results from inflammation-associated changes in the environment surrounding the nerve terminal, that typically occur due to tissue damage[160]. Following tissue damage, activated nociceptors, non-neuronal cells (keratinocytes, fibroblasts and endothelial cells) and immune cells (macrophages, neutrophils and mast cells) release a wide range of endogenous factors that accumulate around the peripheral nerve terminal[161]. These factors consist of a wide range of signalling molecules including cytokines, chemokines, proteases, protons, neutrophins, eicosanoids and related lipids (endocannabinoids, leukotrienes, thromboxanes,

prostaglandins), peptides (substance P, CGRP, bradykinin) and neurotransmitters[161]. Nociceptor neurons express receptors like TRPV1 and TRPV3, that can be activated by these endogenous factors, which in turn heightens the excitability of the nerve fibre and increases sensitivity to temperature[161,162].

In the context of TRPV1, the channel itself can be modulated directly via protons and lipids, or conversely, other factors such as bradykinin, ATP and NGF bind to other receptors which results in downstream intracellular signalling pathways which in turn leads to modulation of TRPV1[161,162]. For example, NGF binds TrkA, which triggers downstream signalling pathways including PI3K-Src and ultimately phosphorylation of TRPV1, which lowers its activation threshold[162]. Activation of nociceptors results in the release of glutamate at the nociceptors central terminal which generates excitatory-post synaptic currents in the second order neurons in the dorsal horn[161]. In the case of injury, nociceptors also release increased levels of neurotransmitters, including substance P, which results in depolarisation of post-synaptic neurons and transmission of the pain signal to the brain[161].

Changes in expression of both *TRPV1* and *TRPV3* are associated with pain. TRPV1 expression is increased in injured DRGs in mouse models[163] and in several human pain related conditions including irritable bowel syndrome[164], chronic pancreatitis[165] and vulvodynia[166]. *TRPV1* and *TRPV3* mRNA and protein expression also increases in human peripheral nerves and DRG after injury[167], and in keratinocytes of women with breast pain[168]. Activation of TRPV1 in keratinocytes has been demonstrated to activate nociception in mice[169]. It is hypothesised that keratinocytes release molecules upon activation of TRPV1 which subsequently activate sensory neuron terminals in a paracrine fashion, however the direct mechanism responsible for this is yet to be elucidated[169]. The role of a keratinocyte is not specifically to detect noxious stimuli however they are juxtaposed to sensory neurons in the skin, highlighting the role of this cell type in cross talk between cell types that contributes to

pain signalling. TRPV1 mRNA expression is elevated in blood of patients with chronic pain[170]. *TRPV3* expression is elevated in the skin of burn scars[171]. The mechanisms leading to increased *TRPV1* and *TRPV3* mRNA expression in these conditions are not currently well understood. Mice lacking TRPV3 also have impaired thermosensory function in response to noxious and innocuous heat (30–42°C)[172]. This was measured based on behavioural changes in *TRPV3* knockout mice compared to wild type mice, in which wildtype mice showed a preference to occupy warmer surfaces (35°C) compared to knockout mice.

TRPV3 has the highest expression level in human and mouse keratinocytes. TRPV3 stimulation in keratinocytes (from environmental/inflammatory stimuli) induces ATP release which results in calcium influx in DRG neurons, thereby playing a role in nociceptive function in the skin of both species[173]. However, expression of *TRPV3* between species is not conserved. *TRPV3* has only been documented in mouse keratinocytes[152,174] yet is expressed in a variety of tissues in human, including; brain, placenta, skin, small intestine, spinal cord, stomach, testes, trachea and DRG[9,10]. This was further validated in the RNA-seq study published by Ray *et al.,* who identified *TRPV1* mRNA expression as strongly conserved between human and mouse DRG but not *TRPV3*[6]. These data indicated that the transcription of *TRPV3* is differentially regulated between human and mice, suggesting potential species-specific CREs at the *TRPV1* and *TRPV3* locus.

Effective therapeutics targeting these channels are limited in the clinic. Capsaicin, applied topically, acts as an agonist of TRPV1 inducing desensitisation and is routinely used for treating neuropathic pain associated with arthritis[175] and diabetes[176]. TRPV1 antagonists that have struggled to make it into the clinic are include AMG517, A-1165901 and AMG7905. AMG517 failed at clinical trial phase II due the side effect of hyperthermia in human subjects[123]. A-1165901 and AMG7905 failed in rodent models due to hypothermia[132,177]. Therapeutic targeting of *TRPV3* is not well documented in the literature. A single inhibitor of

TRPV3, named GRC15300 (originally developed by Glenmark), reached clinical trials for the treatment of pain in 2012, but this was discontinued by Sanofi following its phase II failure in 2013 (the results of this trial have not been published). No other clinical trials targeting TRPV3 have since been reported. Current research is failing to deliver safe effective therapeutics targeting these channels therefore it is important to better understand their transcriptional regulation in the human genome.

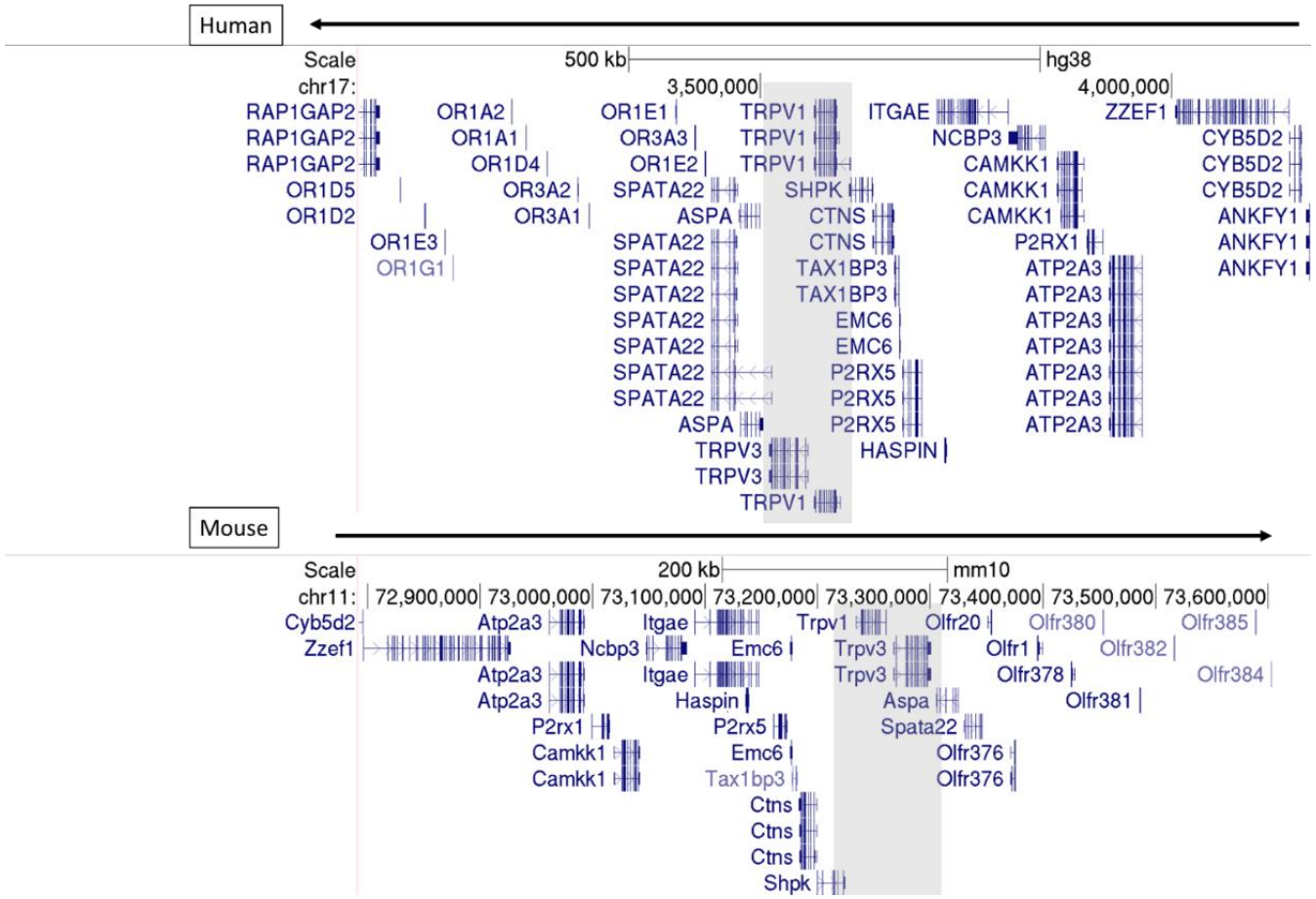


Figure 1.10. Synteny between human and mice *TRPV1* and *TRPV3* loci. A screenshot from UCSC browser (hg38) showing the organisation of genes in a syntenic block in the human genome on chromosome 17 (above) and mouse genome on chromosome 11 (below). The block runs from right to left in the human genome and left to right in the mouse genome. Gene transcripts are shown by blue tracks. *TRPV1* and *TRPV3* transcripts are highlighted in grey boxes demonstrating their close proximity to each other.

1.3.5 Epigenetics and neuroinflammation in the development of chronic pain

An underlying cause of chronic pain is attributed to neuroinflammation resulting from tissue or nerve injury. There are multiple contributors to the activation of neuroinflammation in addition to trauma including age[178], underlying autoimmune or inflammatory conditions like multiple sclerosis or osteoarthritis[179,180], or even external triggers such as stress[181], diet[182,183] and drug treatments[184] (e.g. chemotherapy). Neuroinflammation drives the transition from acute to chronic pain by the process of peripheral and central sensitisation (increased sensitivity to stimuli)[185]. An example of peripheral sensitisation can be demonstrated with TRPV1 and response to the proinflammatory cytokine tumour necrosis factor-α (TNFα). In rat DRG neurons, p38 MAP kinase activates TRPV1 in response to TNFα[186]. This also increases TRPV1 mRNA levels, resulting in more TRPV1 channels in the peripheral terminal of nociceptors, which maintains thermal hypersensitivity[187].

Epigenetic modification in response to neuroinflammation is an underlying mechanism behind sensitisation[188,189]. This can occur in the DRG (peripheral sensitisation) or in pain-associated brain regions (central sensitisation) but as the work in this thesis is focused on the peripheral nervous system, we will present some examples from rodent models that occur in DRG. Rodent nerve injury models have demonstrated changes in histone modifications and DNA hypomethylation in DRG[190-192]. The increase of DNA (cytosine-5)-methyltransferase 3A (DNMT3a) and methyl CpG binding protein 2 (MeCP2) has also been observed in rat DRG in response to nerve injury[193,194]. These examples demonstrate DNA methylation changes in response to neuroinflammation induced by peripheral nerve injury.

In the case of pain gene *TRPV1*, chronic stress has been shown to induce histone acetylation at the *TRPV1* promoter in mice DRG, resulting in increased mRNA expression[195]. Furthermore, knockout of *HDAC4* (a histone deacetylase) in the mouse reduced *TRPV1* expression in DRG injury models, reducing responsiveness to capsaicin and thermal

sensitivity, implicating the role of epigenetic modulation in transcriptional responses that initiate pain[196].

A large number of studies describing epigenetic changes in the DRG and their contribution to pain phenotypes have been reviewed extensively[197] however many of the studies use rodent model of nerve injury. Expression of inflammatory genes also increases with ageing in the DRG however little is known the epigenetic mechanisms that contribute to this[198]. One study in humans estimated the epigenetic age of humans (from profiling the methylation status at genomic sites in blood, a technique termed Horvath's epigenetic clock) and identified that an increased epigenetic age was more associated with chronic pain[199], yet this area, specifically in the DRG, requires more attention.

As discussed in the previous section ""L1 activity in disease, ageing and neuroinflammation" (section 1.2.7), epigenetic changes are well documented in other age-related diseases like cancer and neurodegenerative conditions in the CNS. These diseases share a common biological theme – the activation of the TE family L1 and the occurrence of neuroinflammation. The recent work demonstrating the accumulation of cytoplasmic L1 cDNA and induction of an inflammatory response in ageing tissues in the absence of cancer or neurodegenerative highlights their role as a driver of sterile inflammation, which is a hallmark of ageing and plays a role in many age-related pathologies. To the best of our knowledge, no study of retrotransposons in the peripheral nervous system (specifically the DRG) has yet been published, highlighting an unexplored area of research which could yield new insights into the mechanisms that drive chronic pain.

Chapter 2.

Materials and methods

2.1 Materials

2.1.1 Cloning materials

2.1.1.1 LB broth and agar

LB broth and agar was required for use as growth medium and was prepared accordingly. LB broth was prepared using 25 g/L in distilled water. LB agar was prepared using 40 g/L in distilled water. All LB reagents from Fluka Analytical. Preparations autoclaved before use.

2.1.1.2 Agar plates for bacterial growth

Following transformation, bacteria were plated for growth on agar plates containing an antibiotic. This selected for bacteria that were transformed with a vector containing an antibiotic resistance gene. To make antibiotic containing agar plates, LB agar was heated in the microwave until melted. Once cooled, yet still liquid, 400 µl of ampicillin solution was added to the 400 ml LB agar. To prepare the ampicillin solution, Ampicillin sodium salt (Sigma, A0166) was dissolved in nuclease free water at a concentration of 1000 µg ml^{-1} and filter sterilised.

2.1.1.3 Bacterial cell freezing media

Bacterial cultures containing relevant plasmids were frozen down at -80 °C in LB broth supplemented with 15% glycerol. Glycerol was filtered prior to use to ensure sterility.

2.1.2 Human cell lines and culture materials

2.1.2.1 Culture media reagents

Table 2.1. List of commonly used reagents in cell culture.

Reagent	Supplier (catalogue number)	Cat. no.
PBS pH 7.2 (1x)	Gibco	20012-019
Sodium pyruvate solution 100 mM	Sigma	S8636
Fetal Bovine Serum, heat inactivated	Gibco	10500-064
Trypsin-EDTA solution	Sigma	T3924
Penicillin-Streptomycin solution	Sigma	P0781
L-glutamine 200 mM	Gibco	25030081
DMSO	Sigma	PHR1309

2.1.2.2 HAP1 cells and growth media

HAP1 is a near haploid cell line, derived from male chronic myeloid leukemia cell line KBM-7 by Horizon (catalogue no. C631). Further information is available at https://horizondiscovery.com/products/gene-editing/cell-line-models/PIFs/HAP1-parental-cell-lines. Culture medium was composed of Iscove's Modified Dulbecco's Medium (IMDM) (1X) (Gibco, 12440-053) supplemented with 10% FBS, 1% penicillin-streptomycin and 1% L-glutamine.

2.1.2.3 SH-SY5Y cells and growth media

SH-SY5Y (ATCC® CRL-2266™) is a cell line derived from the neuroblastoma SK-N-AS, originally derived from a metastatic bone tumour. Culture media was a 1:1 ratio of Minimum Essential Medium Eagle (MEME) (Sigma, M2279) and Nutrient Mixture F-12 Ham (Sigma, N4888), supplemented with 10% FBS, 1% penicillin-streptomycin, 1% sodium pyruvate and 1% L-glutamine.

2.1.2.4 HEK293 cells and growth media

HEK293 (ATCC® CRL-1573™) is a hypotriploid human cell line derived from human embryonic kidney cells. Culture media was made from Dulbecco's Modified Eagle Medium (DMEM) containing 4.5 g/L D-Glucose and L-glutamine, supplemented with 10% FBS, 1% penicillin-streptomycin and 1% sodium pyruvate.

2.1.3 Murine DRG purification and primary cell culture materials

2.1.3.1 Commonly used solutions and materials

Table 2.2. Materials used for DRG dissociation and primary cell culture.

Item	Supplier	Cat. no.
Hank's Buffered Saline Solution	Gibco	14175095
HEPES (1M)	Gibco	15630080
Neurobasal™-A Medium	Gibco	10888022
Penicillin-Streptomycin (10,000 U/mL)	Gibco	15140122
L-Glutamine (200 mM)	Gibco	25030081
B-27™ Supplement (50X), serum free	Gibco	17504044
Fetal Bovine Serum, certified, heat inactivated	Gibco	10082139
Poly-L-lysine solution	Sigma	P4707
Collagenase, Type 2	Worthington	LS004176
Glass Pasteur pipette	Appleton Woods	RC022
BD Microlance™ Stainless Steel Needle 0.29mm	Fisher Scientific	10442014
Falcon™ Cell Strainer	Fisher Scientific	10282631

2.1.3.2 Fire polishing glass Pasteur pipettes

In order to dissociate the DRG into single cells, they were passed through a glass Pasteur pipette with a modified pore diameter to achieve complete dissociation. An unmodified pipette was also required. These were autoclaved before use. The pore was modified as follows:

1. Insert a 0.29mm needle into the Pasteur pipette pore.
2. Hold this over a flame to shrink the Pasteur pipette pore around the needle.
3. Allow to cool and remove the needle.

2.1.3.3 Preparing DRG cell culture plates

To aid the attachment of cells to culture plates, glass cover slips were placed at the bottom of each well and coated with poly-L-lysine. Once coated, plates were incubated overnight at 37 °C and stored at 4 °C prior to use. Plates were washed with PBS and air dried before plating cells.

2.1.3.4 DRG dissociation solution

DRG harvested from mice were placed in a dissociation solution and kept on ice prior to culture preparation. This was to preserve DRG during the lengthy harvesting protocol and during transport from the Biomedical Services Unit to the laboratory where culture preparation took place. Dissociation solution was prepared as follows:

1. Remove 8.5 ml of Hank's Buffered Saline Solution (HBSS) from 500 ml bottle.
2. Add 3.5 ml HEPES to HBSS.
3. Add 5ml Penicillin/Streptomycin.

2.1.3.5 DRG collagenase solution

Collagenase solution was used to break down the structure of the DRG and allow plating of single cells. This was prepared fresh during each dissociation, as follows:

1. Aliquot 1 ml of dissociation solution into an Eppendorf tube.
2. Dissolve 10 mg of collagenase powder in dissociation solution by incubating at 37 °C for 10 mins.

2.1.3.6 DRG primary cell culture growth medium

DRG culture medium was used to maintain the primary cell culture for up to one month after plating. The media was prepared as follows:

1. Add 38.6 ml of Neurobasal™-A Medium (supplemented with 10% FBS) to 50 ml falcon
2. Add 200 µl L-glutamine
3. Add 800 µl B-27 supplement
4. Add 400 µl Penicillin/Streptomycin solution

DRG culture medium was stored at 4 °C and prewarmed to 37 °C before use.

2.1.4 Mouse strains

2.1.4.1 Thy1-YFP mice used for DRG samples

Male and female Thy1-YFP mice[200] (Jackson Laboratories) (JAX stock #003709) of various ages were housed at the Biomedical Services Unit at the University of Liverpool. Both male and female mice were included as increasing evidence supports sex-specific differences in dorsal root ganglia pain models[201]. These transgenic mice are bred from the C57BL/6 strain and express yellow fluorescent protein in sensory neurons. Mice were kept between 22-25 °C and subject to a 12-hour light/dark cycle. Mice were fed a standard laboratory diet (SDS, 801722). Food and water were *ad libitum*. No procedures were performed on these animals (therefore no Home Office license was required) however their use was in accordance with the UK Animals (Scientific Procedures) Act 1986 and were ethically approved by the University's Animal Welfare Committee. Mice were culled using cervical dislocation.

2.1.4.2 C57BL/6 mice used for brain samples

Male and female C57BL/6 mice (Enivgo) were housed at the Central Research Facility at the University of Glasgow. Mice were kept at 22 °C and subject to a 12-hour light/dark cycle. Mice were fed a normal chow diet until 10 weeks of age, then either continued on a normal chow diet or switched to a high fat diet (Table 2.3). From 9 weeks of age, mice undergoing treatment trials were given either PBS or 1 µg of ES-62 via subcutaneous injection. ES-62 was purified from spent culture medium used for growth of the filarial nematode *Acanthoceilonema viteae*, as previously described in McInnes *et al.*[202]. Mice received Fenbendazole (150 ppm) used as an anthelminthic to control pinworm. Mice were fasted overnight prior to culling. Mice were culled using cervical dislocation at 56 days (8 weeks) or 500 days (71 weeks). All animal handling was conducted by Felicity Lumb and Jennifer Crowe at the Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, UK. All procedures were undertaken under a UK Home Office Project Licence and in accordance with local ethical committee guidelines.

Table 2.3. Dietary components for C57BL/6 mice.

Component	Chow		High fat	
	Proportion (%)	Atwater fuel energy (%)	Proportion (%)	Atwater fuel energy (%)
Fat	3.36	9.08	21.4	42
Protein	18.35	22.03	17.5	15
Fibre	4.23	n/a	3.5	n/a
Sucrose	3.9	68.9	33	43

2.1.5 Human DNA samples

2.1.5.1 Samples used for assay optimisation

Genomic DNA from healthy members of the general population (all of German/European descent) were provided by Prof. Dan Rujescu at the Department of Psychiatry at the Martin Luther University Halle-Wittenberg, Germany. The average age of individuals was 46 years. Healthy controls were screened for psychiatric disorders and had medical records reviewed, including those of first-degree relatives, to rule out health issues. All individuals provided written informed consent and the study was approved by the ethics committee of Ludwig Maximilians University in Munich, Germany.

2.1.5.2 Osteoporosis cohort used for genotyping study

Genomic DNA from patients with osteoporosis and being treated for bone pain was provided by Daniel Carr at the Department of Molecular and Cellular Pharmacology, University of Liverpool, UK. Patients were recruited from the Royal Liverpool University Hospital, UK. DNA was purified from blood using the Chemagic™ Magnetic Separation Module I (Chemagen). This cohort contained both males and females, from ages ranging 44 – 91 years. Patients were split into 2 groups: (A) currently undergoing treatment with tramadol and (B) previously experienced adverse drug reactions (ADRs) to tramadol. Patients in group A were prescribed 50 mg – 400 mg tramadol per day (in divided doses). Patients in group A were assessed for degree of pain intensity using the Visual Analogue Scale (VAS) ranging from 1 – 10, the McGill Pain Questionnaire (MPQ) and Medical Outcomes Study (MOS) SF-36. Pain scores from the visual analogue scale were taken prior to drug administration and three hours afterwards, upon which the difference was calculated. All patients were monitored for ADRs All patients gave written informed consent. This study was approved by the Liverpool Local Research Ethics Committee at the University of Liverpool, UK.

2.1.5.3 Dyne-Steele cohort used for genotyping and methylation studies

Genomic DNA from aged participants belonging to the Dyne-Steele cohort was provided by Prof. Antony Payton and Prof. Neil Pendleton at the Institute of Collaborative Research on Ageing at the University of Manchester, UK. This is a cohort of individuals that represent healthy cognitive ageing, all aged 50+ years, with a median age of 63 years at recruitment. The age at which the DNA samples were taken was not disclosed. DNA was purified from blood using DNAce MaxiBlood Purification System (Bioline) according to manufacturer's instructions. DNA samples provided were from both sexes. Patients were given a questionnaire that included self-reported pain measures. Ethical approval was obtained from the University of Manchester. All individuals gave written informed consent.

2.1.6 Other materials

2.1.6.1 TBE buffer

TBE buffer was used to make agarose gels. Reagents in Table 2.4 were used to make a 5X stock of TBE buffer in a final 2 L volume of distilled water. TBE buffer was diluted to a 0.5X working concentration prior to use using distilled water.

Table 2.4. TBE buffer reagents.

Reagent	Supplier	Cat. no.	Weight (g)
Tris base	Sigma	10708976001	55
Boric acid	Sigma	B6768	55
EDTA	Sigma	EDS	5.84

2.1.6.2 DNA markers

Commonly used DNA markers are listed in the table below.

Table 2.5. DNA size markers.

DNA marker	Included markers	Supplier	Cat. no.
100 bp	100-1500 bp	Promega	G2101
100 bp	50-1500 bp	Cleaver	CSL-MDNA-100BP
1 Kb	250-10000 bp	Promega	G5711
1 Kb	100 – 10000 bp	Cleaver	CSL-MDNA-1KBPLUS

2.2 Methods

2.2.1 Bioinformatic tools

2.2.1.1 UCSC genome browser

UCSC genome browser was used to visualise genomic arrangements of transcripts (e.g. isoforms, non-coding RNAs), identify potential regulatory domains using ENCODE data (based on histone marks and conservation patterns), identify insertions of transposable elements, profile genetic variation (e.g. SNPs, CNVs) and to obtain DNA sequences for regions of interest in both human and mouse genomes for primer design [203]. Genome assemblies used are specified throughout. All primer sequences are listed in appendix A, with reference numbers specified throughout.

2.2.1.2 Primer design for amplification of gDNA and cDNA templates

For amplification of gDNA using regular PCR, UCSC genome browser (http://genome.ucsc.edu/) was used to obtain the DNA sequence for the desired genomic target[203]. Sequences approximately 50-100 bp upstream and downstream of the target region were input into Primer3 (http://primer3.ut.ee/) to identify potential oligonucleotide sequences[204]. For amplification of cDNA using RT-PCR or qPCR, mRNA sequences were obtained and primers were designed within separate exons to enable detection of genomic DNA contamination (Figure 2.1). Alternatively, for qPCR, predesigned primers listed in the PrimerBank database (https://pga.mgh.harvard.edu/primerbank/) were used if validation data including agarose gels, dissociation curves and amplification plots were available [205]. Potential primers were analysed separately using the OligoAnalyzer tool (https://www.idtdna.com/pages/tools/oligoanalyzer) to assess specific criteria[206] (Table 2.6). Suitable primers were then input into BLAT available through UCSC (https://genome.ucsc.edu/cgi-bin/hgBlat) to check each sequence only complimented the desired target sequence. This minimised the occurrence of non-specific products and

increased likelihood of product specificity. A final check input primer sequences into the *in silico* PCR tool available through UCSC to predict amplicon size.

Table 2.6. Criteria used for primer design.

Parameter	Definition	Criteria
Length	Number of base pairs	18 bp – 22 bp
GC content	Percentage of G or C bases	40 % – 60 %
Melting temperature	Temperature at which 50% of primers dissociate from template	55 °C – 65 °c Both primers in pair must have melting temperature within 2 °C range
Secondary structure	Intramolecular interactions producing structures such as hairpins	Secondary structure must have a melting temperature greater than 10 °C below primer melting temperature to ensure it is broken before annealing during PCR
Self-dimer tendency	Likelihood of primer binding to itself	Avoid runs of inverted complimentary bases (e.g. CCTCNNNNAGAA)
Hetero-dimer tendency	Likelihood of forward primer binding to reverse primer	No runs of >3 of same base (e.g. AAA in forward primer and TTT in reverse primer)

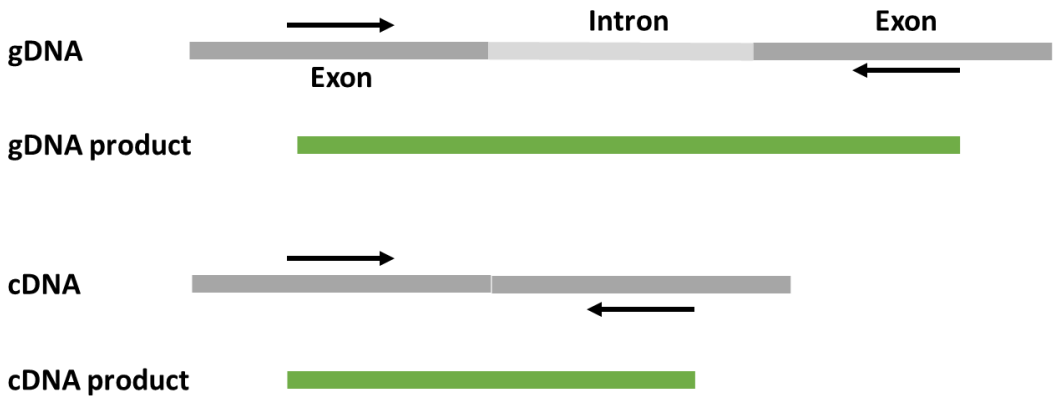


Figure 2.1. Example of primer design for qPCR. (Above) Amplification from gDNA template results in large products containing introns that can be distinguished from smaller cDNA products. (Below) Amplification from cDNA template containing no introns results in small product easily amplified during qPCR cycles.

2.2.2 Polymerase chain reaction techniques

2.2.2.1 Standard PCR/RT-PCR

All PCR reactions were performed using the GoTaq® Hot Start Polymerase enzyme (Promega) according to manufacturer's instructions. An antibody is bound to this enzyme which blocks polymerase activity prior to denaturation at 95 °C for 2 minutes. This can reduce primer dimer and non-specific amplification. This enzyme was used for standard PCR amplification from genomic DNA and reverse transcription PCR (RT-PCR), using cDNA as a template. It was also used for the amplification of inserts used for cloning however sequencing validation was used to confirm the sequence was intact. Specific PCR protocols for each target are listed in appendix A. The details of all materials used for PCR are given in Table 2.7. Thermal cycles were performed using the SimpliAmp™ Thermal Cycler (Applied Biosystems).

Table 2.7. Reagents used in PCR/RT-PCR reactions.

Reagent	Supplier	Cat. no.	Volume (µl)
5X Green GoTaq® Reaction Buffer	Promega	M791B	4.0
$MgCl_2$ (25mM)	Promega	A351B	2.0
Deoxynucleotide mix (10mM)	Sigma	DNTP100A	0.4
Primers (20mM)	Sigma	n/a	0.1
GoTaq® Hot Start Polymerase	Promega	M500A	0.1
gDNA/cDNA template	n/a	n/a	1.0
UltraPure™ DNase/RNase Free Distilled Water	Thermo Fisher Scientific™	10977035	12.3

2.2.2.2 qPCR reaction setup

For all qPCR reactions, GoTaq® qPCR Master Mix (Promega, A6001) was used according to manufacturer's instructions. This contains all necessary components for amplification except primers, water and sample. This mix contains BRYT Green® Dye, which binds to double stranded DNA and exhibits a strong fluorescent signal upon excitation. CXR reference dye was also included in this mix to act as a reference dye between samples. The standard qPCR reaction components are given in Table 2.8. All reactions were set up in triplicate. Thermal cycles and fluorescent readings were measured using the Mx3005P® Real-Time PCR System (Stratagene). All primer sequences and thermal cycling conditions are specified in appendix A and are referred to throughout the text with entry numbers.

Table 2.8. Reagents in qPCR reaction.

Reagent	Supplier	Cat. no.	Volume (µl)
GoTaq® qPCR Master Mix (2X)	Promega	A6001	5.0
Primers (10 µM)	Promega	n/a	0.1
CXR	Promega	C5411	0.1
cDNA	n/a	n/a	2.0
UltraPure™ DNase/RNase Free Distilled Water	Thermo Fisher Scientific™	10977035	2.7

2.2.2.3 Relative quantification using $2^{-\Delta\Delta Ct}$ method

Ct (cycle threshold) values represent the cycle number at which the fluorescence within the sample exceeds the background level. It is representative of the relative abundance in mRNA, therefore the threshold changes within each experiment. For example, a target gene with a Ct value of 30 would contain less cDNA templates than a sample with a Ct value of 20 because it takes more cycles for fluorescence to exceed the background. $2^{-\Delta\Delta Ct}$ is a method to calculate the relative difference in expression of the target gene in experimental groups versus controls, when normalised against the reference gene of choice[207]. This value is representative of fold change of target gene expression in experimental samples versus biological controls. Three Ct values were available for each sample due to setting up triplicate reactions. Any sample with a standard deviation >0.3 between Ct values was visually inspected and the outlier removed. The $2^{-\Delta\Delta Ct}$ was calculated in Microsoft Excel as follows:

1. Calculate the mean Ct (reference) and mean Ct (target) for each sample.
2. Mean Ct (reference) – mean Ct (target) = ΔCt.
3. Calculate the mean ΔCt for all samples in the biological control group.
4. Mean ΔCt of each sample - mean ΔCt (control group) = ΔΔCt.
5. Transform ΔΔCt using equation =2^(-ΔΔCt).

2.2.2.4 qPCR primer validation and efficiency testing

To ensure qPCR data was reliable, we calculated the efficiency of all primer sets used. This is a requirement of the $2^{-\Delta\Delta Ct}$ method. A primer with 100% efficiency should double the number of PCR templates at each cycle, however this is very unlikely therefore it is recommended that primer efficiencies are between 90 – 100 %. To test each primer set, a qPCR test run was set up using a ten-fold serial dilution of cDNA template. The mean Ct values from each dilution were then plotted to generate a standard curve (Figure 2.2). The slope of the curve was then used to calculate the efficiency using the following equation:

$$Efficiency(\%) = \left(10^{\frac{-1}{slope}} - 1\right) \times 100$$

Finally, melt curves were performed after qPCR thermal cycles. A melt curve indicates at which temperature the double stranded DNA product denatures, resulting in loss of the intercalating BRYT Green® Dye. A single distinct peak indicates the presence of a single product and confirmed specificity within the reaction (Figure 2.2). The detection of two peaks indicated the presence of two products or complex secondary structure that did not immediately denature during the melting stage, resulting in some fluorescence being retained at a lower temperature. Any product indicating more than one peak were run on agarose gels to confirm the specificity of the primer set. If multiple products were detected, primers were redesigned.

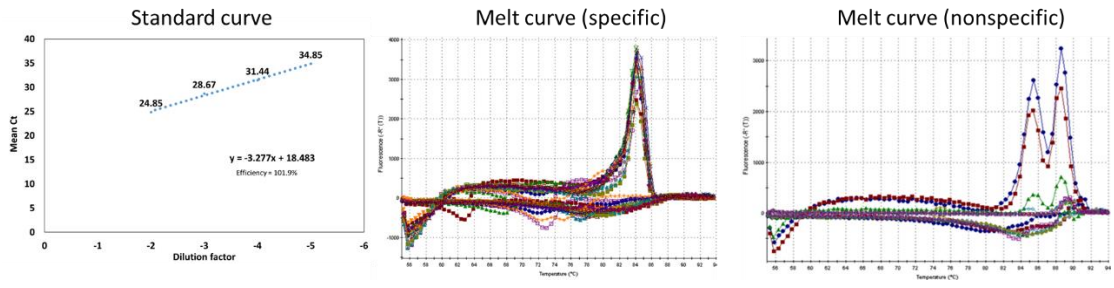


Figure 2.2. Primer validation. (Above) Example of standard curve used to calculate primer efficiency. (Below) Examples of specific (left) and nonspecific (right) melt curves used to measure amplification specificity.

2.2.2.5 Agarose gel electrophoresis

Agarose gels were used to visualise and analyse nucleic acids. 1-2% agarose gels were used throughout, with higher percentage gels being used to generate higher resolution for smaller PCR products (< 500 bp) and lower percentage being used to separate larger PCR products (>1 kb). To make each gel, 0.5X TBE buffer was mixed with UltraPure™ Agarose (Invitrogen) and heated in a microwave until boiling and all agarose was dissolved. This solution was left to cool until no vapour could be seen, before being stained with ethidium bromide solution (Sigma) at a final concentration of 50 µg per 100 ml. Staining enabled visualisation of nucleic acids under UV light. Samples that did not contain any loading dye (e.g. restriction digest products or PCR products with a clear master mix) were mixed with an appropriate volume of 6X Gel Loading Dye Purple (NEB) to aid loading into wells. Sizing of nucleic acids was achieved by loading an adjacent 100 bp or 1 kb DNA marker (Cleaver). All gels were imaged under UV light using the BioDoc-It™ Imaging System (UVP).

2.2.2.6 Capillary electrophoresis

Capillary electrophoresis was performed to resolve PCR products that could not be resolved using agarose gel electrophoresis. We used the QIAxcel automated capillary electrophoresis system (QIAGEN). We used the DNA screening cartridge (QIAGEN, 929004) in conjunction with size markers ranging from 50 bp to 800 bp, and an alignment marker ranging from 15 bp to 1 kb. Samples were screened using default parameters, the AM320 run method and injection time of 10 seconds. Products were verified on a 2.5% agarose gel ran for at least 3 hours at 100V.

2.2.3 Nucleic acid preparation techniques

2.2.3.1 Purification of DNA from agarose gels

PCR and restriction enzyme digest products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega) according to manufacturer's instructions. This is a column-based kit that binds double stranded DNA whilst other components are washed away and discarded. Following analysis of products using agarose gel electrophoresis, products were then excised from the gel with a scalpel under UV light. Exposure to UV light was kept to a minimum to prevent degradation to the sample. Gel pieces were dissolved in the supplied column binding solution at 60°c, before being passed through the DNA binding column. All samples were eluted in 30 µl of nuclease free water (Invitrogen) to produce the highest possible DNA concentration without compromising on efficiency of DNA recovery from the column. All samples were assessed for quantity and quality before being used for sequencing or cloning. PCR products, digested inserts/plasmids that were used for cloning were purified and utilised within 24 hours of amplification to prevent loss of 3' ends required for ligation.

2.2.3.2 Miniprep DNA for subcloning applications

To verify plasmids contained the correct inserts following cloning, plasmid DNA was required for use in diagnostic restriction enzyme digests, PCR and DNA sequencing. Bacteria colonies previously grown on agar plates were selected. A small amount of DNA was required for therefore colonies were transferred into a 5 ml suspension of LB growth media supplemented with 1% antibiotics, and grown overnight in a shaking incubator at 225 rpm at 37°C. 2 ml of each culture was pelleted, providing enough cells for sufficient DNA yield whilst leaving enough culture for further growth stages and freezing of cells stocks. Cultures were stored for maximum of 1 week at 4°c. Plasmid DNA was then purified from the cell pellet using the Wizard® Plus SV Miniprep DNA Purification System (Promega) according to manufacturer's instructions. All samples were eluted in 30 µl of nuclease free water (Invitrogen) to produce the highest possible DNA concentration without compromising on efficiency of DNA recovery from the column. All plasmid resources were stored at -20 °C unless stated otherwise.

2.2.3.3 Maxiprep DNA for transfection in human cell lines

To use plasmid DNA for transfection into human cell lines (for use in reporter gene assays) a large amount of DNA was required. A 50 µl aliquot of bacteria was transferred from remaining suspension cultures (previously validated) into an Erlenmeyer flask containing a 100 ml suspension. The total 100 ml volume was pelleted. Plasmid DNA was purified from the cell pellet using the Plasmid Maxi Kit (QIAGEN) according to manufacturer's instructions. Following ethanol precipitation, the DNA pellet was air dried for 10 minutes to allow evaporation of residual ethanol before being resuspended in 200 µl of nuclease free water (Invitrogen). All plasmid resources were stored at -20 °C unless stated otherwise.

2.2.3.4 Genomic DNA purification from human cell lines

Genomic DNA from human cell lines was purified using the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma) according to manufacturer's instructions. Cells were first dissociated from plates using methods described in section 2.2.5.1. Cells were pelleted by centrifugation at 150 x rcf for 5 minutes at room temperature. Pellets were washed in room temperature PBS (Invitrogen) before being lysed using the supplied lysis buffer. DNA was eluted in 200 µl of nuclease free water (Invitrogen). All samples were assessed for quantity and quality before being used for PCR.

2.2.3.5 RNA extraction from human cell lines

For RNA extraction from human cell lines, the Monarch® Total RNA Miniprep Kit (New England Biolabs) was used following manufacturer's instructions. This column-based purification kit and was used instead of TRI Reagent™, as it removed the possibility of losing small RNA pellets during ethanol precipitation and ensured a sufficient yield of RNA from small cell numbers. All cell lines used were adherent, therefore cells were washed twice with PBS, prewarmed to 37 °c, to ensure complete removal of cell culture medium. Cells were then lysed directly in the plates in which they were grown using the supplied lysis buffer. This was to prevent any changes to RNA profiles or RNA degradation resulting from trypsinisation [208]. Lysed cell samples were first passed through a genomic DNA removal column before collecting the flow through containing the RNA. The flow through was then passed through an RNA binding column and treated for 15 minutes with the supplied DNase I (whilst bound to the column) to remove any residual genomic DNA that may have been carried over. RNA was eluted in 50 µl of nuclease free water (Invitrogen) to produce the highest possible RNA concentration without compromising on efficiency of RNA recovery from the column. All samples were assessed for quantity before being used for cDNA synthesis.

2.2.3.6 Preserving RNA in harvested DRG

All mouse DRG that were harvested for cDNA analysis were preserved to prevent degradation of RNA prior to extraction. DRG were first washed in PBS to remove any excess blood and then submerged in at least 500 µl of RNAlater (Sigma). This is a solution that penetrates tissues, stabilises RNA and inhibits the activity of RNase to preserve the RNA profile of the sample. All tissues were stored overnight at 4 °C to enable RNAlater to fully penetrate tissues before being frozen at -20 °C until RNA extraction (samples were not stored for longer than 1 month prior to RNA purification).

2.2.3.7 RNA purification from DRG

DRG harvested from Thy1-YFP mice (section 2.1.4.1) had previously been stored in RNAlater® (Sigma) were washed twice in PBS (Invitrogen) to remove any traces from the sample and were placed into an RNase free Eppendorf tube. Tissues were submerged in the smallest volume of TRI Reagent® (Sigma) appropriate for the size of the tissue sample (1 ml per 50 – 100 mg of tissue). TRI Reagent® is a solution that lyses cells/tissues, inhibits the activity of RNase and dissociates nucleoprotein complexes enabling the isolation of RNA, DNA and proteins from a biological sample. Tissues were then homogenised using a handheld electronic pellet mixer (VWR). Samples were then left at room temperature for 15 minutes to allow complete cell lysis and dissociation on nucleoprotein complexes. All centrifugation steps were at 12000 x g at 4°c. Samples were spun for 10 minutes to separate lipids (present in clear supernatant) resulting from accidental carryover of fatty tissue. Once the supernatant was removed, chloroform (Sigma) was added in a 10:1 ratio (chloroform: TRI Reagent®). Samples were shaken vigorously and spun for 15 minutes to separate the mixture into three phases (Figure 2.3). The aqueous phase was harvested for RNA precipitation and the remaining phases were stored at -20 °C for future isolation of gDNA and/or proteins. An equal volume of isopropanol (Sigma) was added to each sample and allowed to stand for 10 minutes. Samples were spun for 10 minutes to precipitate the RNA. The RNA pellet was then washed in 1 ml of ethanol (Sigma) and spun again for 5 minutes. The RNA pellet was then allowed to air dry for 10 minutes before being resuspended in 50 µl of nuclease free water (Invitrogen). All samples were assessed for quantity and quality before being used for cDNA synthesis. Prior to cDNA synthesis, all samples were treated with DNaseI (Thermo Scientific) according to manufacturer's instructions. This is an endonuclease that degrades any remaining genomic DNA potentially carried over from phase separation.

2.2.3.8 RNA purification from brain

RNA was isolated from the right hemisphere of the brain of C57BL/6 mice (section 2.1.4.2) using the RNeasy Lipid Tissue Mini Kit (Cat. No. 74804, Qiagen). Genomic DNA (gDNA) was removed from the preparation using a gDNA eliminator column and DNaseI treatment as described in previous section. RNA quality was assessed using a Microplate Spectrophotometer with Take3 plates (Epoch). RNA was reverse transcribed into cDNA using the High-Capacity cDNA Reverse Transcription Kit (Cat. No. 4368814, Applied Biosystems™).

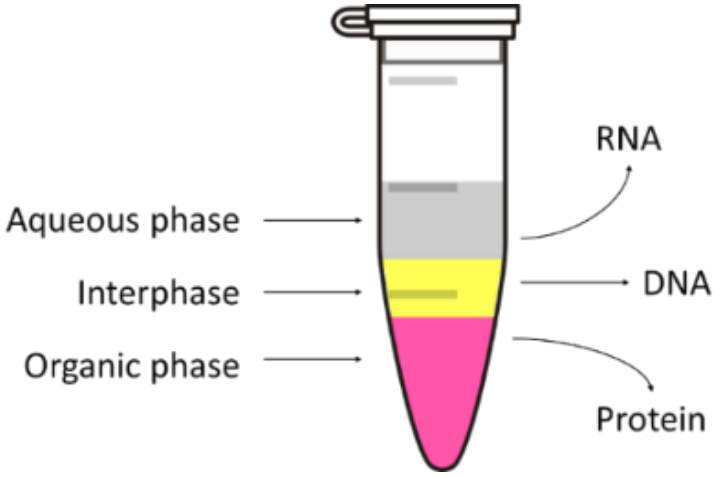


Figure 2.3. Phase separation of biological samples using TRI reagent®. Centrifugation of biological samples following addition of TRI Reagent® and chloroform results in separation of cellular components into 3 phases. The upper aqueous phase contains RNA, the central interphase contains genomic DNA and the lower organic phase contains protein.

2.2.3.9 cDNA synthesis

To generate cDNA templates suitable for gene expression applications such as RT-PCR and qPCR, reverse transcription was performed on up to 1 µg of purified RNA to generate single-stranded cDNA. All RNA samples within specific experiments were normalised to the same concentration (typically 100 ng/µl, however this did vary between experiments) prior to reverse transcription to ensure the same quantity of cDNA was produced for gene expression analyses, and efficiency of the reverse transcription reaction was not affected. RNA templates were reverse transcribed using the GoScript™ Reverse Transcriptase Kit (Promega) according to manufacturer's instructions. Specific volumes of reagents are given in Table 2.9. The reaction was supplemented with both oligo(dT) and random primers to increase the likelihood that all PCR targets could be detected, regardless of being located at the 3' or 5' of the transcript (Figure 2.4). Following reverse transcription, all cDNA samples were diluted 1:10 with nuclease free water and stored at -20°c.

Table 2.9. Reagents used to synthesise cDNA in reverse transcription reaction.

Reagents	Supplier	Cat. no.	Volume (µl)
GoScript™ 5X Reaction Buffer	Promega	A500B-C	4.0
$MgCl_2$ (25mM)	Promega	A3551	3.0
$Oligo(dT)_{15}$ primers (500 µg/ml)	Promega	C1101	1.0
Random primers	Promega	C1181	1.0
dNTPs (10mM)	Promega	U1205	3.0
Recombinant RNasin® Ribonuclease Inhibitor (40 u/µl)	Promega	N2111	0.5
GoScript™ Reverse Transcriptase	Promega	A5002	1.0
RNA (100 ng/µl)	n/a	n/a	3.0
UltraPure™ DNase/RNase Free Distilled Water	Thermo Fisher Scientific™	10977035	3.5

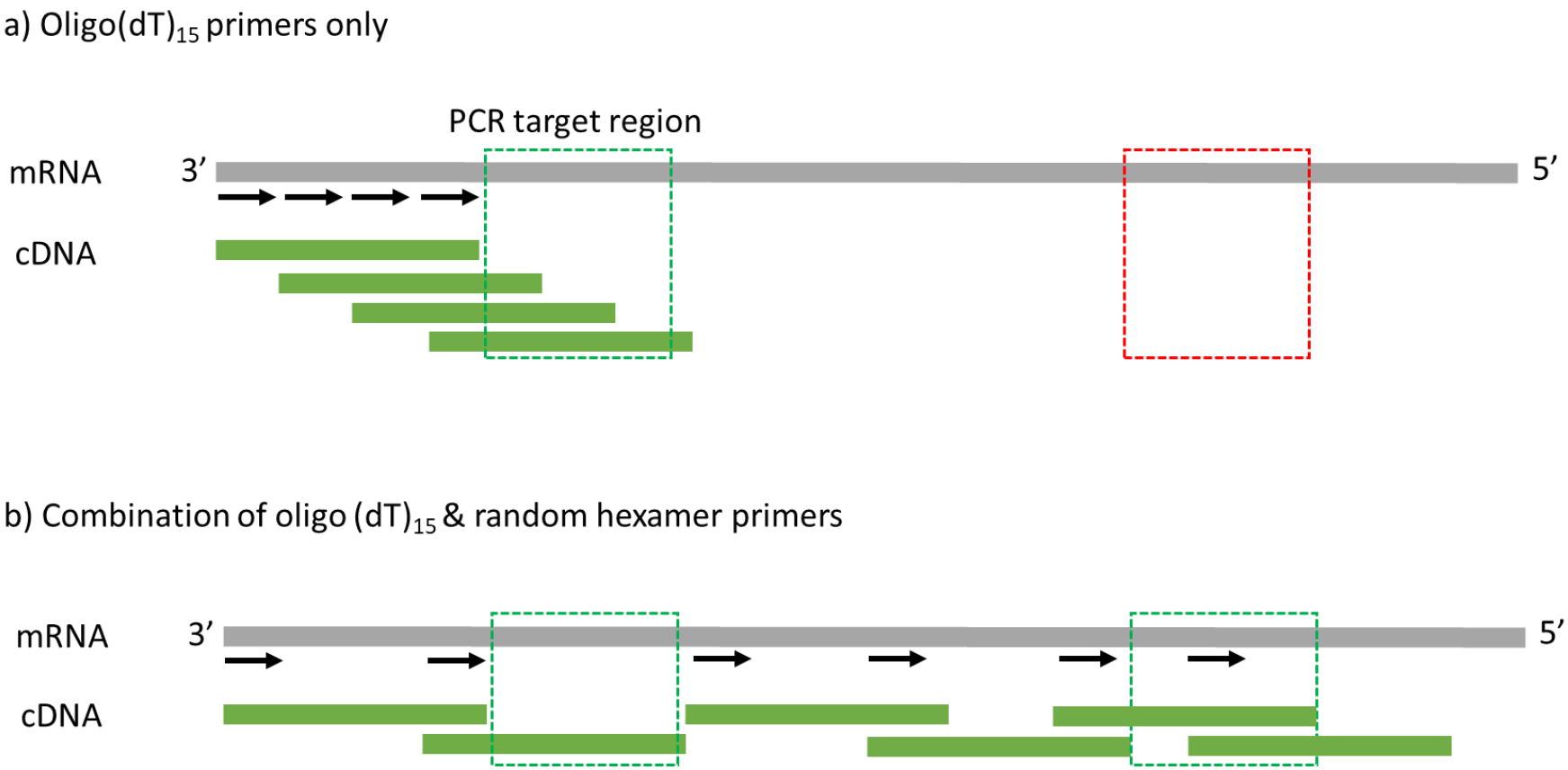


Figure 2.4. Generation of cDNA templates based on different primer combinations. (Above) Use of only oligo(dT) primers begins reverse transcription at the 3’ end of the mRNA. Reverse transcriptase enzyme drops off the template after several thousand bases so for long transcripts, cDNA may not be generated that includes the 5’ end of the transcript. If the PCR target is at the 3’ end of the mRNA, this is not an issue and products will be amplified (green box), however a target at the 5’ end may not be amplified and could falsely indicate no expression (red box). (Below) Use of a combination of oligo(dT) and random hexamer primers ensures that cDNA products are generated across the entire length of the mRNA transcripts and increases the likelihood that targets at the 5’ end will be amplified during PCR (green boxes).

2.2.3.10 Quantification and quality controls

Quantity and quality of nucleic acids were assessed using the NanoDrop 8000 Spectrophotometer (Thermo Scientific) prior to downstream processing (e.g. cDNA synthesis) or use in assays (e.g. genotyping, cloning or sequencing). 260/280 ratios reflect the purity of a sample and are used to assess protein contamination. 260/230 ratios can be used to assess residual phenol from TRI reagent-based methods of nucleic acid purification or residual guanidine carryover from column-based purification methods. Ratios should be above 1.8 for DNA sample and 2.0 for RNA samples. Agarose gel electrophoresis was also used to visualise RNA integrity. 1 µg of total RNA was mixed with loading dye and ran on an agarose gel to visualise the presence of distinct 18S and 28S rRNA bands. Clear bands without smearing indicated RNA was intact and had not been degraded by contaminating RNase.

2.2.4 Molecular cloning techniques

2.2.4.1 Ligation of inserts into intermediate vector pCR®II

The desired insert to be cloned was amplified using PCR. The quantity of PCR product required to perform a 1:1 ligation reaction was calculated using the equation below. Ligation reactions were always set up in a 3:1, 5:1 and 10:1 (insert:vector) ratio to cover a range of potential ligation efficiencies. Ligations were incubated overnight at 14 °C using the SimpliAmp™ Thermal Cycler (Applied Biosystems). All vector maps are given in appendix B.

$$x\ (ng\ of\ PCR\ product) = \frac{(bp\ of\ insert) \times (50ng\ of\ vector)}{size\ of\ vector\ (bp)}$$

Inserts were initially ligated into the Dual promoter pCR®II vector using the TA Cloning® Kit (Invitrogen, K207020) according to manufacturer's instructions. The reagents used are given in Table 2.10. Ligation reactions were incubated overnight at 14 °C prior to transformation. This method relies on the presence of 3' A-overhangs in the PCR product to anneal to the 3' T-overhangs in the vector (Figure 2.5). Taq enzymes add a 3' A-overhangs to PCR products hence were always used for cloning purposes. 3' A-overhangs are also prone to removal over time therefore PCR products were always purified and ligated immediately after amplification. The vector backbone contained the LacZα gene fragment, making it suitable for X-gal blue white screening of positive clones following transformation.

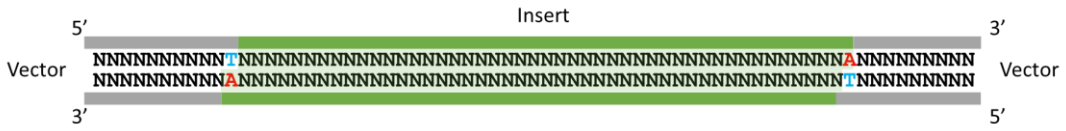


Figure 2.5. Schematic of TA cloning. DNA insert previously amplified using PCR highlighted in green. Note the addition of A nucleotides at the 3' end of the insert, generating A overhangs. The vector is lineated with 5' T overhangs therefore complimentary base pairing between the A and T aids ligation of insert into the vector.

Table 2.10. Reagents used for ligation of insert into intermediate vector pCR®II.

Reagent	Supplier	Cat. no.	Volume (µl)
PCR product	n/a	n/a	X
5X Express Link™ T4 DNA Ligase Buffer	Invitrogen	46300018	2.0
pCR®II vector (25 ng/µL)	Invitrogen	K207040	2.0
ExpressLink™ T4 DNA Ligase (5 u/µl)	Invitrogen	15224041	1.0
UltraPure™ DNase/RNase Free Distilled Water	Thermo Fisher Scientific™	10977035	final volume 10 µl

2.2.4.2 Ligation of inserts into expression vector pGL3-P

Upon successful cloning of the desired inserts into the intermediate vector, inserts were digested out of the vector , purified and ligated into the reporter gene vector pGL3-Promoter (pGL3-P) (Promega, E1751) using T4 DNA ligase (NEB, M0202) according to manufacturer's instructions. Ligation reactions were setup using reagents given in Table 2.11. and incubated overnight at 14°C. DNA mass to be used in ligation was calculated using equation in previous section.

Table 2.11. Reagents used for ligation of insert into expression vector pGL3-P.

Reagent	Supplier	Cat. no.	Volume (µl)
T4 DNA Ligase Buffer (10X)	NEB	B0202S	2.0
Vector DNA (50ng)	n/a	n/a	X
Insert DNA	n/a	n/a	Y
T4 DNA ligase (40 u/µl)	M0202S		1.0
UltraPure™ DNase/RNase Free Distilled Water	Thermo Fisher Scientific™	10977035	final volume 20 µl

2.2.4.3 Transformation of chemically competent cells

To generate enough plasmid DNA for applications such as subcloning and *in vitro* use, they were propagated using bacterial host cultures. Transformation of bacteria was conducted using Subcloning efficiency™ DH5α competent cells (Invitrogen, 18265017) according to manufacturer's instructions. 200 µl of cells were spread onto agar plates containing ampicillin. Plated were then inverted and incubated overnight at 37°C. Clones containing the intermediate vector were selected for using ampicillin and/or X-gal blue white screening. Clones were picked, grown and plasmid DNA was purified. If the DNA to be transformed was highly repetitive or had a high GC content, Endura™ Chemically Competent Cells (Lucigen, 60240) were transformed following manufacturer's instructions. Colonies were selected and grown up for DNA purification using methods described in section 2.2.4.5.

2.2.4.4 Blue white screening

Blue-white screening was utilised when the cloning site of the vector contained the gene encoding the enzyme β-galactosidase, which metabolises X-gal into 5-bromo-4-chloro-indoxyl. This metabolite dimerises to produce a blue pigment called 5,5'-dibromo-4,4'-dichloro-indigo, turning the colony containing the nonrecombinant vector blue. Colonies containing recombinant plasmids have an insert in the cloning site, therefore the β-galactosidase gene is disrupted and consequently do not turn blue in the presence of X-gal. This enabled quick screening of colonies transformed with recombinant plasmids. If blue white screening was required, plates were coated/spread with X-gal prior to plating of bacteria. 20 µl of X-gal solution at a concentration of 50 mg ml^{-1} (Promega, V3941) was mixed with 50 µl of LB broth and spread over prepared agar plates. Plates were then incubated at 37 °C for 30 mins to allow the X-gal to soak into the plates before use.

2.2.4.5 Plasmid DNA purification from bacteria cultures

Colonies were picked using a sterile pipette tip, which was then placed into a 5 ml starter culture and grown overnight at 37 °C in a shaking incubator rotating at 225 rpm. Bacteria cultures were grown in a LB broth suspension supplemented with 100 µg/ml ampicillin. Sufficient growth was indicated by the culture turning turbid overnight. A 100 ml culture was then set up using a 50 µl aliquot from the starter culture if a large amount of high-quality DNA was required for transfection. The smaller 5 ml cultures were used to isolate small amounts of DNA used for diagnostic purposes or subcloning (Figure 2.6). This was carried out overnight at 37 °C in a shaking incubator rotating at 225 rpm. Bacteria were then pelleted in a centrifuge at 4 °C at 500 x rcf for 5 minutes. LB broth was removed from the pellet. Pellets were then used for plasmid DNA purification using either miniprep or maxiprep. All DNA samples were assessed for quantity and quality before being used for downstream applications.

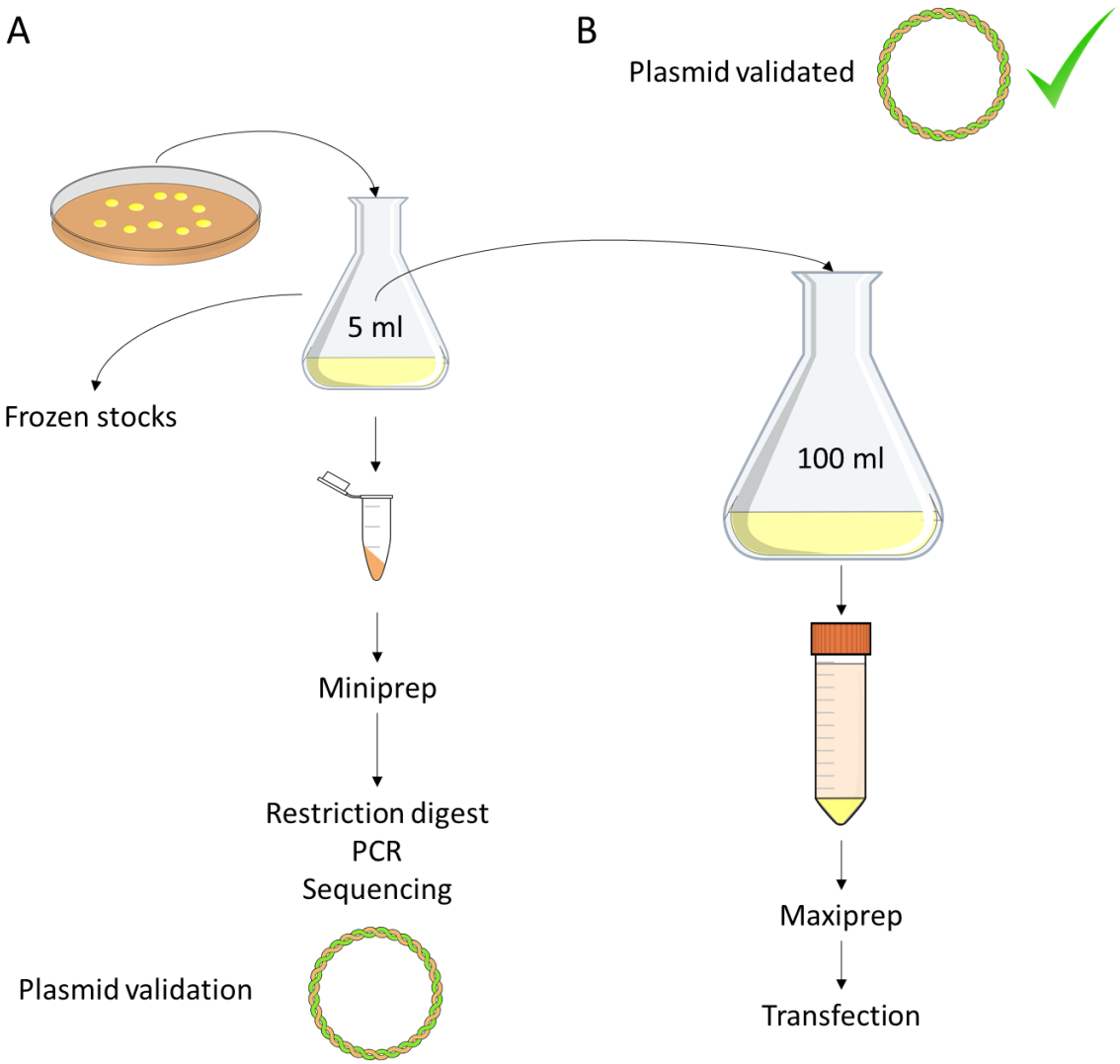


Figure 2.6. Methods used to purify plasmid DNA constructs from bacterial cultures for use in different downstream applications. (A) Methods used to select colonies for growth in small volume cultures and purification of plasmid DNA using miniprep protocol; suitable for plasmid validation applications including restriction digest, PCR and sequencing. (B) Methods used to purify large amount of high-quality plasmid DNA using maxiprep; suitable for transfection *in vitro.*

2.2.4.6 Restriction enzyme digests

Restriction enzyme digests were used as a diagnostic tool to confirm the presence and orientation of cloned inserts, and to enable subcloning of inserts from intermediate to expression vectors. Enzymes for diagnostic digests were chosen based on having one cut site near one end of the DNA insert and one cut site in the backbone of the SVA. Agarose gel electrophoresis can then be used to assess product size and determine which plasmids are correct. Enzymes for subcloning were chosen based on ability to generate compatible nucleotide overhangs. The basic set up of a restriction enzyme digest is given in Table 2.12. Each specific enzyme required a specific buffer; however all incubations were conducted at 37 °C for at least 2 hours to allow complete digestion followed by a heat inactivation step at >65 °C for 15 minutes. Incubations were performed used the SimpliAmp™ Thermal Cycler (Applied Biosystems). Following digestion, samples were run on a 1% agarose gel for 1 hour at 100V to enable enough separation for diagnostic or subcloning use. If subcloning was required, digests were set up in triplicate and loaded into the same well to enable sufficient yield following gel purification.

Table 2.12. Reagents used in restriction enzyme digests.

Reagents	Supplier	Cat. no.	Volume (µl)
Multicore buffer (10X)	Promega	R9991	2.0
BSA (10 mg/ml)	Promega	R3961	0.2
Enzyme (10 U/µl)	Promega	n/a	0.5
Template DNA (500 ng)	n/a	n/a	X
UltraPure™ DNase/RNase Free Distilled Water	Thermo Fisher Scientific™	10977035	Final volume 20 µl

2.2.5 Human cells line techniques

2.2.5.1 Maintenance and passaging of cell cultures

All new cell lines introduced to the laboratory (from nitrogen stores or collaborators) were kept in isolation and screened for mycoplasma using the MycoAlert™ Mycoplasma Detection Kit. This was repeated every six months. All cell lines were grown at 37 °C in 5% CO_2. All media used contained phenol red, which changes colour according to pH. Complete media was replaced when pH change induced a yellow colour in the culture, or more frequently. When cells reached 70 – 80% confluency, they were passaged to allow continuation of growth. All reagents were prewarmed to 37 °C prior to use. To passage, cells were washed twice with PBS before being dissociated using trypsin to detach cells from the growth surface. After addition of trypsin, cells were incubated for 5 minutes before trypsin was inactivated by adding 10X the volume of cell culture media. Cells were triturated gently to ensure complete dissociation. A proportion of the cell suspension was taken into a new flask. Cell were typically passaged using a 1:10 or 1:20 ratio, dependent upon application.

2.2.5.2 Freezing and thawing cells in liquid nitrogen stores

Cells to be used as future resource were stored in liquid nitrogen. Prior to freezing, cells were grown to 70-80% confluency in a T25 flask to ensure sufficient numbers could be frozen and recovered upon thawing. To freeze cells, cultures were dissociated using trypsin and pelleted to remove excess media. Cells were resuspended in 5 ml of freezing media. 1 ml of cell suspension was transferred to cryovials (at least 3 replicates) before being gradually frozen overnight at -80 °C using a Mr Frosty™ Freezing Container (ThermoFisher Scientific). 24 hours later, cells were transported on dry ice to the liquid nitrogen stores for long term storage. To thaw cells, vials were removed from liquid nitrogen and thawed in a water bath at 37 °C for at least one minute. Cells were resuspended in 10 ml of prewarmed media and pelleted to remove excess freezing media. Pellets were resuspended in an appropriate volume of fresh media and added to an appropriate size flask.

2.2.5.3 Transfection of reporter gene constructs into human cell lines for luciferase assays

Transient transfection of plasmid DNA into human cell lines was carried out using TurboFect™ Transfection Reagent (Thermo Scientific™) according to manufacturer's instructions. This is a solution of cationic polymers in water that form stable complexes with plasmid DNA, protecting it from degradation in culture, an ensuring efficient delivery into cells. High quality DNA is essential for effective transfection therefore only DNA purified using maxiprep protocols was used. 24 hours prior to transfection, cells were seeded into a 24-well plate at a density of 10^5 cells/well, containing 1 ml of media. Volumes of reagents used to generate the transfection mix are given in Table 2.13. pRL-TK is a *Renilla* luciferase control vector used for normalisation during luciferase assays and was included alongside constructs being tested. 100 µl of complete transfection reagent was added to each well in a dropwise manner. A minimum of 4 biological repeats were used for each construct being tested. A full media change was conducted 4 hours after transfection reagent was added to minimise toxicity and cell death. Cells were then incubated for 48 hours before luciferase assays were conducted.

Table 2.13. Components used in complete transfection reagent.

Reagent	Volume (µl)
Serum-free media	100
TurboFect™ Transfection Reagent	2.0
Plasmid DNA (500 ng)	X
pRL-TK (20 ng/µl)	1.0

2.2.6 Dorsal root ganglia techniques

2.2.6.1 Dissection of DRG from mouse

To obtain dorsal root ganglia for gene expression studies and *in vitro* assays, they were surgically dissected from mice. All mice were culled using cervical dislocation. Mice were culled individually as needed to prevent loss of RNA integrity[209], as the DRG harvest protocol is very time consuming per animal (up to one hour). The dissection protocol has been adapted from one published online at bio-protocol.org (www.bio-protocol.org/e1785), to increase efficiency and further preserve RNA. It is essential that all surgical equipment is clean, sharp and mirror finished tweezers used to pull the DRG away from tissue are not blunt. Surgical equipment is listed in Table 2.14. Dissection procedure is given on next page.

Table 2.14. Surgical equipment required for DRG harvest.

Item	Supplier	Catalogue number
Standard scissors	Fine Science Tools	14002-12
Spring scissors	Fine Science Tools	91500-09
Dumont #7b medical forceps	Fine Science Tools	11270-20
Dumont #5 mirror finish forceps	Fine Science Tools	11252-23

DRG dissection procedure is as follows:

1. The animal is decapitated using standard scissors.
2. The spinal column is immediately dissected using incisions on the ventral side of the animal. All tissue is trimmed away from the spinal column using spring scissors.
3. Using spring scissors, horizontal incisions were made at the neck opening of the spinal column (Figure 2.7). This was performed in an alternating manner, whilst simultaneously lifting the ventral side of the column away using medical forceps to exposed the spinal cord. The spinal cord was then removed with tweezers.
4. Using a dissection microscope, the DRG are apparent situated in between the vertebrate. Using mirror finish surgical tweezers, the DRG were gently pulled from their sockets and trimmed at the base to release the bulb (mirror finish tweezers must be used to limit tissue damage). If DRG were required for *in vitro* applications, all visible DRG along the length of the spinal column were harvested to maximise cell numbers. If DRG were required for RNA extraction, only larger lumber DRG were harvested as these are easier to isolate than the smaller cervical DRG whilst providing sufficient RNA yield.
5. DRG were washed in PBS before being transferred to their storage or transport solutions. DRG required for gene expression studies were placed immediately into RNA later prior to freezing or immediate RNA extraction. DRG required for *in vitro* assays were placed in Hank's Buffered Saline Solution prior to dissociation into a single cell suspension.

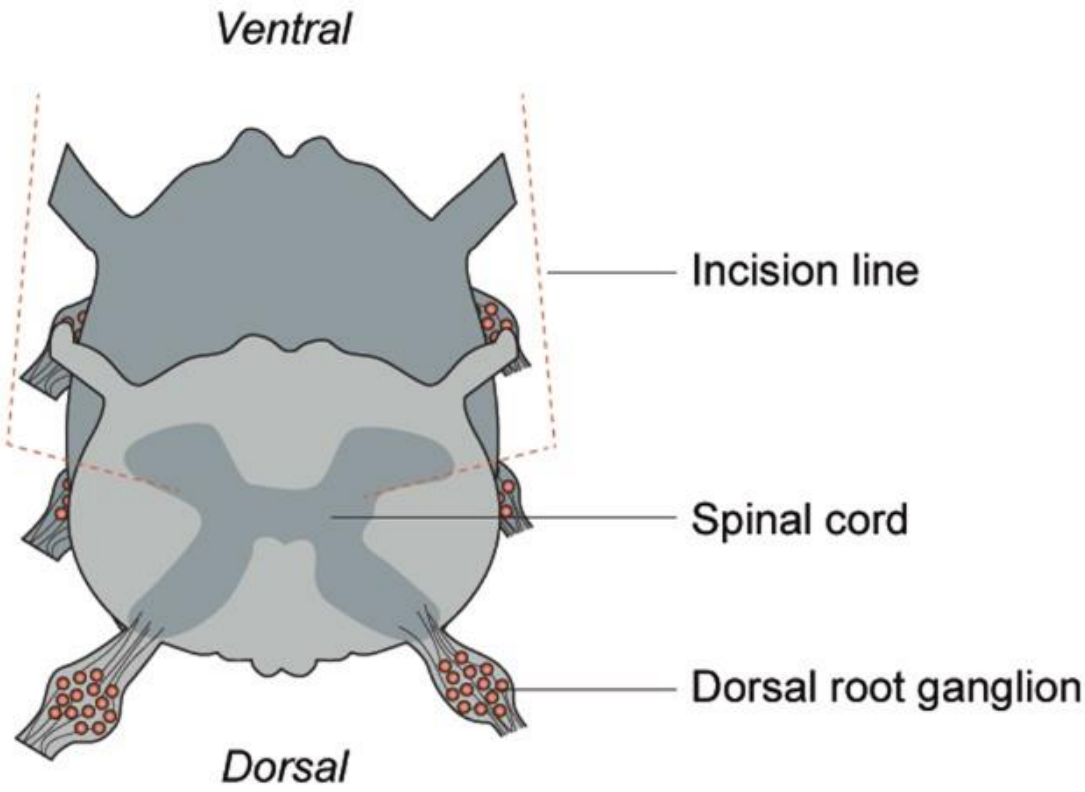


Figure 2.7. Schematic of DRG dissection. Image taken from www.bio-protocol.org/e1785

2.2.6.2 Dissociation of DRG into single cell suspension

To dissociate the DRG tissue into a single cell suspension that could be used for in vitro assays the following steps were taken:

1. Prepare collagenase solution and preheat for 10 minutes at 37°c.
2. Allow DRGs to settle to the bottom of the tube and gently remove the supernatant.
3. Wash the DRGs using 5 ml of dissociation solution (repeat 3 times).
4. Remove 3 ml of dissociation solution. Add 1 ml of collagenase solution. Incubate for 1 hour at 37°c.
5. Following the incubation, invert to resuspend the DRGs and centrifuge at 160 rcf with a ramp speed of 3.5.
6. Remove the supernatant. Wash DRGs with 5 ml of dissociation solution (repeat 3 times).
7. Remove the supernatant and add 1.5 ml of dissociation solution to the DRGs.
8. Triturate the DRGs using an unmodified Pasteur pipette approximately ten times until the suspension looks homogenous.
9. Triturate the DRGs gently using the modified Pasteur pipette approximately ten times.
10. If the mixture has large pieces of debris remaining, pass through a cell strainer.
11. Centrifuge the suspension for 5 minutes at 160 rcf with a ramp speed of 3. This will prevent the cell pellet from being dispersed as the centrifuge accelerates and decelerates.
12. Remove the supernatant and resuspend the cell pellet in 5 ml of DRG culture media. Centrifuge again.
13. Remove supernatant and resuspend the cell pellet in 1 ml of DRG culture media and perform cell count using a haemocytometer.

2.2.6.3 Transfection of DRG using nucleofection

DRG were transfected using nucleofection. This method uses electroporation, which involves applying an electrical field to primary cells to increase the permeability of the cell membrane and allow the introduction of DNA. The Mouse Neuron Nucleofector™ Kit (Lonza, VPG-1001) was used according to manufacturer's instructions. The DRG yielded from a single mouse following purification was approximately 60,000 cells. This protocol is most suited for cell numbers >100,000 therefore DRG were pooled from multiple animals (at least n=3). Prior to setup, cuvettes were prechilled to 4 °C to prevent the sample from overheating during nucleofection. Per sample, 100,000 cells were counted and suspended in 100 µl of Nucleofection® Solution and combined with 1 µg of the prepared vector. 2 µg of pmaxGFP® vector was used as a transfection control. The solution was then transferred to a cuvette (care taken to avoid bubbles, as this disrupts current) and placed into the nucleofection device. The Nucleofector™ 2b device was used to administer the voltage, using programme 0-003. Cells from each transfection were then resuspended in 4 ml of fresh prewarmed media and 500 µl (approximately 25,000 cells) was plated into the wells of a 24-well plate. After 2-4 hours (depending on cell adherence) media was replaced. Media was replaced again after 24 hours. Luciferase assays were performed 48 hours after transfection.

2.2.7 Luciferase reporter gene assays

Luciferase assays were conducted 48 hours after transfection using the Dual-Glo® Luciferase Assay System (Promega, E2920) according to manufacturer's instructions. The transfected plasmid, either pGL3-P or pGL3-B, encode firefly luciferase, which is measured to assess for changes in transcriptional activity dependent on the insert. Cells were cotransfected with pRL-TK, a plasmid that encodes Renilla luciferase, which functions as an internal control to normalise data between reactions. All luciferase reactions were set up in quadruplicate. All luciferase readings were made using the GloMax® 96 Microplate Luminometer and the "DLRwithTwoInjections" protocol. This instructs the machine to automate the addition of luciferase assay reagents (described in the manufacturers protocol) to enable high throughput and take fast readings. To calculate differences in luciferase activity, a ratio of the RLU (relative light units) values from firefly luciferase and Renilla was calculated. The fold change was then calculated against pGL3-P as the control group.

2.2.8 CRISPR

2.2.8.1 gRNA design

UCSC genome browser (https://genome-euro.ucsc.edu) was used to retrieve DNA sequences 250 bp flanking each side of the SVA. These sequences were input into a bioinformatic tool (http://crispr.mit.edu/) used to select ideal 20 bp nucleotide sequences upstream of PAM sites. This tool assessed sequences for potential off target effects and ranked them according to sequence specificity. Multiple guides were selected to test efficiency *in vitro*. To clone these sequences into the Cas9 vector and generate a functional gRNA molecule, a double stranded insert was needed which required several sequence modifications to generate the necessary overhangs. The PAM sequence was removed from each oligonucleotide before the complimentary sequence was generated. CACC was added to the 5' end of the original oligo and an AAAC was added to the 5' end of the complimentary oligo. This generated overhangs that were compatible with the Cas9 vector during the cloning stage. To generate double stranded inserts to encode a functional gRNA molecule, both complimentary oligonucleotides were annealed together. 6 µl of both oligonucleotides (100 µM each) were combined with 83 µl of nuclease free water and 5 µl of T4 DNA Ligase Reaction Buffer (NEB). This mix was heated at 95 °C for 5 minutes and allowed to cool back to room temperature gradually. Annealed guides were stored at -20 °C.

2.2.8.2 Golden gate cloning of gRNA inserts into pSpCas9(BB)-2A-GFP

The golden gate method was used to clone guides into the as described in Ran *et al.*[210]. This method relied on using BbsI to digest a small insert out of the plasmid and replacement with the guides. If the desired insert is ligated with the vector, the BbsI recognition site is lost and the enzyme will not cut, if the original insert is religated with the vector and BbsI can cut again. 50 ng of annealed guides were mixed with 150 ng of pSpCas9(BB)-2A-GFP, 2µl of 10X ligase buffer (NEB), 1µl of BbsI-HF (NEB), 1µl of T4 ligase (Promega) and nuclease free water to a final volume of 20ul. This was incubated for 10 cycles at 37 °C for 5 minutes and 16 °C for 10 minutes, then 37 °C for 30 minutes and 80 °C for 20 minutes to heat inactivate enzymes. 2 µl of the ligation mix was used to transform DH5α *E. coli* competent cells. Colonies were selected and DNA was miniprepped for screening. A restriction digest was performed using BbsI-HF (NEB), according to manufacturer's instructions, to identify an insertion of annealed guides and therefore elimination of the BbsI restriction site. Clones containing the desired insert were grown overnight and DNA was maxiprepped. The presence of each guide within the plasmid was verified by Sanger sequencing from the U6 promoter and (primers are listed in appendix A; see entry 7 for gRNA sequences and entry 8 for U6 sequencing primer).

2.2.8.3 Transfection of CRISPR-Cas9 constructs into human cells

Transfection was conducted using TurboFect™ (as described in section 2.2.5.3) however the transfection reagent was prepared as shown in Table 2.15 to enable the transfection of two constructs encoding different gRNAs.

Table 2.15. Transfection reagent for dual transfection of CRISPR-Cas9 constructs.

Reagent	Supplier	Volume (µl)
TurboFect™ Transfection Reagent	ThermoFisher Scientific (R0533)	2.0
DMEM, high glucose	Gibco™	100
CRISPR-Cas9 vector with 5'gRNA (250 ng/µl)	n/a	2.0
CRISPR-Cas9 vector with 3' RNA (250 ng/µl)	n/a	2.0

2.2.8.4 Cell sorting using flow cytometry

48 hours post transfection, GFP expressing cells were isolated using the BD FACSAria™ III (BD Biosciences). Prior to cell sorting, cell culture medium was removed, and cells were washed with PBS. Cells were then dissociated into a single cell suspension using trypsin. Cells were then pelleted at 150 rcf for 5 minutes (at room temperature) before being resuspended in 3 ml PBS supplemented with 1% FBS in Falcon® Round-Bottom Polystyrene Tubes.

2.2.8.5 Growth of clonal cell lines

Transfected cells were seeded at a density of 1000 cells into 10 cm dishes. This enabled the growth of single cells into colonies without risking contamination from other colonies. When colonies were grown enough to be seen by eye (this could take up to 2 weeks), they were scraped using a p20 pipette tip and transferred into individual wells in a 96-well plate format. Clonal cell lines were grown until 70 – 80% confluent and split into two 96-well replicate plates at a 30:70 ratio. The less confluent plate was used for growth and the more confluent plate was used for gDNA extraction and PCR to check for the presence of a modification.

2.2.8.6 Identification of modified cell lines

Initial genotyping was performed using direct cell lysis as a template for PCR. Cells were washed twice using PBS and lysed directly in wells using DirectPCR Lysis Reagent (Viagen, 301-C) supplemented with Proteinase K solution (Sigma, P4850) according to manufacturer's instructions. Plates were sealed with parafilm and placed in a rotating hybridisation over set at 55 °C overnight. Proteinase K was heat inactivated at 65 °C for 45 minutes prior to PCR. PCR primers and thermal cycles used to screen for deletions are listed in appendix A (entry 9). PCR reactions were set up as described in Table 2.16. To generate enough cells for gDNA purification, RNA purification and creation of frozen stocks, cells were grown up in a T25 flask. Validation of positive clones was repeated using purified gDNA from clones grown in T25 flasks, followed by DNA sequencing across predicted breakpoints using the same primers that were used for PCR.

Table 2.16. PCR used to check for CRISPR modification.

Component	Volume (µl)
5X Green GoTaq® Flexi Buffer	4.0
$MgCl_2$ (25mM)	2.5
dNTPs (10mM)	0.4
Forward primer (20µM)	0.1
Reverse primer (20µM)	0.3
GoTaq® Hot Start Polymerase	0.3
Crude lysate/gDNA (10ng/µl)	1.0
Nuclease free water	11.4

2.2.9 DNA sequencing

DNA sequencing was conducted externally by Source Bioscience, UK (www.sourcebioscience.com). Sequencing primers are listed in appendix A and specific primers are referred to throughout results chapter with entry numbers. 5 µl of plasmid DNA (per read) normalised to 100 ng/µl with nuclease free water was used per read. 5 µl of PCR products (per read) normalised to 10 ng/µl with nuclease free water. All sequencing primers were prepared at a concentration of 3.2 µM in a volume of at least 5 µl per read. Electropherograms were visualised too assess quality using Chromas (https://technelysium.com.au/wp/chromas/) (Figure 2.8). FASTA files were exported to BLAT to check for the correct genomic region in UCSC genome browser.

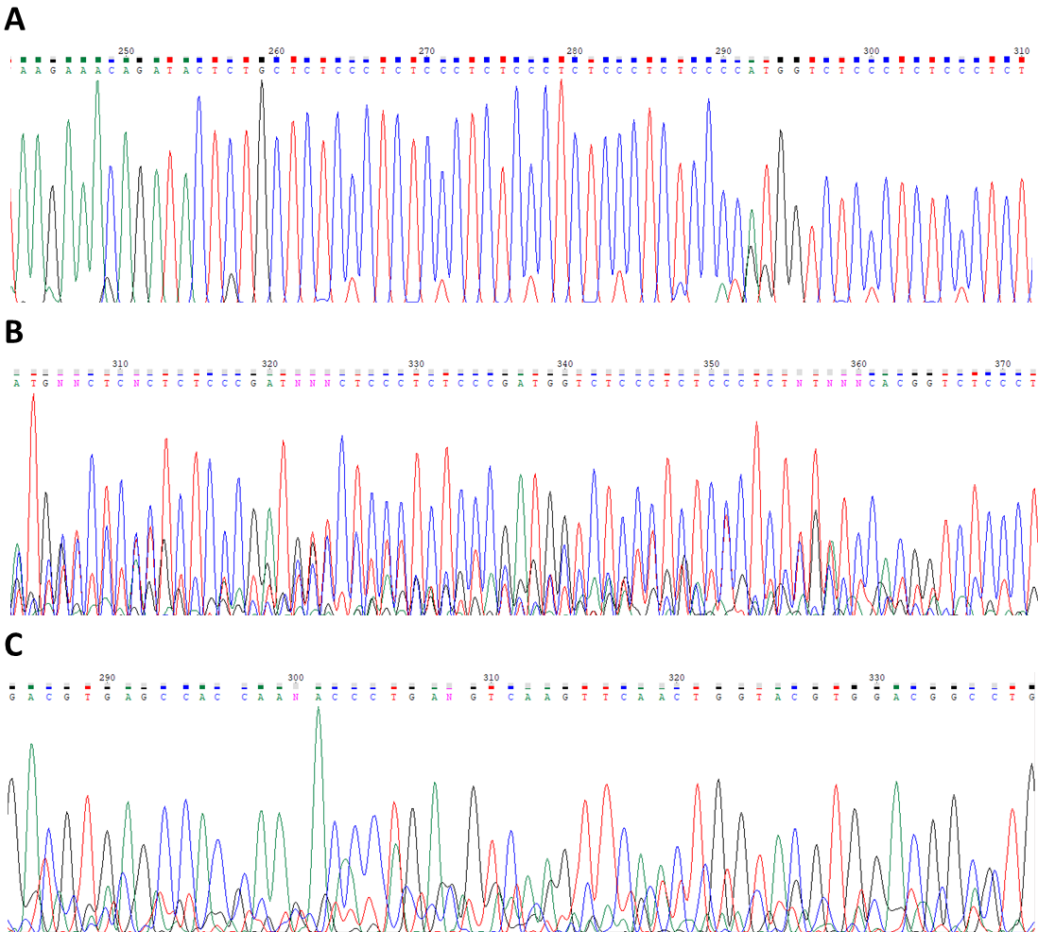


Figure 2.8. Examples of electropherograms used for quality control. (A) Good quality sequencing trace. (B) Poor quality sequencing trace from contamination. (C) Poor quality sequencing trace from low sample concentration.

2.2.10 MBD pulldown

2.2.10.1 Isolation of methylated and unmethylated DNA

Methylated and unmethylated DNA was isolated from cells using the CpG MethylQuest kit (Millipore) according to manufacturer's instructions. 500 ng of total genomic DNA was used. DNA was sonicated to <500 bp fragments using the S220 focused-ultrasonicator (Covaris). Sonicated DNA was then concentrated using Agencourt AMPure XP Beads (Beckman Coulter) using a 1:1.1 ratio and eluted in 25 µl water. DNA concentration was then measured using the Qubit dsDNA BR Assay Kit (Fischer Scientific). Capillary electrophoresis was used to validate the size of DNA fragments following sonication.

2.2.10.2 Densitometry to estimate relative DNA methylation

PCR products were loaded onto a 1% gel and ran for 1 hour at 100V. All products were loaded onto the same gel therefore all bands were imaged under the same conditions, reducing the effects of variability. The image was uploaded into ImageJ (https://imagej.nih.gov/ij/), a freely available software commonly used to relatively quantify the intensity of PCR product bands using densitometry. Each band on the gel was converted into a peak. The area of both peaks, representing methylated and unmethylated DNA from each sample, were combined to generate a total value. The percentage of methylated DNA was calculated by dividing the area of the methylated band by this total value.

Chapter 3.

Characterisation of a human specific SVA as a transcriptional regulator of *TRPV1* and *TRPV3*

3.1 Introduction

Mobilisation of transposable elements (TEs) leads to novel insertions which contributes to genome evolution and genetic variation between humans[211,212]. New insertions can be co-opted for new functions which drives species-specific gene regulation[213]. Within individual genomes, TEs are mediators of epigenetic mechanisms that influence gene regulation in response to stimuli, for example in response to stress[214]. Analyses of non-LTR retrotransposons across the genome have revealed a positive correlation between SVA density (SVAs per Mb) and gene density (genes per Mb), indicating that SVAs preferentially insert into genic (gene dense) regions of the genome and thus may carry more capacity to exert greater effects on gene regulation[15,215]. L1 and *Alu* density do not positively correlate with gene density[10]. The reason behind this is currently unknown, however this potentially highlights a preference for SVAs to integrate into open chromatin[204]. More recently, bioinformatic analysis by Gianfrancesco *et al.* identified 30 loci that were enriched for SVA insertions (based on SVA copy number Mb), highlighting certain regions of the genome that were overrepresented for SVAs compared to other gene dense loci that were already expected to be enriched for SVA insertions[216]. The most enriched regions were KRAB-ZNF (zinc nuclease finger) gene clusters, however of interest to this project was a 1 Mb region on chromosome 17 (chr17:3000001-4000000), named chr17p13.2, which encodes pain associated genes *TRPV1* and *TRPV3*[148,150-152]. All SVAs were SVA D and 5 out 6 were human specific, suggesting this region has undergone significant modification in recent human evolutionary history[15,25].

TRPV1 and *TRPV3* mRNA and proteins are coexpressed in human DRG, yet TRPV3 expression is not conserved. In mice, *TRPV1* is well documented in DRG but *TRPV3* is not expressed in DRG and current literature describes it being predominantly in keratinocytes (with no evidence to support expression in DRG) [9,10,152,174]. *TRPV3* expression has further been documented in a variety of human tissues (but not in mice) including brain, placenta, skin,

small intestine, spinal cord, stomach, testes and trachea, suggesting a wider role for *TRPV3* in human physiology compared to the mouse[6,9,10,152,174]. This suggests that the mechanisms which regulate tissue-specific expression of *TRPV1* and *TRPV3* are different in mice and human genomes - which may be important in the development of therapeutics targeting these channels by utilising preclinical mouse models.

CREs are defined as non-coding DNA sequences that regulate transcription of neighbouring genes. Regulatory sequences and mechanisms controlling *TRPV1* and *TRPV3* gene expression have not been well characterised in the human genome. A recent epigenomic study using primate genomes identified that SVAs are enriched in human specific CREs in the liver[217]. This highlights SVAs as ideal candidates to explore human specific gene regulation patterns in relation to differences observed between preclinical mouse models and humans. Closer inspection of the SVA enriched locus encoding *TRPV1* and *TRPV3* using UCSC genome browser (hg19) revealed an SVA insertion at the intergenic region between *TRPV1 and TRPV3*. The SVA was directly adjacent to an evolutionary conserved region (ECR) which we predicted to function as a CRE. The proximity of the SVA to the ECR could potentially disrupt conserved function via multiple mechanisms including transcription factor binding, DNA methylation and/or changes to secondary structure (e.g. G4-quadruplex), which may also potentially be modified in response to stimuli[26]. This highlighted the SVA as a candidate CRE and epigenetic modulator at this locus.

In this chapter we provide an overview of the *TRPV1* and *TRPV3* locus which forms the basis of the research conducted in this chapter and chapter 4. We demonstrate using reporter gene assays that the SVA is functional as transcriptional regulator in the human cell line HEK293. We further demonstrate its capacity to function as a transcriptional regulator in a sensory neuron background, using murine DRG primary cell culture. Finally, we present an optimised protocol to delete candidate regulatory element using CRISPR and show in

modified HEK293 cell lines (carrying SVA deletions) direct changes in total *TRPV3* mRNA expression - highlighting this SVA as a human specific CRE.

3.2 Hypotheses and aims

The hypotheses in this chapter are that the ECR identified at the intergenic region between *TRPV1* and *TRPV3* functions as a CRE, contributing to regulation of *TRPV1* and *TRPV3* mRNA expression. To further this, the human specific SVA insertion adjacent to the ECR is also functional as a CRE. We propose that the proximity of the human specific SVA to this ECR may exert species-specific regulatory properties which may contribute to human specific regulation of *TRPV1* and *TRPV3* mRNA expression.

The aim of this chapter was to characterise if the ECR and SVA at *TRPV1* and *TRPV3* functions as a CRE and was achieved by completing the following:

1. Generation of reporter gene constructs.
    a. Generate ECR reporter gene construct and assess function as transcriptional regulator in human cell line.
    b. Generate SVA reporter gene constructs and assess function as transcriptional regulator:
        i. Test in human cell line.
        ii. Test in mouse DRG primary cell culture.
2. Generation of CRISPR modified cell lines to assess function of DNA sequence as CRE affecting *TRPV1* and *TRPV3* mRNA regulation.
    a. Optimise CRISPR protocol to delete entire regulatory domains
    b. Generate clonal cell lines containing ECR deletion
    c. Generate clonal cell lines containing SVA deletion
    d. Optimise RT-PCR/qPCR assays to measure *TRPV1* and *TRPV3* protein coding mRNA transcript variants and total mRNA levels.

3.3 Results

3.3.1 Assessing transcriptional regulatory domains using luciferase reporter gene assays

3.3.1.1 Bioinformatic analysis of TRPV1 and TRPV3 locus using UCSC genome browser

Tang et al. recently published an article describing the contribution of mobile DNA to the human genome[15]. Using the supplementary data, we identified the average number of human specific SVA insertions per Mb was 0.53 genome-wide, and 0.95 for chromosome 17. This highlighted chromosome 17 as enriched for SVAs compared to other chromosomes, however this was expected as it is the second most gene dense chromosome in the genome at 38 genes/Mb (behind chromosome 19 at 53 genes/Mb)[15].

Gianfrancesco *et al.* recently published an analysis of the distribution of SVAs in the human genome[216]. Analysis of this supplementary data identified that there were 6 SVAs at chr17p13.2 (Figure 3.1A); 5 were human specific. This highlighted a 10-fold enrichment in human specific SVA insertions compared to the genome average and a 5-fold increase compared to the chromosome 17 average. We further analysed the supplementary data from Gianfrancesco *et al.* and split the enriched loci based on SVA subclass. We identified that chr17p13.2 was the only enriched region in the genome that contained SVA D exclusively (Figure 3.1C).

A recent comparison of repetitive elements in primate genomes identified that SVA D are the most active transposable element in the human genome[218]. This highlighted a region of the genome that had undergone a significant restructure in recent evolutionary history therefore was a target for further investigation. We conducted a full literature search of the genes encoded within this Mb to identify potential roles in pain pathways (Supplementary table 1). Several genes were reported to have roles in pathways relevant to pain however the literature was not as comprehensive as compared to *TRPV1* and *TRPV3*. We focused on the SVA at the coding region for *TRPV1* and *TRPV3* (Figure 3.1B).

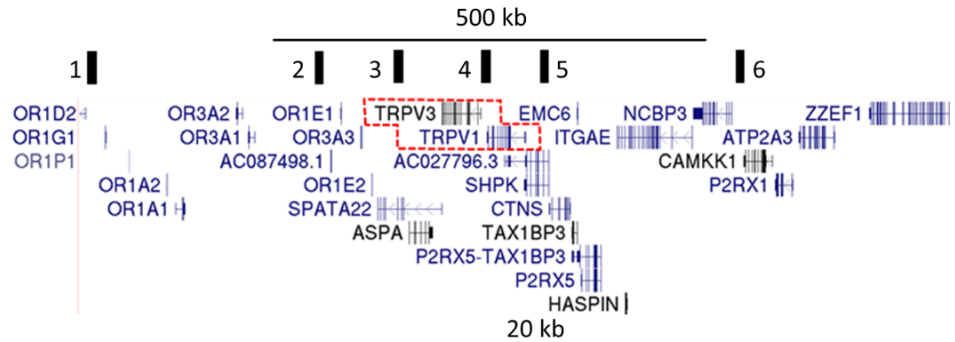
We used UCSC genome browser to explore the intergenic region in more detail. The intergenic region is 7452 bp in length and contains other repetitive DNA sequences that are conserved with other primate species (Figure 3.2A). 73% of the intergenic sequence was composed of TEs. The most common elements were hominid-specific *Alus* (36% of total sequence) followed by LINEs (12% of total sequence), none of which were full length. The SVA was the largest transposable element in this region and accounted for 19% of the intergenic sequence. The lack of conservation tracks corresponding to the SVA establish it as a human-specific repeat, therefore it emerged in the genome following divergence from the last common ancestor with chimpanzee.

The SVA at the intergenic region between *TRPV1* and *TRPV3* is encoded on the antisense strand of DNA, like *TRPV1* and *TRPV3* (Figure 3.2A). The SVA is approximately 27 kb downstream of the nearest *TRPV1* TSS and 400 bp downstream of the *TRPV1* 3'UTR. The SVA is 5.7 kb upstream of the nearest *TRPV3* TSS. The SVA is listed in RepeatMasker as 1385 bp in length (chr17:3466973-3468357), however closer inspection of the SVA DNA sequence and flanking DNA sequences at the 5' and 3' end identified that the 5' TCTCCC hexamer domain extends 17 bp upstream of the 5' end, therefore the full SVA sequence is 1402 bp in length (chr17:3466973-3468374), containing an additional 3 TTTCCC hexamers than described in the RepeatMasker annotation. The manually annotated SVA sequence was used throughout this project.

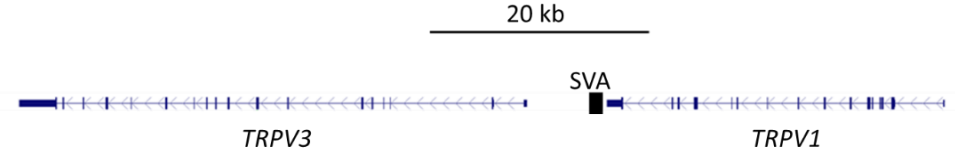
The SVA is adjacent to an evolutionary conserved region (chr17:3,466,258-3,466,820, hg19) which is conserved with the mouse (Figure 3.2A). This sequence is 600 bp in size and contains a MIR (Mammalian-wide Interspersed Repeat). MIRs are the most ancient family of TE in the human genome, enriched for TFBS and have been shown to act as conserved CREs in K562 and He-La cell lines[219]. ENCODE data overlaid across this ECR showed histone marks H3K27Ac and H3K4Me1 and ChIP-seq data highlighting numerous transcription factors –

which was suggestive of transcriptional enhancer/repressor activity[220]. No histone data was available for the SVA as they are larger repetitive sequences which are difficult to map and often excluded from analyses. The data suggested this ECR may be functional as a CRE. However, we hypothesised that its modulation in humans may differ from the mouse due to the adjacent SVA insertion, forming the basis for this chapter.

A



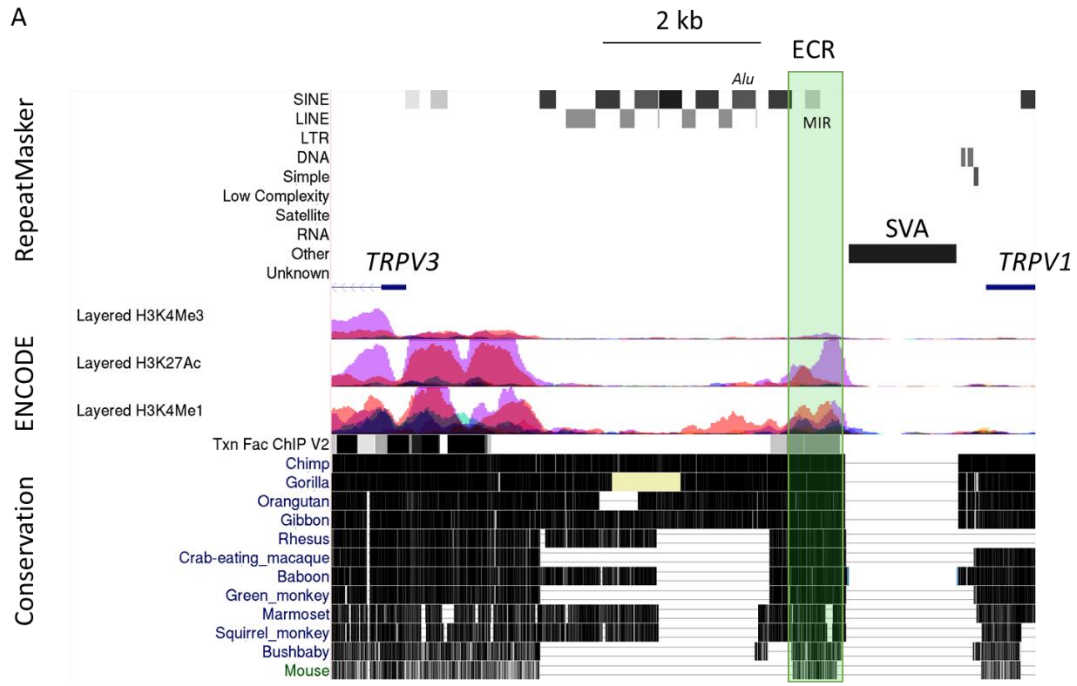
B

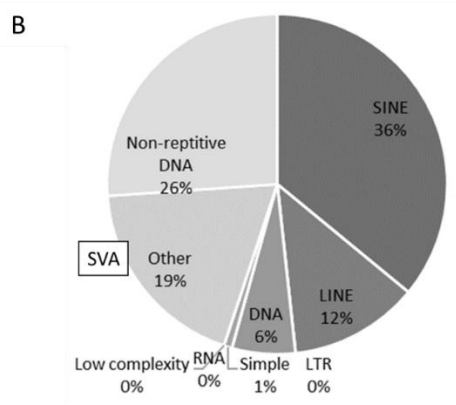


C

SVA	Subclass	Location	Start position	End position	Size (bp)	Human specific
1	SVA_D	Intergenic *OR1D2/OR1G1*	3015148	3016938	1790	Yes
2	SVA_D	Intergenic *OR3A4P/OR1E1*	3273928	3276203	2275	Yes
3	SVA_D	Intronic *SPATA22*	3367208	3369090	1882	Yes
4	SVA_D	Intergenic *TRPV3/TRPV1*	3466972	3468371	1399	Yes
5	SVA_D	Intronic *SHPK*	3537000	3537574	574	No
6	SVA_D	Intergenic *C17orf85/CAMKK1*	3756886	3758383	1497	Yes

Figure 3.1. SVA insertion identified at the intergenic region between *TRPV1* and *TRPV3*. (A) SVA enriched megabase at chr17:3000001-4000000 encoding TRPV1 and TRPV3 (hg19). (B) Overview of TRPV1 and TRPV3 locus containing SVA insertion at intergenic region. (C) Table summarising SVA insertions.





Class	Count	Combined size (bp)	Percentage of sequence
SINE	10	2680	36%
LINE	6	918	12%
LTR	0	0	0%
DNA	2	448	6%
Simple	1	59	1%
Low complexity	0	0	0%
RNA	0	0	0%
Other	1	1402	19%
Non-repetitive DNA		1945	26%

Figure 3.2. Overview of intergenic region between *TRPV1* and *TRPV3*. (A) UCSC screenshot of chr17:3,461,289-3,468,740 (hg19) highlighting all repetitive DNA elements listed in RepeatMasker. The SVA is annotated. The ECR is highlighted in green box which contains MIR element and is marked by Txn Fac ChIP v2 and histone marks from ENCODE data. The conservation track shows conserved regions in black and a break across the SVA highlighting specificity to the human genome. (B) Summary of transposable elements including count, combined size and percentage of total intergenic sequence.

3.3.1.2 Generation of reporter gene constructs

To test the regulatory potential of the ECR and the SVA, we cloned them separately into the reporter gene construct pGL3-P (section 2.2.4). Both targets were amplified from human genomic DNA using PCR (methods section 2.2.2.1). Primer sequences and PCR conditions are listed in appendix A (entry 1 for SVA primers and entry 2 for ECR primers). PCR products were initially ligated into the intermediate vector, pCRII, using the TA cloning strategy described in section 2.2.4.1.

A restriction enzyme digest, using *HindIII* (Promega, R6041), was used to cut the DNA insert from the plasmid backbone to confirm the presence and orientation of inserts in selected clones (section 0) (Figure 3.3A). The empty pCRII vector was cut once therefore generated a linearised product 3971 bp in length (Figure 3.3B). Recombinant pCRII vectors containing the SVA insert were cut twice, generating product sizes of 4831 bp and 1101 bp if the SVA was in the forward orientation, and product sizes 5861 bp and 744 bp if in the reverse orientation. Similarly, for the ECR, if cloned in the forward orientation product sizes were 5029 bp and 139 bp, and 3997 bp and 1171 bp in the reverse orientation (Figure 3.3C). We generated four constructs in total containing the SVA or ECR in forward and reverse orientations (Figure 3.3).

Restriction enzyme digests were then performed on recombinant plasmids to cut out the insert, and also cut out a small sequence in the reporter gene vector to generate compatible ends that would enable subcloning (section 0). We digested the target insert from the intermediate pCRII vector using SacI (Promega, R6061) and XbaI (Promega, R6181). The reporter gene vector pGL3-P was digested using SacI (Promega, R6041) and NheI (Promega, R6501) to produce compatible ends (Figure 3.4). XbaI and NheI are isoschizomers, which recognise different (but similar) sequences yet generate compatible overhangs. This also prevented the orientation of the cloned inserts changing when subcloned from the intermediate vector to the reporter gene vector.

Sequencing was used to verify the correct inserts and assess any mutations (e.g. indels), as this could affect transcription factor binding compared to the native form. Sequencing primers are listed in appendix A (entry 6).

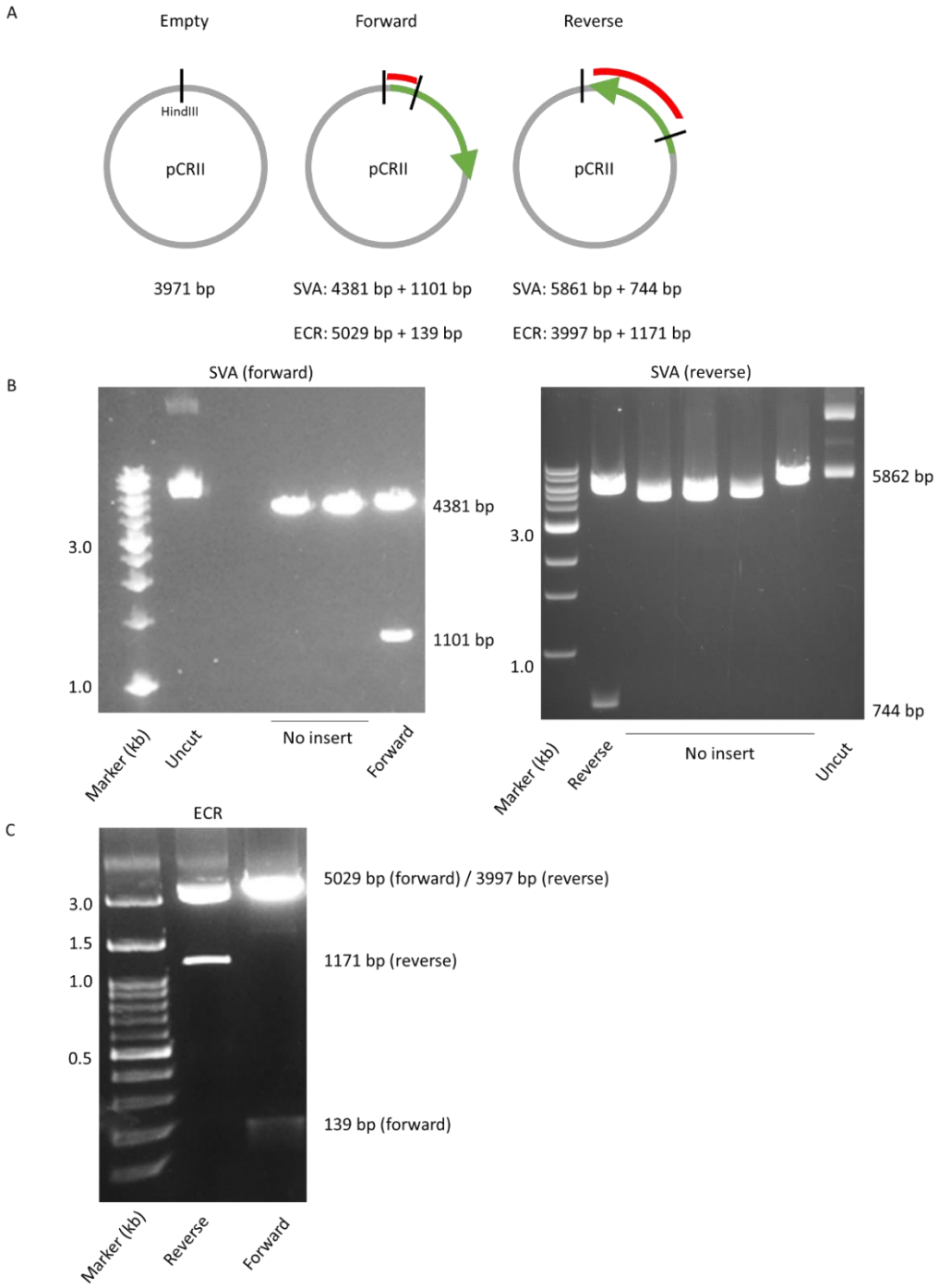


Figure 3.3. Diagnostic restriction enzyme digests showing target sequences cloned in both orientations into intermediate vector pCRII. (A) Schematic showing predicted digested product sizes following digest with *HindIII*. (B) Restriction enzyme digest showing cloned SVA insert in forward orientation (left) and reverse orientation (right). (C) Restriction enzyme digest showing cloned ECR insert in forward orientation (right) and reverse orientation (left).

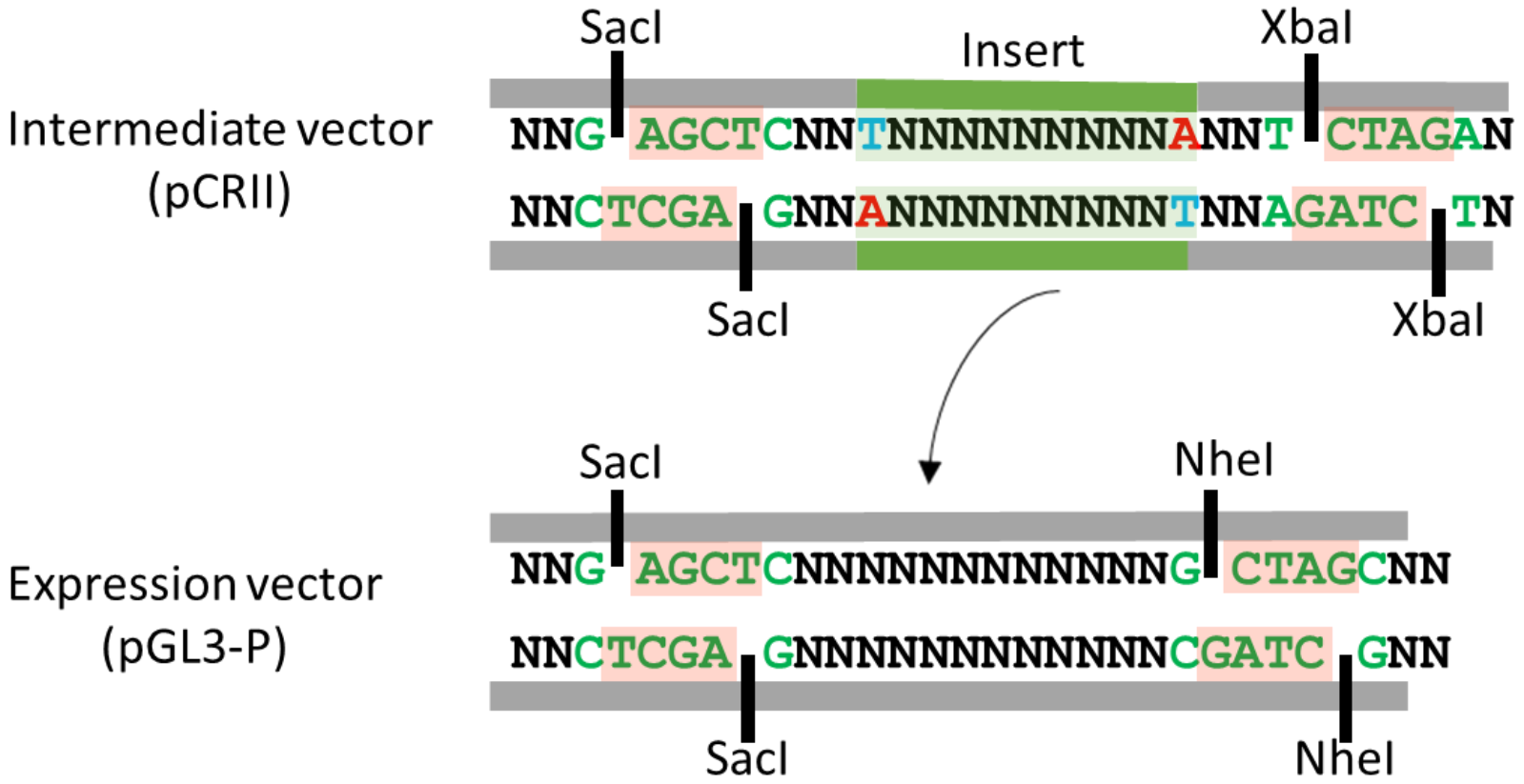


Figure 3.4. Subcloning strategy using restriction enzyme digests to generate compatible nucleotide overhangs. (Above) The intermediate vector pCRII was digested using *SacI* and *XbaI* to produce 5' overhangs. The green insert represents the initial sequence ligated into pCRII using TA cloning. (Below) The expression vector pGL3-P was digested using *SacI* and *NheI* to generate compatible 5' overhangs to enable subcloning.

3.3.1.3 SVA is functional as transcriptional repressor and ECR shows minimal activity in HEK293 cell model

SVA and ECR constructs were transfected into HEK293 cells (section 2.2.5.3) and assessed using a reporter gene assay 48 hours later (Figure 3.6). Using the pmaxGFP plasmid (Lonza) to visualise fluorescence we estimated a transfection efficiency of 80% (Figure 3.5). Any changes in luciferase activity (comparing the recombinant pGL3-P containing the insert against the original empty vector) could be directly correlated with luciferase expression resulting from the presence of the insert, therefore determining regulatory activity *in vitro*.

All assays consisted of 4 biological replicates, which were defined as independent cell cultures that were transfected separately (n=4). Luciferase activity was averaged from triplicate fluorescent readings. Cells were cotransfected with the pRL-TK vector which expresses Renilla luciferase and was used to normalise data between transfections. The basic vector, pGL3-B did not contain a promoter and did not express luciferase, acting as a negative control within the assay.

pGL3-B showed relative luciferase activity approaching zero, indicating no luciferase activity and this assay was functioning correctly (0.03±0.00, n=4, P=0.03) (Figure 3.6). The SVA cloned into pGL3-P in the forward orientation showed a 2-fold decrease in luciferase activity (0.48±0.03, n=4, P=0.03). Similarly, the pGL3-P construct containing the SVA cloned in the reverse orientation showed a 1.6-fold decrease (0.62±0.03, n=4, P=0.03) (Figure 3.6). The reduction in luciferase activity directly correlates with luciferase expression, therefore the insertion of the SVA induced a repressive effect on gene expression in this HEK293 model. Thus, we concluded that this SVA can function as a transcriptional regulatory element. This could potentially be due to transcription factors directly binding the SVA sequence, or perhaps the SVA sequence induced a DNA structure which interfered with transcription factor binding at the SV40 promoter.

The ECR cloned in the forward orientation showed a 1.3-fold increase in relative luciferase activity (1.3±0.02, n=4, P=0.03) (Figure 3.6). This was statistically significant however this was a very minimal increase therefore we could not consider this as biologically relevant in this case. The ECR cloned in the reverse orientation showed a 1.08-fold increase however this was not biologically or statistically significant (1.08±0.05, n=4, P=0.19) (Figure 3.6). From this data we concluded the ECR was showed minimal activity as a transcriptional regulatory domain in this HEK293 model. As this was not tested in any other cell line, we cannot rule out the possibility that this ECR is regulatory.

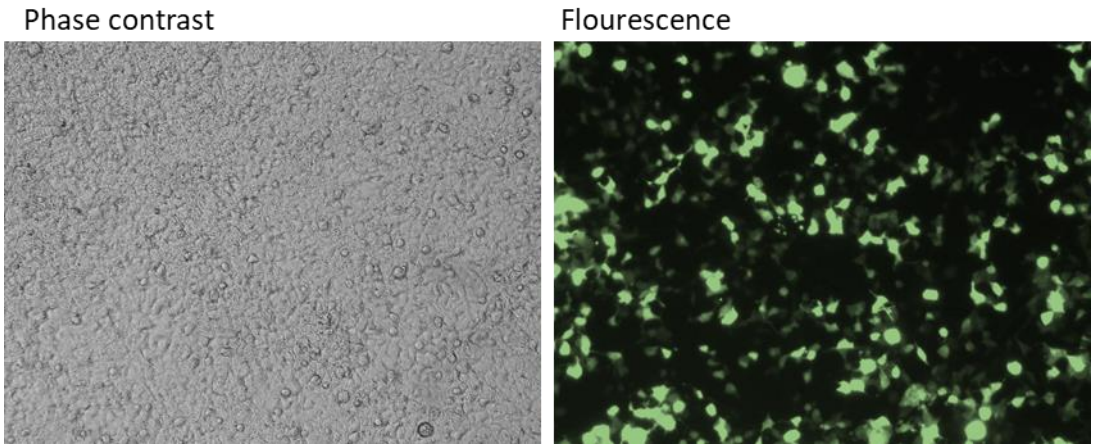


Figure 3.5. Transfection of HEK293 with pmaxGFP. Images showing expression of GFP detected under fluorescent microscope 48 hours post transfection. (Left) Phase contrast images showing cell confluency. (Right) Fluorescent images showing GFP. Estimated transfection efficiency 50%.

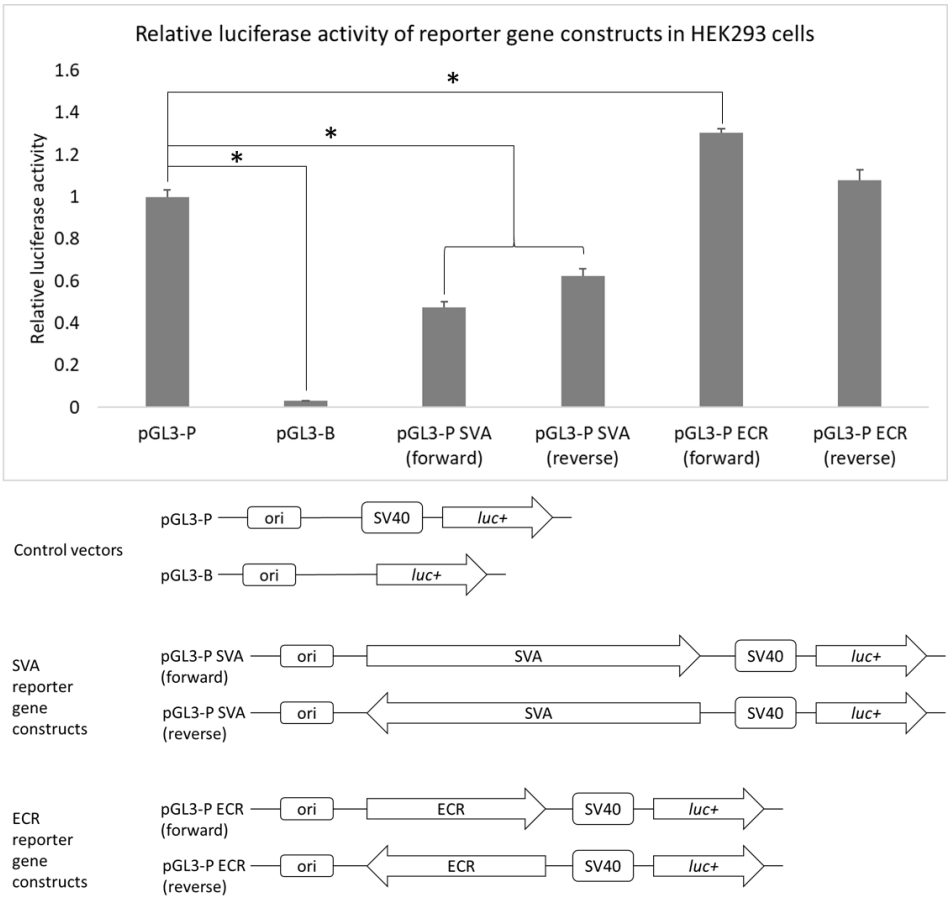


Figure 3.6. Results for luciferase reporter gene assay testing SVA and ECR reporter gene constructs in HEK293. Relative luciferase activity in empty (pGL3-P) vector was compared against baseline (pGL3-B) and reporter gene constructs (pGL3-P containing an SVA or ECR in forward and reverse orientations). Luciferase activity was normalised to the internal control vector (pRL-TK). Error bars represent standard error from the mean. Mann Whitney tests were used to measure the significance of fold change between vectors. Significance p-value <0.05*.

3.3.1.4 SVA is functional as transcriptional repressor in mouse DRG model

Using the same constructs in DRG primary culture it was possible to assess if the SVA could co-opt transcription factors present within sensory neurons and function as a transcriptional regulator in this tissue. It was not possible to obtain human DRG during this project therefore mouse DRG were used as a model. A recent study of mouse and human DRG transcriptomes highlighted a conserved enrichment of transcription factors and supports the use of mouse DRG in this context[6].

Purified DRG cell suspensions were transfected using nucleofection, which is a harsh treatment and resulted in a lot of cell death and low transfection efficiency (estimated <10%) (Figure 3.7). An example of untransfected purified DRG taken from Thy1-YFP mice grown in culture is given in Supplementary figure 1 to highlight culturing conditions were optimal for DRG growth and the cell death observed in this experiment was due to transfection. 48 hours post-transfection, luciferase activity was assayed (Figure 3.8). Due to low cell numbers resulting from cell death, and to keep the number of mice culled to a minimum, it was decided to only test the SVA reporter gene construct in its forward orientation.

pGL3-B showed a decrease in luciferase activity (0.50±0.01, n=4, P=0.03). The decrease in luciferase from pGL3-B was not as great as observed in HEK293 due to the low number of cells present in the assay, and the relatively high level of background fluorescence. A 1.6-fold decrease in luciferase activity from pGL3-P containing the SVA was observed (0.64±0.06, n=4, P=0.03), consistent with the results observed in HEK293. This indicates that the SVA is responsive to transcription factors within sensory neurons and can function as a transcriptional regulator in DRG.

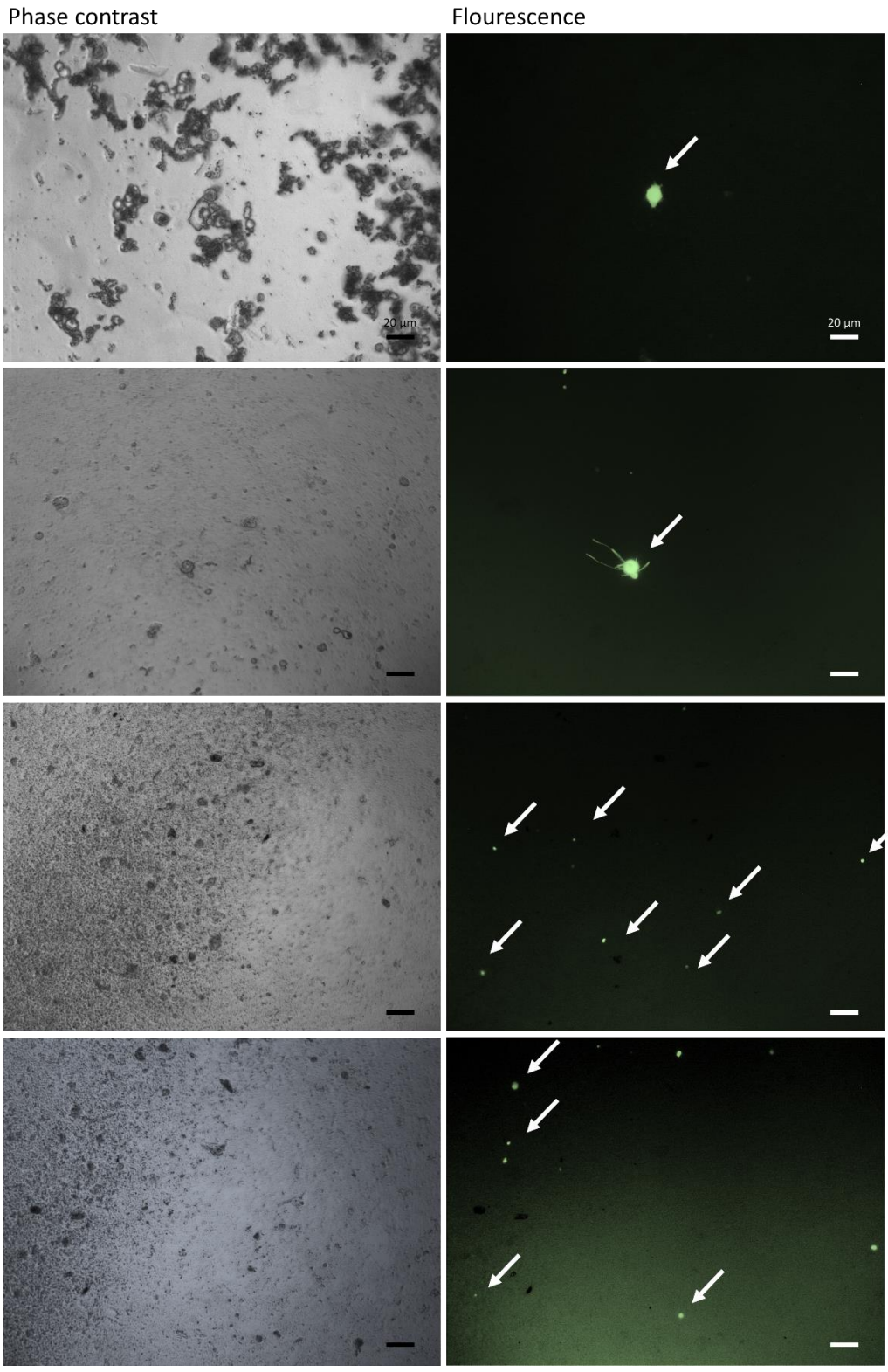


Figure 3.7. Examples of DRG cultures 24 hours following transfection with pmaxGFP. (Left) Phase contrast images showing DRG neurons and cell debris resulting from cell death. (Right) Flourescent images showing transfected DRGs expressing GFP in the same field of view.

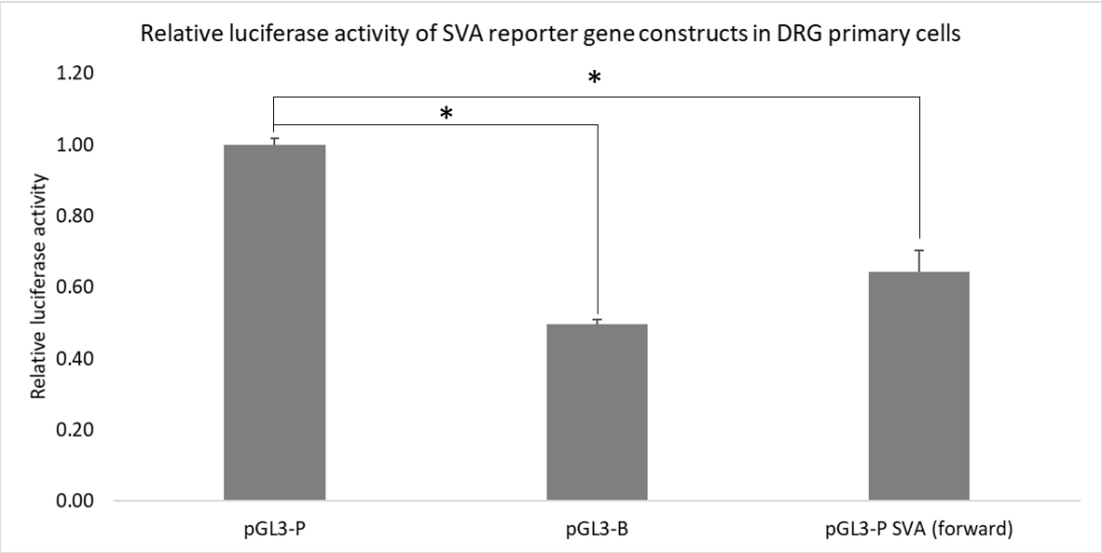


Figure 3.8. SVA at TRPV1 and TRPV3 functions as transcription regulator. Relative luciferase activity in empty (pGL3-P) vector was compared against baseline (pGL3-B) and SVA vectors (pGL3-P containing an SVA in endogenous and reverse orientations). Luciferase activity was normalised to the internal control vector (pRL-TK). Error bars represent standard error from the mean (SEM). Mann Whitney tests were used to measure the significance of fold change between vectors. Significance p-value <0.05*.

3.3.2 CRISPR assay optimisation and generation of modified cell lines

3.3.2.1 *In silico* design of CRISPR assay to delete SVA and ECR at *TRPV1* and *TRPV3*

We designed a CRISPR experiment to delete the SVA and the ECR at the intergenic region between *TRPV1* and *TRPV3* and assess the potential impact on *TRPV1* and *TRPV3* mRNA expression. The Cas9 enzyme utilised in this experiment was from *Streptococcus pyogenes* and encoded by the vector pSpCas9(BB)-2A-GFP (Figure 3.9). The Cas9 enzyme was guided to a target sequence in the genome by a sgRNA tag which contained a specific 20 nt sequence that binded to the complimentary strand of the target sequence in the genome (Figure 3.10). This sequence was directly upstream of a protospacer adjacent motif (PAM) site (5'-NGG-3', where N can be any nucleotide) as this specific Cas9 nuclease creates a double strand break (DSB) 3 bp upstream of the PAM[210] (Figure 3.10). A dual targeting approach was taken which involved co-transfecting cells with two pSpCas9(sgRNA) plasmids, to transcribe the required Cas9 machinery to generate two separate DSBs at the 5' and 3' end of the target sequence – which resulted in total excision of the target sequence and repair of each DSB via non-homologous end joining (NHEJ)[210].

Multiple sgRNAs were designed using *in silico* methods described in section 2.2.8.1. to enable excision of the ECR, SVA (separately) or both the ECR and SVA (together) based upon the combination used in different trials (Figure 3.11). Several guides were designed which differed by a few nucleotides as the predicted off target effects (capacity of the gRNA to bind elsewhere in the genome, resulting in additional DSBs) were dependent upon sequence. The option of having multiple sgRNAs also enabled us to test which combination of gRNAs created the most efficient deletion (discussed in section 3.3.2.3). All gRNA sequences are given in appendix A, entry 7.

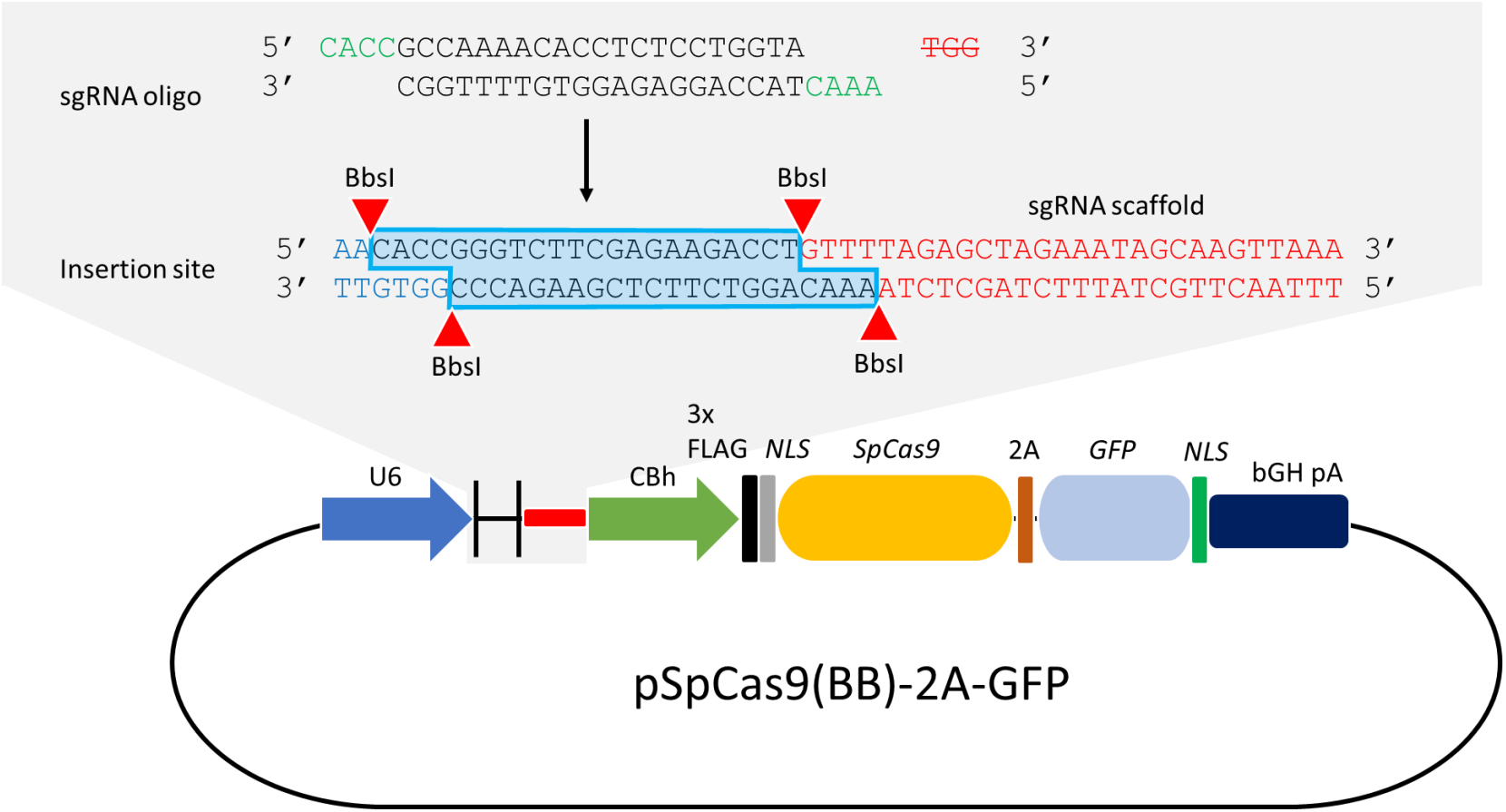


Figure 3.9. Overview of pSpCas9(BB)-2A-GFP. sgRNA oligonucleotides containing the 20 nt target sequence (black sequence) were modified via removal of the PAM sequence (red text) and addition of sequences at (green) to generate compatible 5′ overhangs which were used to clone into the BbsI site of the plasmid. Image adapted from Ran *et al.*[210]

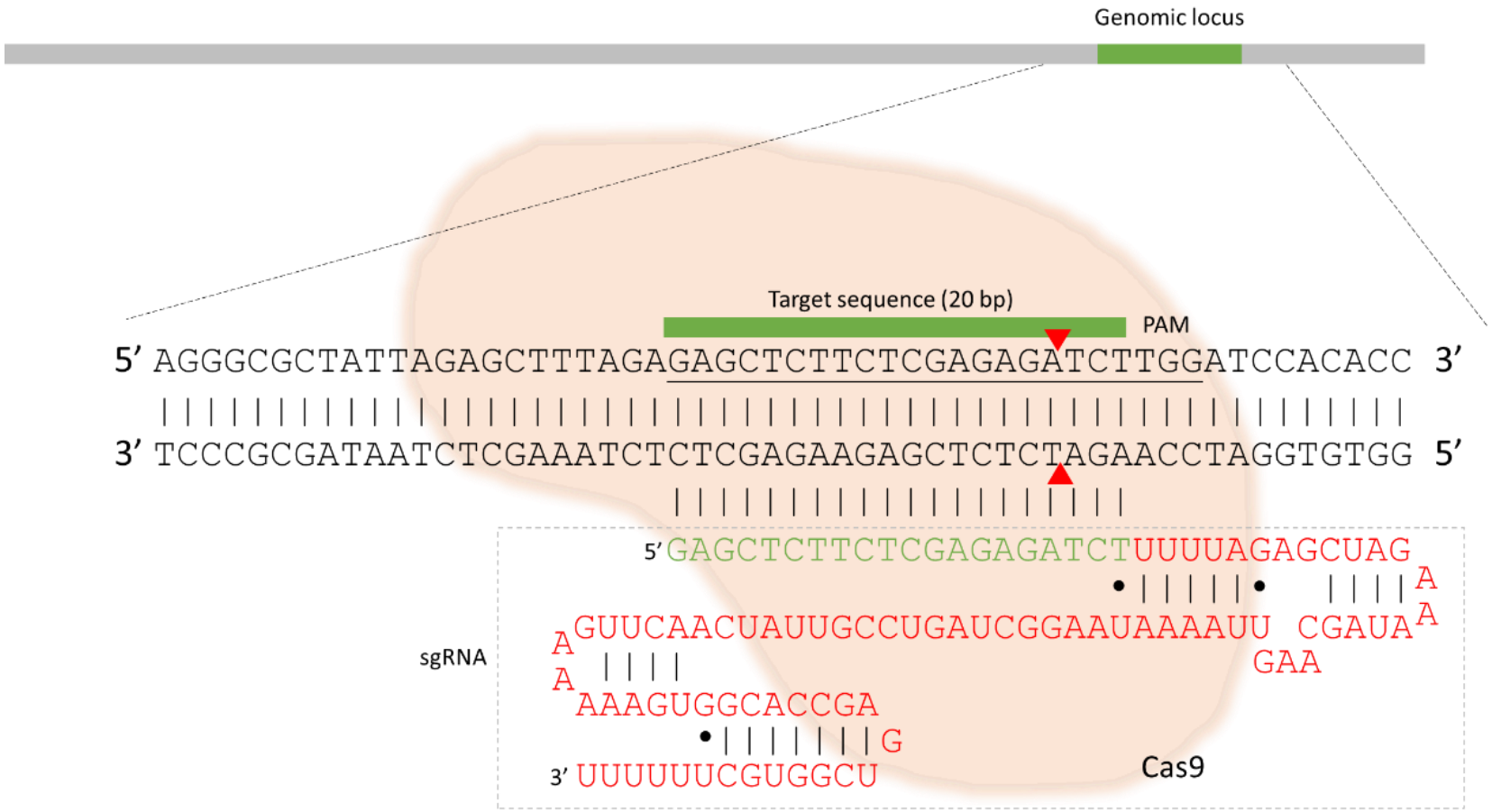


Figure 3.10. Schematic of Cas9 cleavage guided by sgRNA. A 20 nt target sequence (green) directly upstream of a PAM site is cloned into the pSpCas9 vector. When expressed, the Cas9 enzyme has a sgRNA tag (red) which encodes the target sequence (green), which pairs with the complimentary strand of the target sequence. The guided Cas9 then makes a DSB (red triangle) 3 bp upstream of the PAM sequence. Image adapted from Ran *et al.*[210]

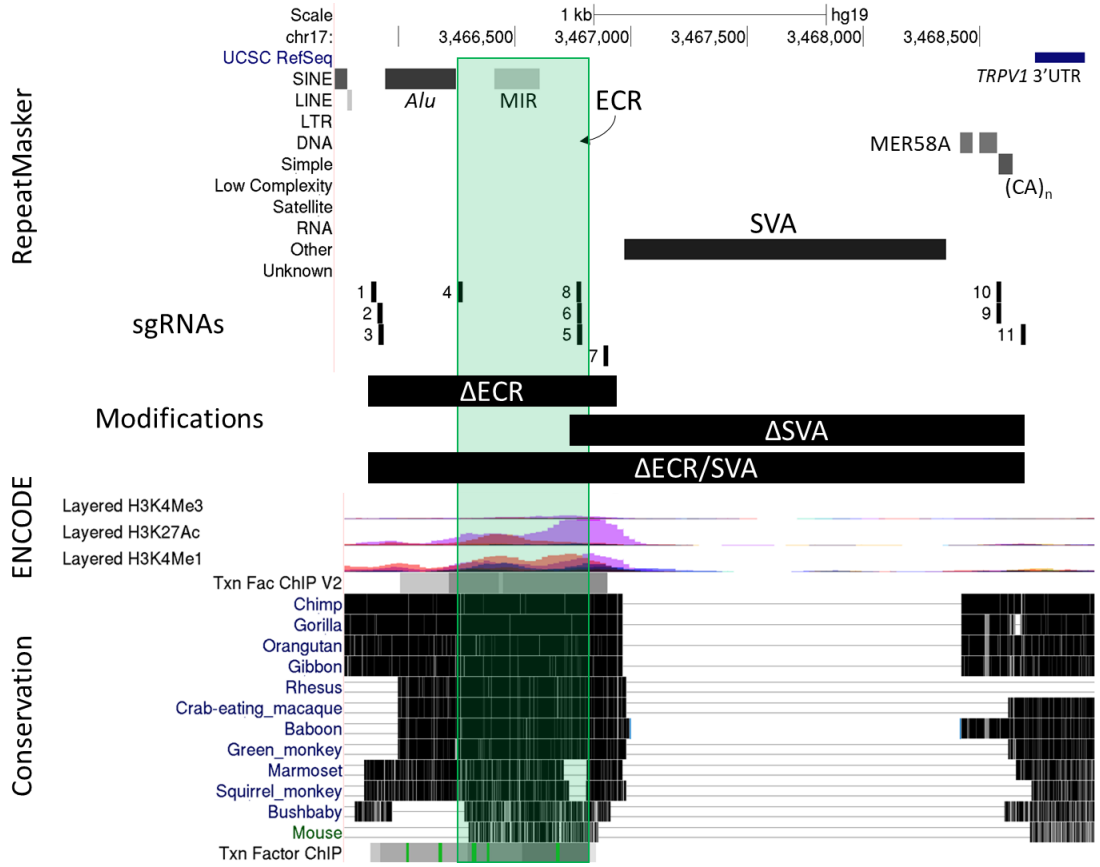


Figure 3.11. gRNA target sequences. Screenshot taken from UCSC genome browser (hg19) showing the SVA at the intergenic region between *TRPV1* and *TRPV3*. gRNA target sequences were selected (shown inside black box). SVA is denoted. ECR is highlighted in green box. gRNA 1, 2 and 3 are upstream of the ECR. 4, 5, 6 and 8 are within the ECR. gRNA 7 is between the ECR and the SVA. gRNAs 9, 10 and 11 are between the SVA and the 3'UTR of *TRPV1*. Examples of modifications to be generated in separate cell lines are annotated by black boxes.

3.3.2.2 Generation of CRISPR-Cas9 vectors containing gRNAs

To clone the gRNA sequences into the Cas9 vector pSpCas9(BB)-2A-GFP, a double stranded oligonucleotide was required with specific sequence modifications to produce compatible ends for cloning (Figure 3.9). The target sequences were modified with the removal of the PAM sequence and the addition of 5'-CACC-3' onto the 5' end (Figure 3.9). The complimentary olidonucleotide to the target sequence was modified with 5'-AAAC-3' onto the 5' end (Figure 3.9). Complimentary oligonucleotides were then annealed and cloned into the BbsI site of the Cas9 vector using golden gate cloning (Figure 3.9). Successful cloning resulted in the disruption of the BbsI cut site and no restriction digest. All cloning was successful using BbsI digest (Figure 3.12). This was further confirmed using DNA sequencing. Plasmids containing gRNAs are referred to as pSpCas9(sgRNA) from this point onwards.

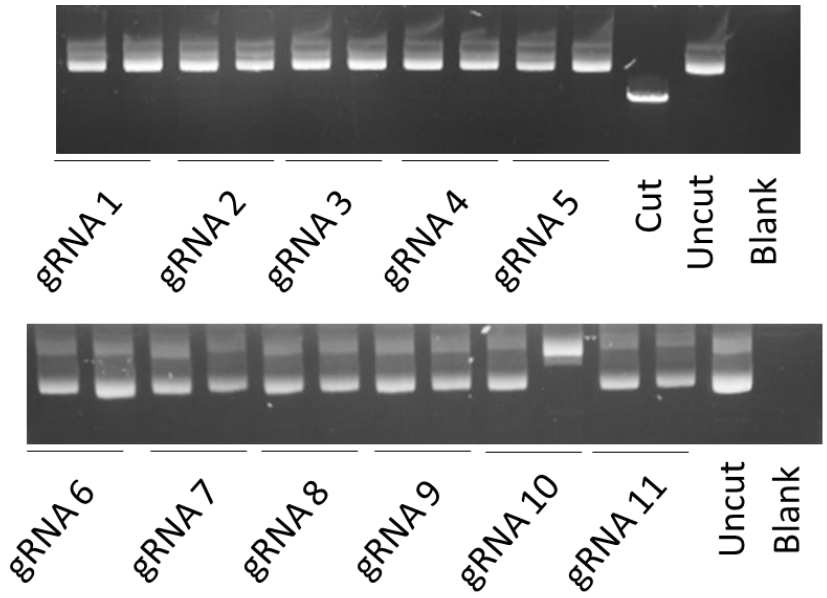


Figure 3.12. BbsI restriction digest results showing successful cloning of gRNA inserts. Failure to digest using BbsI indicates recognition site was disrupted through cloning and therefore represents a recombinant plasmid. All plasmids were verified by sequencing. 2 results are present for each guide as 2 colonies were picked during cloning. Positive control is the linearised plasmid. Negative controls include the uncut plasmid (no enzyme present in digest) and blank.

3.3.2.3 Optimisation of CRISPR-Cas9 strategy for efficient SVA deletion

The CRISPR protocol was optimised by focussing efforts on deleting the SVA at *TRPV1* and *TRPV3*. pSpCas9(sgRNA) plasmids were transfected in various combinations into different cell lines and screened using PCR to identify which gRNAs enabled the greatest modification efficiency and which cell line was most responsive to modification. PCR primers and thermal cycles are listed in appendix A, entry 9. Results from PCR would indicate if a modification had taken place. A large PCR product (2486 bp) indicates an unmodified region the SVA. A smaller PCR product (545 bp – 746 bp, dependent on sgRNAs used) indicates a modified region (not containing an SVA). All PCRs were performed on purified gDNA from pooled populations 48 hours post transfection.

3.3.2.3.1 HAP1

We first trialled HAP1 cells, as they contained a haploid genome[221]. This would have enabled modification of a single allele and ability to measure gene expression in modified cells without effects being masked by a remaining allele.

This cell line was problematic as the GFP from pSpCas9(sgRNA) vectors could not be detected after 24 or 48 hours under a fluorescent microscope, therefore potentially modified cells could not be isolated using FACS (no data to show). GFP from pmaxGFP (used as a positive control) could be detected however, indicating a low transfection efficiency <10% and an issue with the expression of pSpCas9(sgRNA) in HAP1 (Figure 3.13).

The pSpCas9(sgRNA) vector was being expressed as some modifications were observed following genotyping, although this was likely at a low level as modified PCR products were evident on agarose gels at a very low intensity compared to the unmodified PCR product, indicating a poor modification efficiency (Figure 3.18).

The poor modification efficiency may have been due to the poor transfection efficiency, resulting in low levels of pSpCas9(sgRNA) being expressed. Uncharacterised

genetic aberrations in target sequences surrounding the SVA may have also contributed to inefficient modification. The lack of GFP from pSpCas9(sgRNA) was not resolved therefore were unsuitable for cell sorting using FACS. We concluded that HAP1 cells were not suitable for this assay as modification efficiency was poor and the lack of a Cas9-GFP tag would make it increasingly difficult to isolate modified clones.

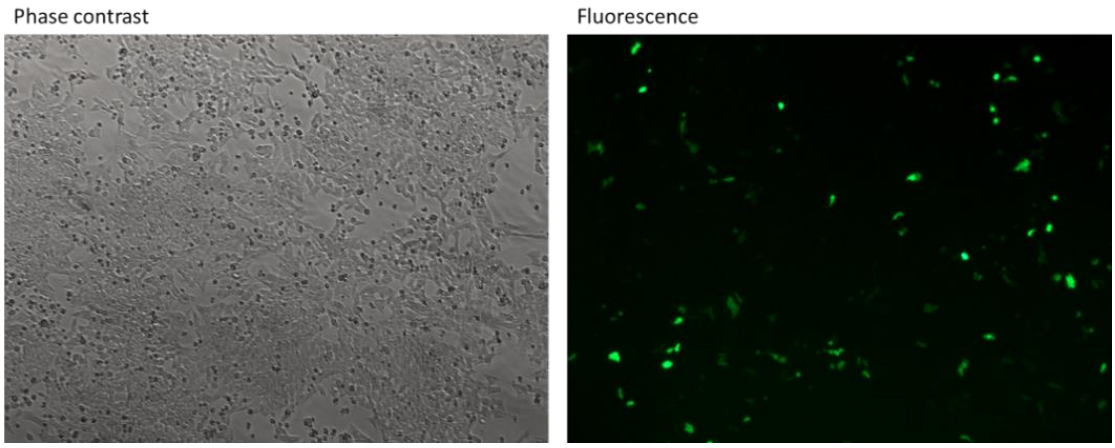


Figure 3.13. Transfection of HAP1 cells with pmaxGFP. (Left) Phase contrast of image showing cells at approximately 80-90% confluency, 48 hours post transfection. (Right) Flourescent microscopy image showing expression of GFP from pmaxGFP control vector in the same field of view. Estimated transfection efficiency <10%.

3.3.2.3.2 SH-SY5Y

SH-SY5Y cells were trialled next, as they are characterised by neuronal markers and are often used for *in vitro* assays within neurobiology[222,223].

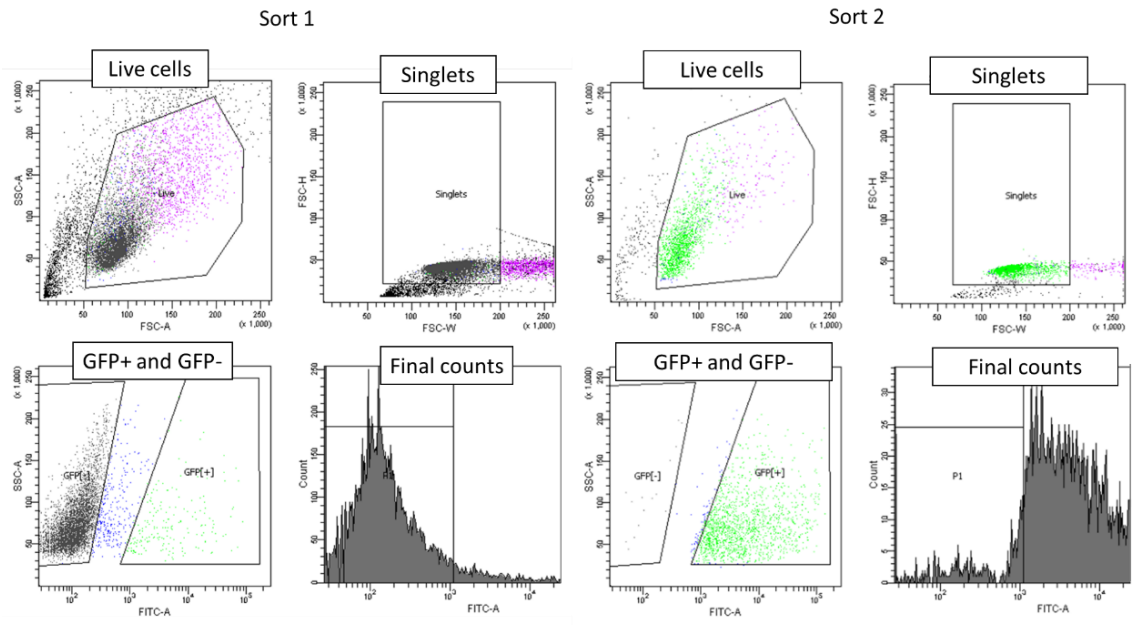
PCR analysis revealed that all guide combinations generated modifications (Figure 3.18), therefore the assay was working more efficiently in this cell line compared to results in HAP1. pSpCas9(sgRNA7) and pSpCas9(sgRNA10) produced generated an SVA deletion with the smallest amount non target DNA also being excised, therefore these constructs were taken forward. GFP was expressed from pSpCas9(sgRNA) and fluorescing well at 48 hours post transfection therefore this cell line was suitable for cell sorting via FACS to isolate GFP positive cells and enrich for cells with potential modifications (no data to show).

SH-SY5Y were then transfected with the selected constructs pSpCas9(sgRNA7) and pSpCas9(sgRNA10) and sorted based on GFP by FACS. The output from FACS with thresholds set to isolate live, single cells expressing GFP is shown (Figure 3.14). We set thresholds to detect cells that were identified as being live, single cells expressing GFP - suitable for single-cell clonal expansions. No microscopy images were taken during this trial however the data from FACS showed more GFP- cells than GFP+ cells, indicating a transfection efficiency <50% (Figure 3.14). We set thresholds to sort the top 2% of GFP expressing cells, as we hypothesised those cells with the highest GFP expression would have the highest Cas9 expression and thus more likely to carry the desired modification. A second sort was performed to remove any GFP negative cells that may have been included in the original sort (Figure 3.14).

Following FACS, GFP positive and negative (controls) cells were plated at different densities across two 96-well plates (per well; 1, 5 and 10 cells) in anticipation of potential growth difficulties of clones in isolation (e.g. cell death resulting from few growth factors) however very few potential modified clones grew using this method (Table 3.1). This did not

improve after repeating with conditioned media (filtered SH-SY5Y cell media taken from overnight growth of unmodified SH-SY5Y). Untransfected cells did not show the same growth difficulties as transfected cells, highlighting that SH-SY5Y cells were particularly vulnerable to death if FACS was used to isolate GFP positive cells following transfection.

The few GFP colonies were grown and genotyped using PCR however all results showed that they did not contain any modified DNA (Figure 3.15). The difficulties isolating enough clones for genotyping highlighted SH-SY5Y cells as unsuitable to conduct CRISPR.



Population	Events	Parent (%)	Total (%)
All	10000	n/a	100.0
Live	7183	71.8	71.8
→Singlets	5975	83.2	59.8
→ GFP+	197	3.3	2.0
→ GFP-	5495	92.0	54.9

Population	Events	Parent (%)	Total (%)
All	2008	n/a	100
Live	1767	88.0	88.0
→Singlets	1607	90.9	80.0
→ GFP+	1522	94.7	75.9
→ GFP-	15	0.9	0.7

Figure 3.14. Results from FACS cell sorting SH-SY5Y for isolation of CRISPR modified clonal cell lines. (Left) First sorting of cells based on live cell detection, single cell dissociation, GFP expression and final counts showing 200 for every 10,000 cells sorted. (Right) Second sorting of cells based on the same parameters to remove any potential GFP negative (untransfected) cells from the population.

Table 3.1. Results from low density seeding GFP of SH-SY5Y cells following FACS. Total cells plated = 196 wells.

	GFP positive		GFP negative (control)	
Seeding density (per well)	Positive wells	%	Positive wells	%
1	0	0.0	4	2.1
5	0	0.0	32	16.7
10	3	1.6	42	21.9

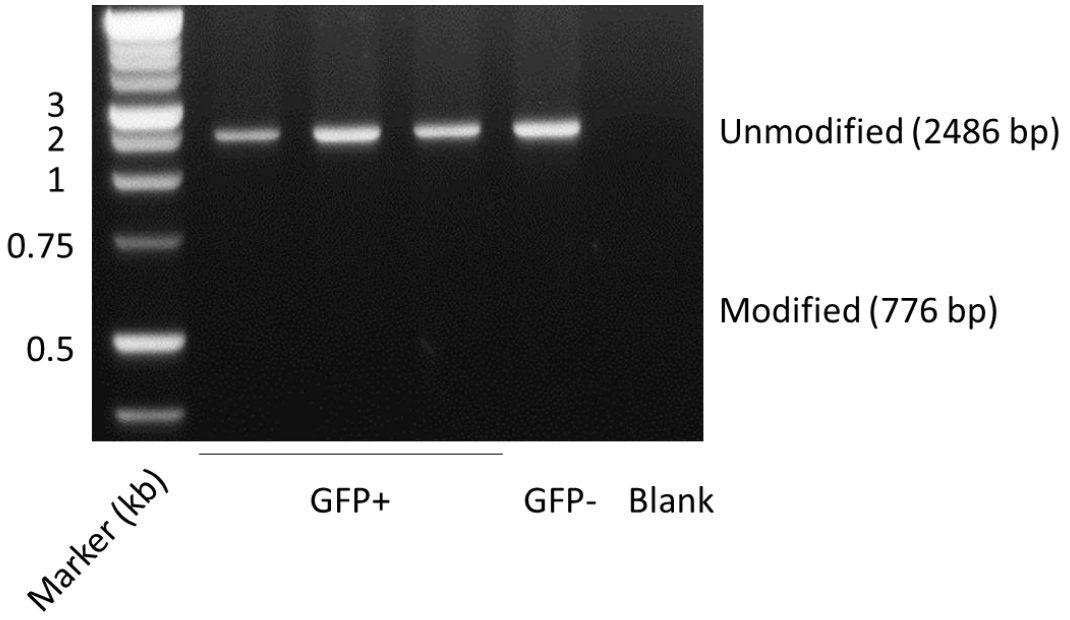


Figure 3.15. No modifications identified in SH-SY5Y following isolation of clones. Clones identified from low density seeding were genotyped to detect SVA presence/absence following CRISPR. GFP+ cells previously isolated using FACS. GFP- cells were used as a negative control. Blank refers to no template control. Unmodified PCR product is 2486 bp and modified product is 776 bp (not identified).

3.3.2.3.3 HEK293

HEK293 cells were the final cell line to be trialled for CRISPR as they are well characterised as being easy to transfect and produce high levels of exogenous proteins[224-226]. The ATCC website describes four copies of chromosome 17 (http://www.lgcstandards-atcc.org/products/all/CRL-1573.aspx?geo_country=gb#characteristics), which is supported by Binz *et al*[227]. A review of publicly available SNP array data revealed no CNVs across chromosome 17 in HEK293, compared to HEK293T (a variant of cell line of HEK293), in which a large-scale deletion was identified (Figure 3.16A)[225]. An earlier study of the HEK293 karyotype also found no aberrations on chromosome 17, supporting this data[224]. Visualisation of the SVA containing region using a HEK293 genome browser highlighted few SNPs, deletions and insertions providing evidence that this region would be readily susceptible to modification using CRISPR (Figure 3.16B). As previously described, HEK293 populations were transfected with each combination of pSpCas9(sgRNA) vectors. PCR analysis (Figure3.18) showed the products corresponding to modified regions (range 545 bp – 746 bp) were much brighter on agarose gels than the unmodified regions (2486 bp). This suggested the modification efficiency in HEK293 was improved compared to HAP1 and SH-SY5Y and was the chosen cell line to generate clonal cell lines. SpCas9(sgRNA7) and SpCas9(sgRNA10) target DNA sequences were closest to the SVA (Figure 3.11) therefore were used.

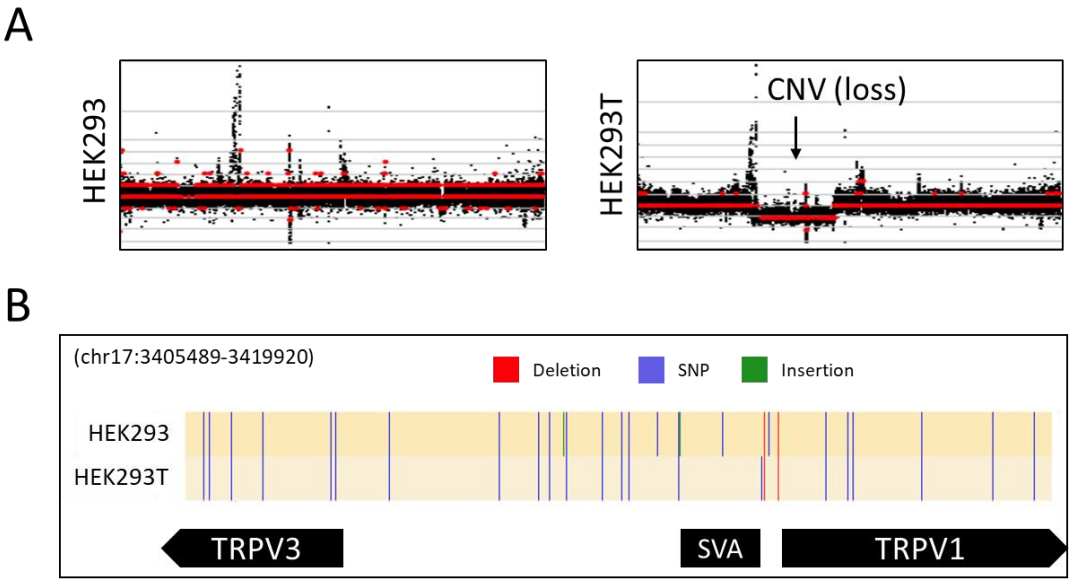


Figure 3.16. Bioinformatic assessment of intergenic region at TRPV1 and TRPV3 in HEK293. (A) SNP array data showing no copy number variation (CNV) in HEK293 and an example of CNV loss in HEK293T. Data taken from Lin *et al*[225]. (B) Detailed analysis of target region containing SVA shows few base pair changes to the target sequence in HEK293.

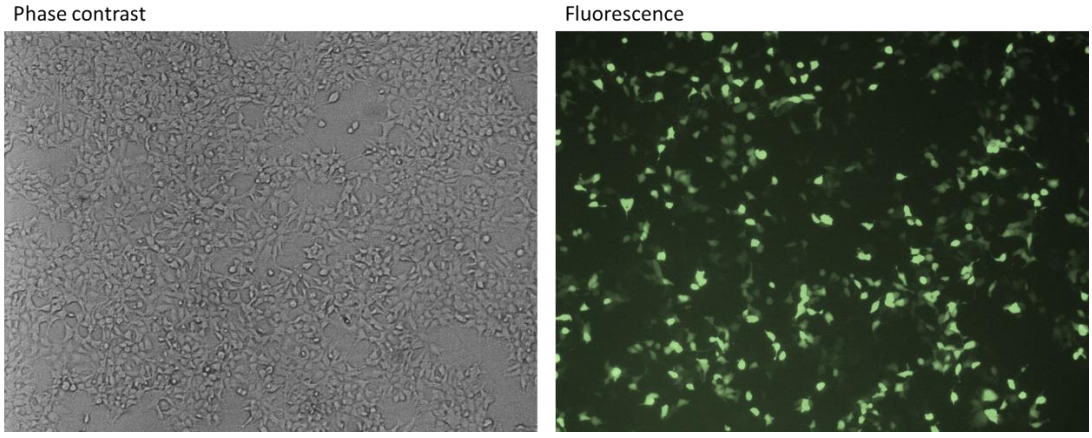


Figure 3.17. Transfection of HEK293 with pSpCas9 GFP. (Left) Phase contrast image showing HEK293 cells at approximately 80-90% confluency, 48 hours post transfection. (Right) Fluorescent microscopy image showing expression of GFP from pspCas9 control vector in the same field of view. Estimated transfection efficiency 50%.

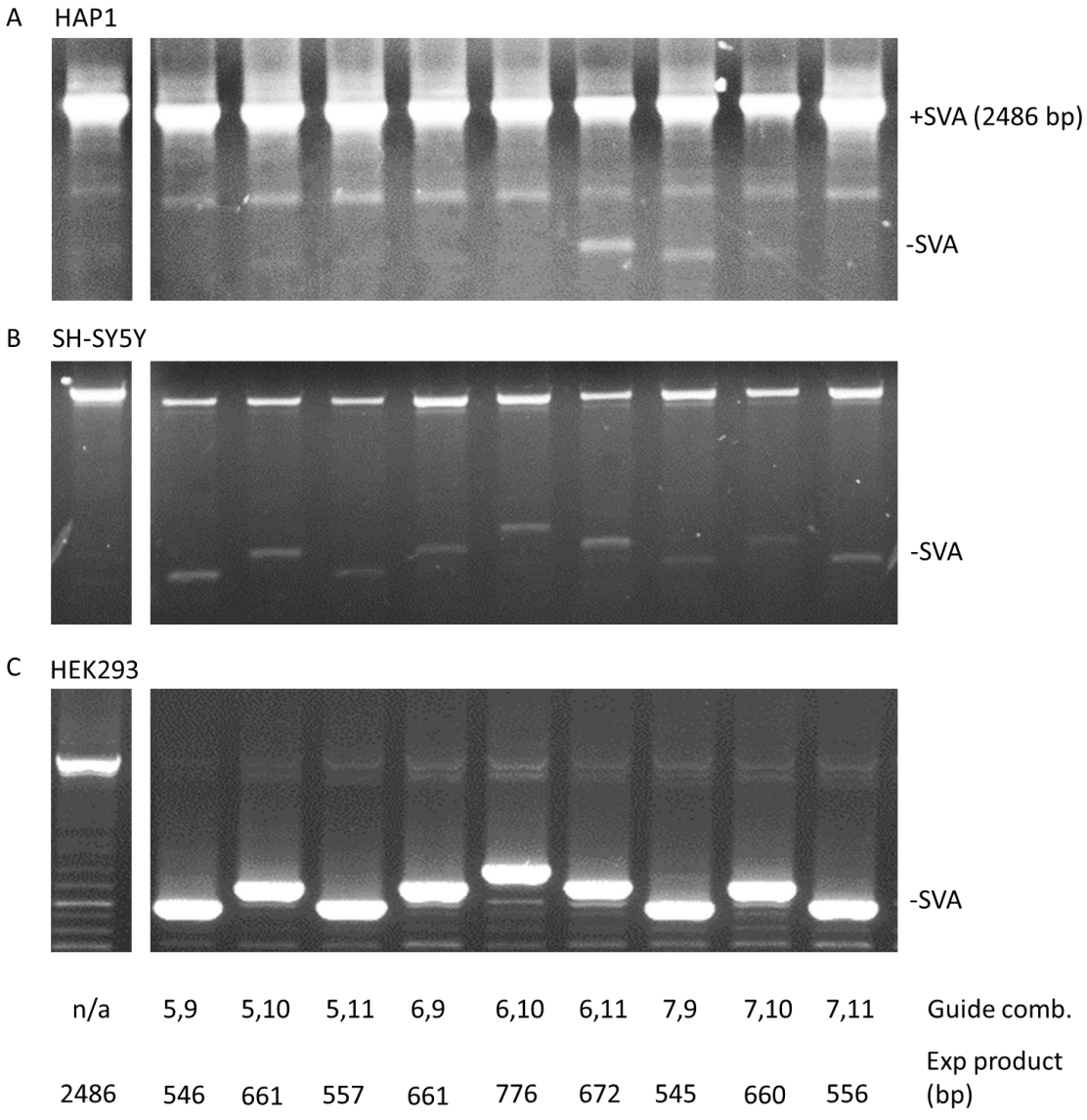


Figure 3.18. Optimising pSpCas9(sgRNA) transfection combinations in human cell lines. Different combinations of pSpCas9(sgRNA) vectors were transfected into human cell lines. Modified genomic region containing deletion of SVA generated product size ranging between 545 and 776 bp. dependent upon combination of sgRNAs. Guide combinations are denoted at the bottom, with expected product sizes. Refer to Figure 3.11 for schematic of sgRNA locations. (A) HAP1 results showing minimal modified product from combination 6,11 and 7,9. (B) SH-SY5Y results showing modification in all combinations, with majority of product corresponding to unmodified product. (C) HEK293 results showing majority of product corresponds to modified region containing SVA deletion.

3.3.2.4 Isolation of HEK293 CRISPR modified clones with SVA deletions

Previous results with SH-SY5Y highlighted difficulties in selecting clones from seeding cells in a 96-well format following FACS (section 0) therefore low-density seeding in 10 cm dishes was tested instead, as this enabled the release and uptake of growth factors between cells, increasing likelihood of survival. FACS was not used to cell sort HEK293 to avoid unnecessary strain on the cells, as the transfection efficiency (Figure 3.17) and modification efficiency (Figure 3.18)) were deemed sufficient to obtain enough modified cells.

A total of 6 plates were seeded with 1000 cells each. Low density seeding allowed the growth of single cells into clonal populations, which were then picked (when visible by eye) and expanded in a 96-well format. Spreading the cells across 6 plates ensured that no single plate was out of the incubator for any length of time, which may have reduced cell viability. A total of 360 clones were picked across the six plates. 24 hours after picking, 215 clones survived (71%) which were then grown in duplicate to provide enough cells for genotyping whilst enabling continuation of growth for further experimentation.

Screening for modifications was performed using crude cell lysate as a PCR template to increase throughput, however this method failed to amplify larger PCR products (perhaps due to PCR inhibitors in crude lysate), therefore could only be used to identify if any SVA deletion had occurred but not identify if this was a partial or complete deletion (Figure 3.19A).

Using this screening method, 33 cell lines (12%) were identified that contained at least one SVA deletion. To determine the genotype of each cell line accurately, genomic DNA was purified from selected clones and the PCR was repeated (Figure 3.19B). Three results could be observed, a single product at 545 bp indicating a homozygous deletion (i.e. all copies of the SVA are deleted), a single product at 2486 bp indicating no SVA copies were deleted or both products indicating a heterozygous deletion (i.e. at least one copy of the SVA is

deleted and at least one copy of the SVA is present). A total of 23 cell lines were confirmed to contain a heterozygous SVA deletion and 3 cell lines that contained a homozygous SVA deletion. 6 false positives were identified in the crude lysis resulting from non-specific PCR products. This represents a 10% modification efficiency for heterozygous SVA deletions and 1% modification efficiency for homozygous SVA deletions.

Sequencing was then used to profile modification breakpoints and assess the accuracy of this CRISPR system (Figure 3.21). Sequencing primers are in appendix A, entry 8. For comparison, we selected all 3 cell lines containing a homozygous SVA deletion and 3 lines containing a heterozygous SVA deletion (Figure 3.20). Based on the PAM sequence chosen when designing the gRNAs, the Cas9 enzyme should cut the DNA exactly 3 bp upstream (referred to as DSB site) (Figure 3.21).

One heterozygous cell line and all homozygous cell lines contained breakpoints that aligned perfectly with the predicted DSB sites at both the 5' and 3' end of the SVA (Figure 3.21). The remaining two heterozygous cell lines contained different breakpoints. At the 5' end the breakpoints were between 2 – 4 bp upstream of the DSB site, representing small deletions (Figure 3.21). At the 3' end, both breakpoints were 18 bp upstream of the DSB site, representing a larger deletion although still relatively small (Figure 3.21). 4 out of the 6 cell lines had breakpoints confirmed at the predicted DSB sites, indicating a high level of editing accuracy and non-homologous end joining (NHEJ) DNA repair using this system. An overview of the entire CRISPR protocol optimised for the deletion of larger DNA sequences in HEK293 is given in Figure 3.22.

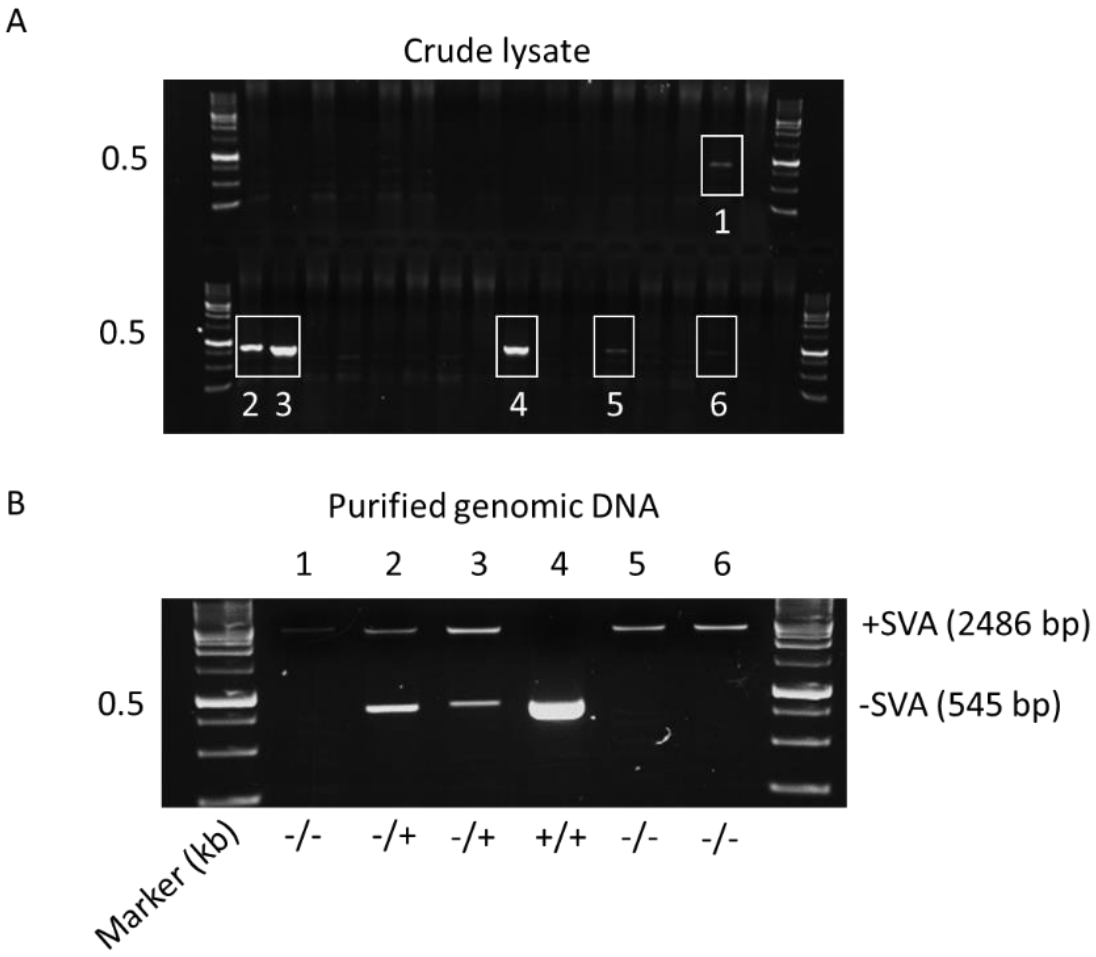


Figure 3.19. Screening HEK293 clonal cell lines to identify SVA deletion. (A) PCR results from crude lysates identify products corresponding to modified region containing SVA deletion at 545 bp. No unmodified products at 2486 bp could be detected. (B) Same samples repeated PCR using purified genomic DNA as template confirmed genotype of clonal cell lines. Sample 1, 5 and 6 were false positives resulting from non-specific PCR products in crude lysate. Samples 2 and 3 were heterozygous containing at least one deleted SVA copy. Sample 4 was a homozygous deleted sample as all SVA copies were deleted.

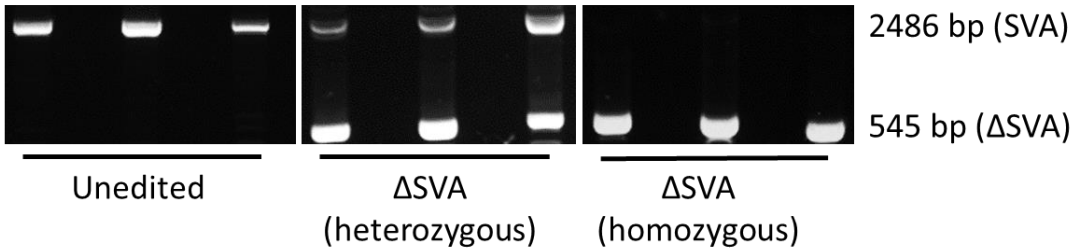


Figure 3.20. Clonal cell lines selected for gene expression analysis. PCR results showing clonal cell lines corresponding to unedited HEK293 (n=3), cell lines carrying a heterozygous SVA deletion (n=3) and cell lines carrying homozygous SVA deletion (n=3). These cell lines were used for RT-PCR/qPCR.

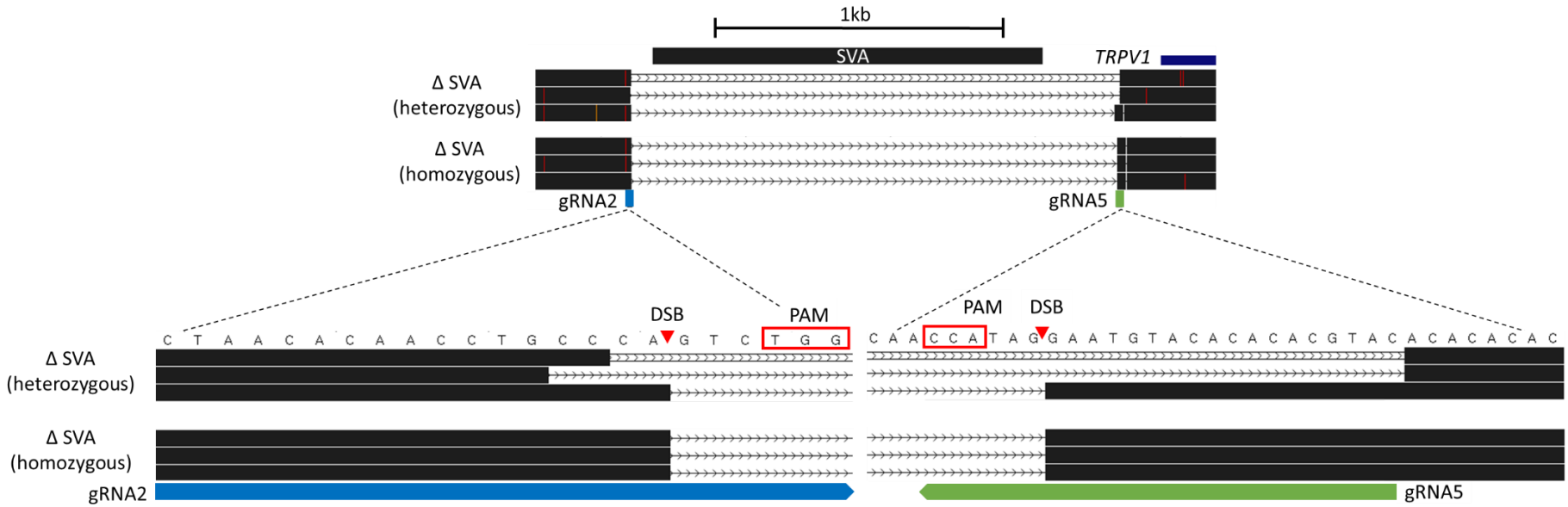


Figure 3.21. Validation of clonal CRISPR modified cell lines. (A) PCR validation of clonal modified cell lines containing heterozygous and homozygous SVA deletions. B) Sequencing validation of DSBs via NHEJ in clonal cell lines.

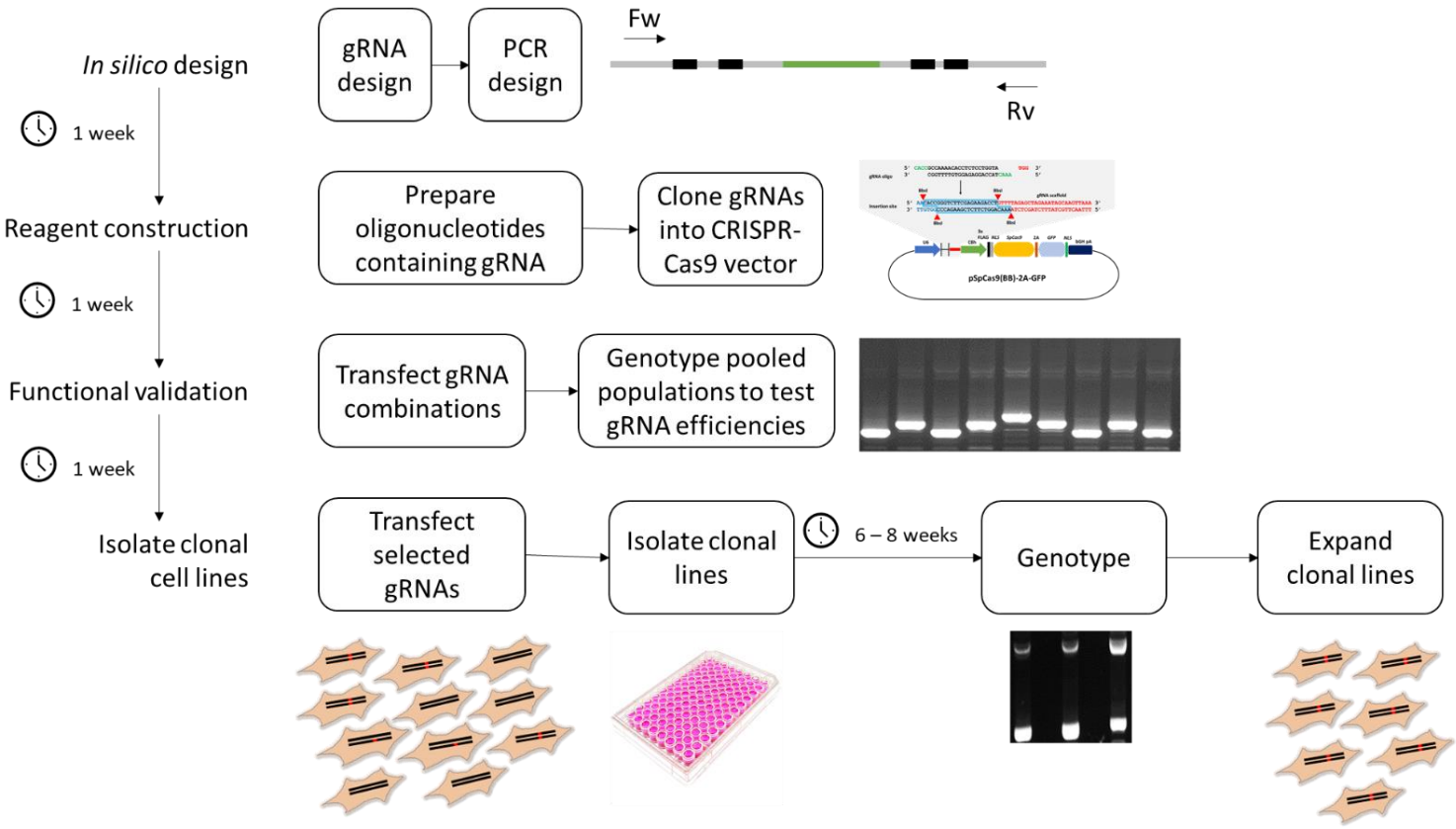


Figure 3.22. Overview of final CRISPR protocol developed to generate cell lines containing large deletions. *In silico* design includes design of gRNAs and PCR primers to test for deletion. Reagent construction includes preparation of oligonucleotides containing gRNAs and cloning into CRISPR-Cas9 vector pSpCas9(BB)-2A-GFP. Functional validation includes transfection of CRISPR-Cas9 vectors containing gRNAs into cell lines, isolating genomic DNA and PCR to test for deletion – which informs which gRNAs are most efficient at generating required modification. Isolation of clonal cell lines includes transfection of selected gRNAs, isolation and growth of clonal cell lines (this can take up to 8 weeks), followed by genotyping to identify positive clones and expansion for functional analysis.

3.3.2.5 Isolation of HEK293 cell lines containing ECR deletions was unsuccessful

We also attempted to delete the ECR independently and generate a double knockout cell line carrying a combined ECR/SVA deletion (Figure 3.11). Following the same protocol as described in the previous section (section 3.3.2.4). HEK293 was transfected using various combinations of pSpCas9(sgRNAs).

To delete the ECR, pSpCas9(sgRNA4) was selected to create the 5' DSB as this was the only guide that would enable deletion of the ECR conserved with the mouse, and not include a deletion of the adjacent *Alu* which was conserved with primates only, not mouse (Figure 3.11). To delete the ECR and SVA together, pSpCas9(sgRNA4) was paired with pSpCas9(sgRNA10) as this had already been used to generate the SVA clonal deletions, therefore would generate a more comparable deletion than using the other constructs.

PCR products corresponding to modified regions were detected in both ECR deletion trial 1 (2733 bp) and ECR deletion trial 2 (2618 bp), indicating the constructs were working (Figure 3.23A). However, the intensity was very low compared to the product corresponding to the unmodified region (3264 bp), indicating the modification efficiency was also low.

Constructs used in ECR trial 2 were taken forward as this appeared to work better. This was also the case for the double ECR/SVA knockout trial which showed a low intensity product corresponding to the region containing neither SVA or ECR (947 bp) (Figure 3.23B). This was repeated to however no ECR deletion could be detected. The double ECR/SVA deletion was replicated therefore was taken forward to isolate clonal cell lines.

After low density seeding, and growth in 96-well format, 44% (n=42) of colonies grew to sufficient numbers for genotyping, the remainder did not survive. From these colonies, crude lysates were made for screening however zero contained any deletion (Figure 3.23C).

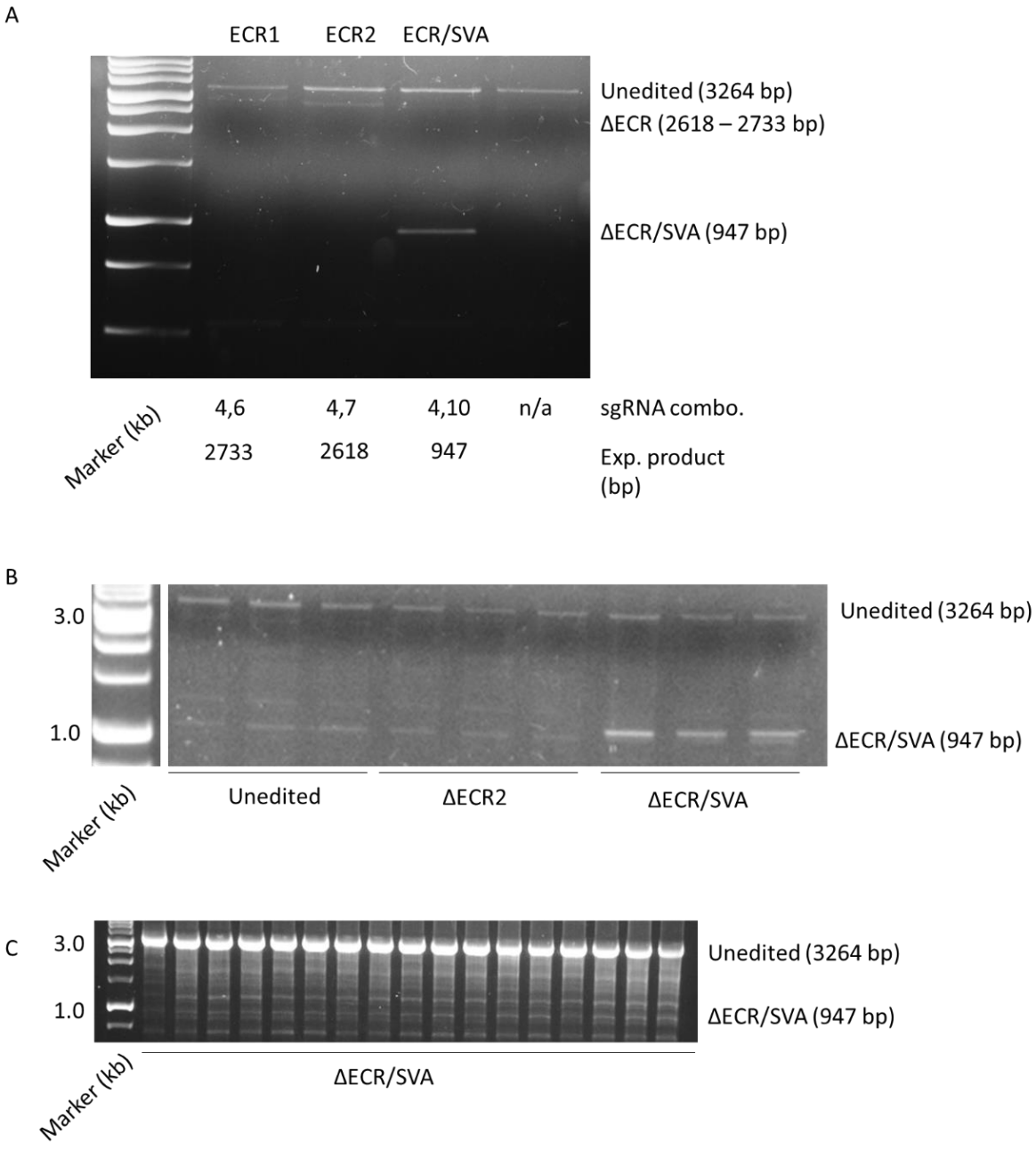


Figure 3.23. CRISPR modification of ECR deletion in HEK293 was unsuccessful. (A) Testing pSpCas9(sgRNA) combinations in ECR deletion trials (ECR1 and ECR2), and double knockout ECR/SVA shows modifications achieved in HEK293. (B) Repeat knockout of ECR and ECR/SVA with selected pSpCas9(sgRNA) show ECR modification was unsuccessful but ECR/SVA medication was achieved. (C) Genotyping of clones isolated from mixed population containing ECR/SVA modification show no modified clones were isolated.

3.3.3 Measuring *TRPV1* and *TRPV3* expression in HEK293 lines carrying SVA deletions

3.3.3.1 Optimisation of qPCR assay to measure total TRPV1 and TRPV3 mRNA and RNA quality controls

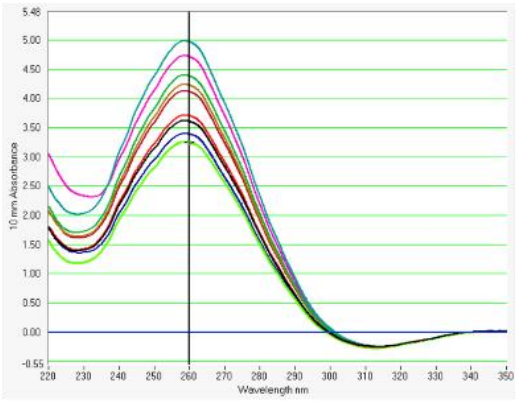
To measure changes in TRPV1 and TRPV3 expression, we designed a qPCR assay to measure total mRNA. Primer sequences and thermal cycling conditions are given in appendix A (entry 10). All data was normalised against β-actin (ACTB), a highly conserved cytoskeletal protein that is commonly used as a reference gene in RT-PCR[228].

Primer efficiencies were calculated using standard curves (Figure 3.25). ACTB primers were 100.8% efficient, TRPV1 primers were 109.57% efficient and TRPV3 primers were 106.28% efficient. It is desirable to achieve a primer efficiency between 90 – 110%. Primer efficiencies of 100% indicates the polymerase is working at maximum capacity. Efficiencies >100% can indicate inhibitors in the reaction including excessive amounts of RNA template or contaminants such as ethanol or phenol, from RNA purification steps. The results were within the desired range therefore we proceeded with the experiment using these primer sequences.

RNA quality control data did not indicate the presence of inhibitors, as all 260/280 and 260/230 ratios were >2.0 and absorbance plots produced expected curves, therefore efficiencies >100% may have been due to pipetting error (Figure 3.24B). All RNA samples were intact and not degraded (Figure 3.24C).

All dissociation curves indicated the presence of a single product (Figure 3.25). *TRPV1* and *TRPV3* controls (including no reverse transcriptase and no template) did not amplify any product therefore we were confident that cDNA preparation and reaction setup had been conducted properly (Figure 3.25). We observed low level amplification in ACTB controls towards the end of the reaction (>35 cycles) but this was attributed to primer dimer (Figure 3.25).

A



B

Sample ID	Concentration (ng/ul)	260/280	260/230
P3.G3	180.2	2.14	2.57
P2.H3	185.1	2.17	2.6
P3.G7	219	2.15	2.54
P3.D12	211.1	2.15	2.56
P3.C5	162.3	2.16	2.72
P3.G2	248.3	2.12	2.44
P3.G7	169.3	2.17	2.49
P3.G8	235.6	2.14	2.02
P3.G11	205.7	2.14	2.54

C

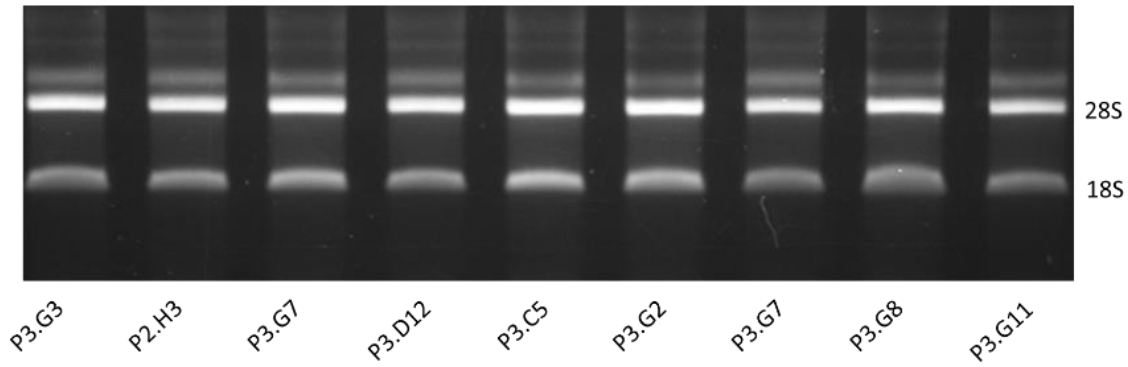


Figure 3.24. Example of RNA quality controls. (A) Absorbance plots from Nanodrop showing peak absorbance at expected 260 nm. (B) Summary of 260/280, 260/230 ratios and RNA concentrations following purification – highlighting pure RNA samples. (C) RNA samples run on agarose gel indicating intact RNA form presence of 28S and 18S bands and lack of smear.

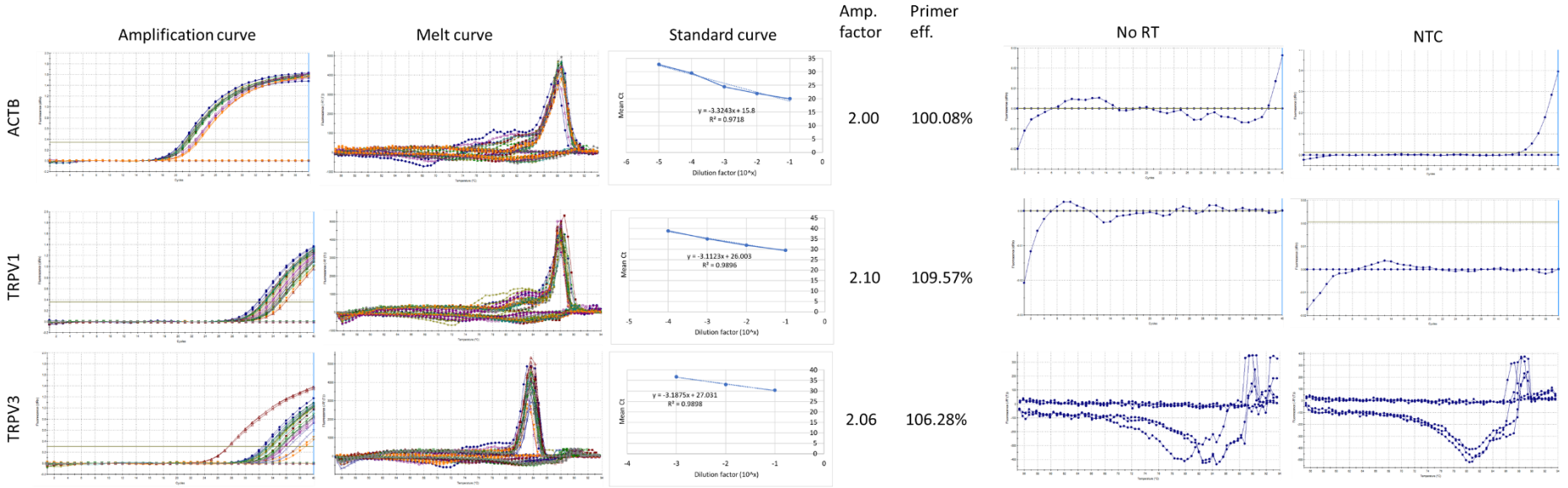


Figure 3.25. qPCR optimisation results and primer efficiencies. Amplification curves showing real time amplification of products from TRPV1, TRPV3 and ACTB primer sets. Dissociation curves show single peaks indicating amplification of single product. Standard curves were used to calculate slope of curve which was then used to calculate amplification factor and primer efficiency. No reverse transcriptase (No RT) and no template controls (NTC) show lack of amplification in TRPV1 and TRPV3 trials, indicating no contamination gDNA is present in preparation. The presence of low-level amplification in ACTB No RT and NTC controls indicates primer dimer formation.

3.3.3.2 Total *TRPV3* mRNA expression increases in cell lines carrying SVA deletion

qPCR was used to measure total *TRPV3* mRNA (normalised against ACTB) in HEK293 cells containing an SVA deletion compared against unedited HEK293. Cells transfected with Cas9 vectors containing non-targeting RNAs (pSpCas9(ntgRNA)) were used as a control, and all data was normalised against this group.

Untransfected cells showed no difference in *TRPV3* expression compared to pSpCas9(ntgRNA) controls (0.97±0.05, n=3, P>0.05) (Figure 3.26). This indicated that transfection with the Cas9 enzyme alone (lacking a specific sgRNA) did not induce any observable change in total *TRPV3* mRNA expression and supported that differences observed in edited cells were a result of the genetic modification.

Cells with a heterozygous SVA deletion had a mean 4.6-fold increase in *TRPV3* mRNA expression (4.56±2.9, range: 1.28-10.2, n=3, P>0.05). *TRPV3* expression increased in all three cultures carrying a heterozygous SVA deletion.

Cell lines carrying a homozygous modification had a mean 1.5-fold decrease in *TRPV3* expression (0.67±0.39, range: 0.01-1.35, n=3, P>0.05). Whilst statistical significance was not reached, changes in *TRPV3* mRNA expression were only observed in cells carrying an SVA deletion, consistent with the results from the reporter gene assays showing the SVA functioned as a regulatory element in HEK293 (Figure 3.6). It was concluded that the SVA insertion functions as a transcriptional regulator of *TRPV3*.

3.3.3.3 Total *TRPV1* mRNA expression does not change in cells with SVA deletion

qPCR was also used to measure total *TRPV1* mRNA expression in modified cells compared against unedited cells transfected with pSpCas9(ntgRNA) as controls (Figure 3.27). Primers and thermal cycles are listed in appendix A, entry 10.

Untransfected cells showed a mean 1.2-fold decrease in *TRPV1* mRNA (0.80±0.03, range: 0.75-0.86) compared to cells transfected with pSpCas9(ntgRNA) (1.01±0.09, range: 0.88-1.19, P>0.05). This indicated that transfection of cells with Cas9 was inducing some observable effect on *TRPV1* mRNA expression. Cells with a heterozygous SVA deletion had a 1.18-fold increase in *TRPV1* mRNA expression (1.18±0.51, range: 0.58-2.19, n=3, P>0.05).

Cells with a homozygous deletion had a 1.12-fold increase in *TRPV1* mRNA expression (1.12±0.12, range: 0.95-1.34 n=3, P>0.05). The results between pSpCas9(ntgRNA) controls and modified cell lines were variable and within the same range, therefore changes to expression in cell lines containing an SVA deletion could not be specifically attributed to the genetic modification.

TRPV1 mRNA expression differences in cell lines carrying a heterozygous modification were the most variable across the experiment, like the trend seen for *TRPV3* mRNA (Figure 3.26). This may have been a genuine reflection of SVA regulatory activity, but the possibility of off target effects could not be dismissed due to changes observed between untransfected and pSpCas9(ntgRNA) controls. For example, introduction of a double strand break in an exon of a protein coding gene could introduce a nonsense mutation and generate a gene knockout. If this gene is involved in the same pathway as *TRPV1* (or a regulator of *TRPV1*), this could induce change in *TRPV1* mRNA. This could hypothetically explain why we see no effect of pSpCas9(ntgRNA) controls on *TRPV3* expression but observed them in *TRPV1* mRNA expression. Ultimately, it was concluded that the SVA does not contribute to the changes in total *TRPV1* mRNA expression in this model.

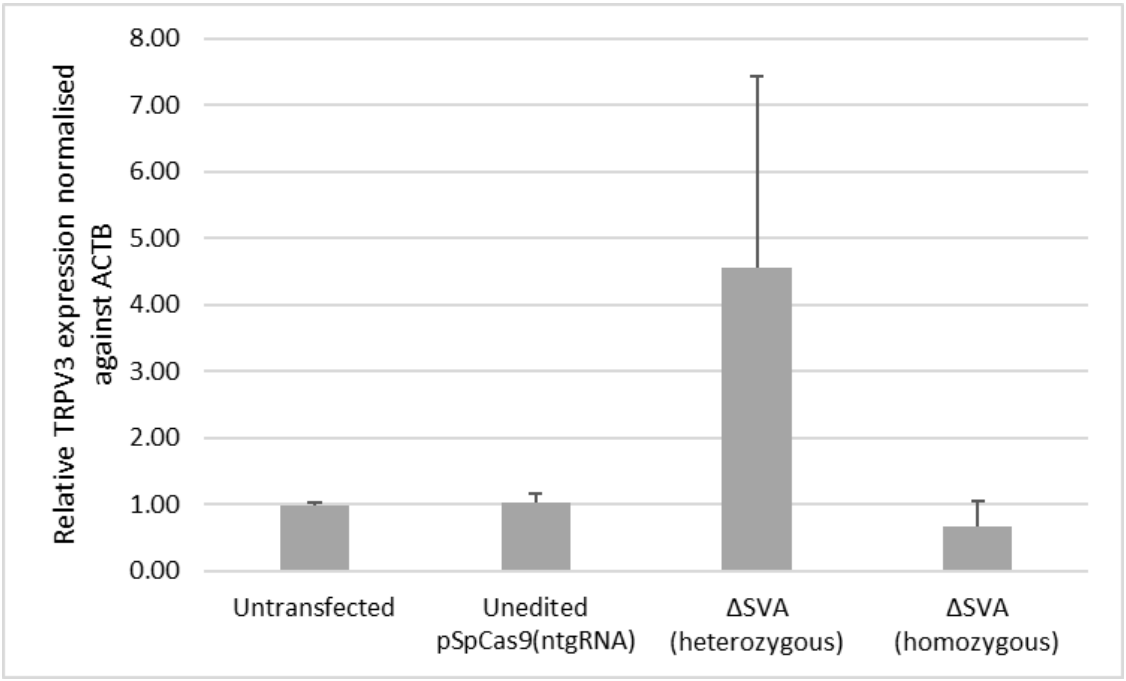


Figure 3.26. TRPV3 mRNA expression increases in CRISPR modified HEK293 cell lines carrying homozygous SVA deletion. All data normalised against ACTB. All results compared against unedited cells transfected with pSpCas9(ntgRNA). No statistical significance detected between edited and unedited cell lines (Mann-Whitney U test). Error bars indicate SEM.

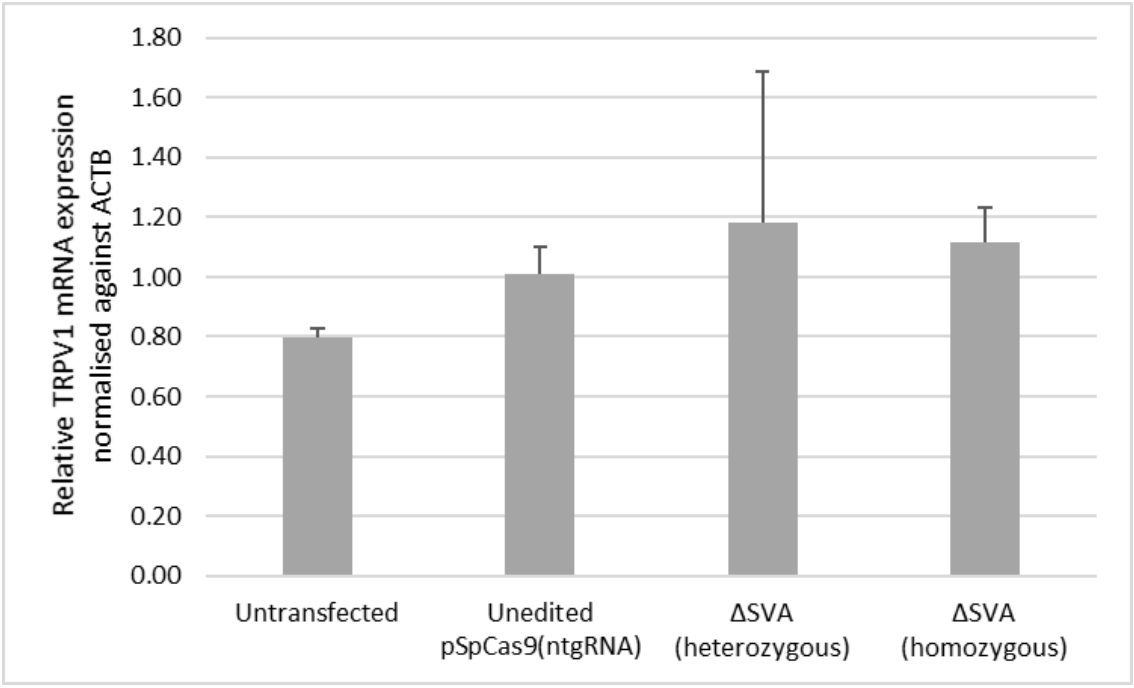


Figure 3.27. TRPV1 mRNA expression did not alter in CRISPR modified HEK293 cell lines with SVA deletion. All data normalised against ACTB. All results compared against unedited cells transfected with pSpCas9(ntgRNA). No statistical significance detected between unedited and edited cell lines (Mann-Whitney test). Error bars indicate SEM.

3.3.3.1 Development of a multiplex RT-PCR assay to detect isoform expression of *TRPV1* and *TRPV3*

Ensemble (hg38) was used to explore protein coding transcripts for *TRPV1* and *TRPV3*. Transcripts predicted to trigger nonsense mediated decay and noncoding transcripts were not analysed in this experiment. There were 7 protein coding transcripts for *TRPV1,* however four of these transcripts encoded an 839aa protein. The other three encoded transcripts encode proteins sized 779aa, 829aa and 850aa. There were 4 protein coding transcripts for *TRPV3*. Each encoded a different sized protein; 765aa, 775aa, 790aa and 791aa. We designed primers to amplify a specific exon belonging to each protein coding transcript, to enable differential detection of isoforms using multiplex RT-PCR, in which each isoform generated a different sized product (Figure 3.28). The multiplex RT-PCR reactions developed for this assay followed the basic PCR protocol described in methods section 2.2.2.1, however nuclease free water was adjusted as multiple primers were included in each mastermix. The primers and thermal cycles are listed in appendix A, entry 11.

Across all unedited, ntgRNA, and edited cell lines, we identified a single 240 bp product corresponding to TRPV1 transcript ENST0000572705.1 (Figure 3.29). According to GTEx, this transcript is expressed across all tissues with the highest levels in tibial nerve (read count 13.2), cerebellum (read count 12.4) and cerebral hemisphere (read count 11.8). We realised it was not possible to specifically detect TRPV1 transcript ENST00000576351.5 (197 bp) in the presence of ENST0000572705.1 (240 bp) due to sharing the same exon content across the amplified region (Figure 3.28). We did not detect any band at 197 bp so we could ignore this error in this case, however it is important to note this for future studies. This data gives no evidence to suggest that the SVA influences protein coding transcript expression in HEK293.

Three *TRPV3* PCR products were identified across all unedited and edited cell lines. These were ENST00000616411.4 (927 bp), ENST00000301365.8 (243 bp) and

ENST00000576742.5 (159 bp) (Figure 3.29). All three isoforms were expressed in all unedited and edited cell lines, indicating no change in the expression of specific transcript variants in cells containing an SVA deletion. We realised that the primers designed to detect ENST00000301365.8 (243 bp) and ENST00000576742.5 (159 bp) were not specific to those transcripts in the presence of ENST00000616411.4 (927 bp), therefore we could not determine if these were specific isoforms or simply products amplified from ENST00000616411.4 (927 bp) (Figure 3.28). It was not possible to design specific primers for the smaller transcripts as the exon content other than at the 3'UTR is shared; therefore this should be taken into account in all future studies. This may explain why the intensity of different transcripts were relatively similar within each sample. From this data, we could only definitively confirm the presence of the larger transcript ENST00000616411.4 (927 bp). According to GTEx, this transcript is the second highest expressed transcript in human skin (read count 2.44). ENST00000381913.8 is the highest expressed *TRPV3* transcript in human, across a broad range of tissues including testis, nerve and brain – however this transcript was not detected in HEK293. This data gives no evidence to suggest that deletion of the SVA effects the expression of protein coding *TRPV3* mRNA transcripts in HEK293.

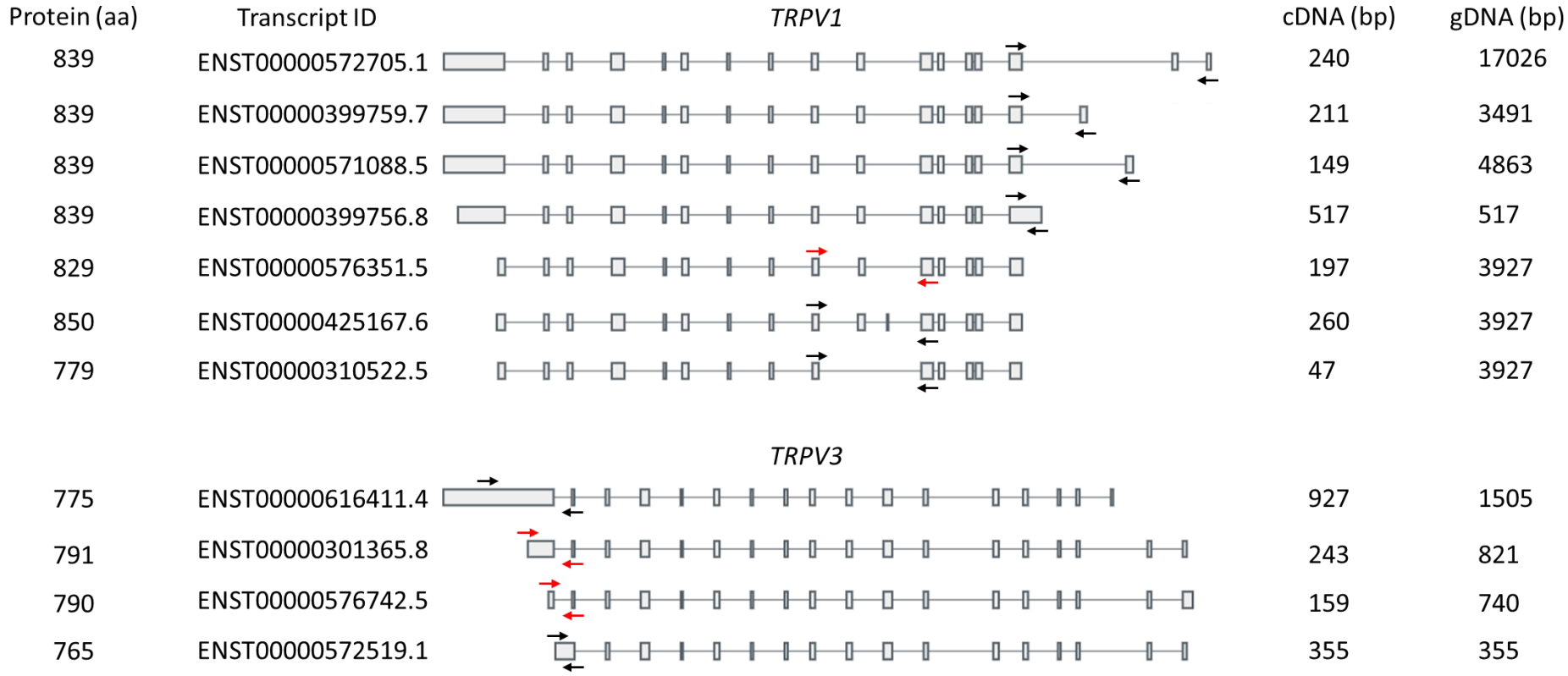


Figure 3.28. Schematic of RT-PCR design to measure TRPV1 and TRPV3 isoforms. Primers were designed within specific exons of each protein coding transcript for TRPV1 (above) and TRPV3 (below). cDNA and gDNA product sizes are given. Protein size and transcript IDs are also provided. Grey boxes represent exons in each transcript. Arrows indicate exons in which primers were designed. Red arrows indicate lack of transcript specificity in the presence of other transcripts. In TRPV1, ENST00000576351.5 cannot be specifically identified in the presence of transcripts above it due to sharing the same exon content in the amplified region. Also, in TRPV3, ENST00000301365.8 and ENST00000576742.5 cannot be specifically identified in the presence of transcripts above it due to sharing the same exon content in the amplified region.

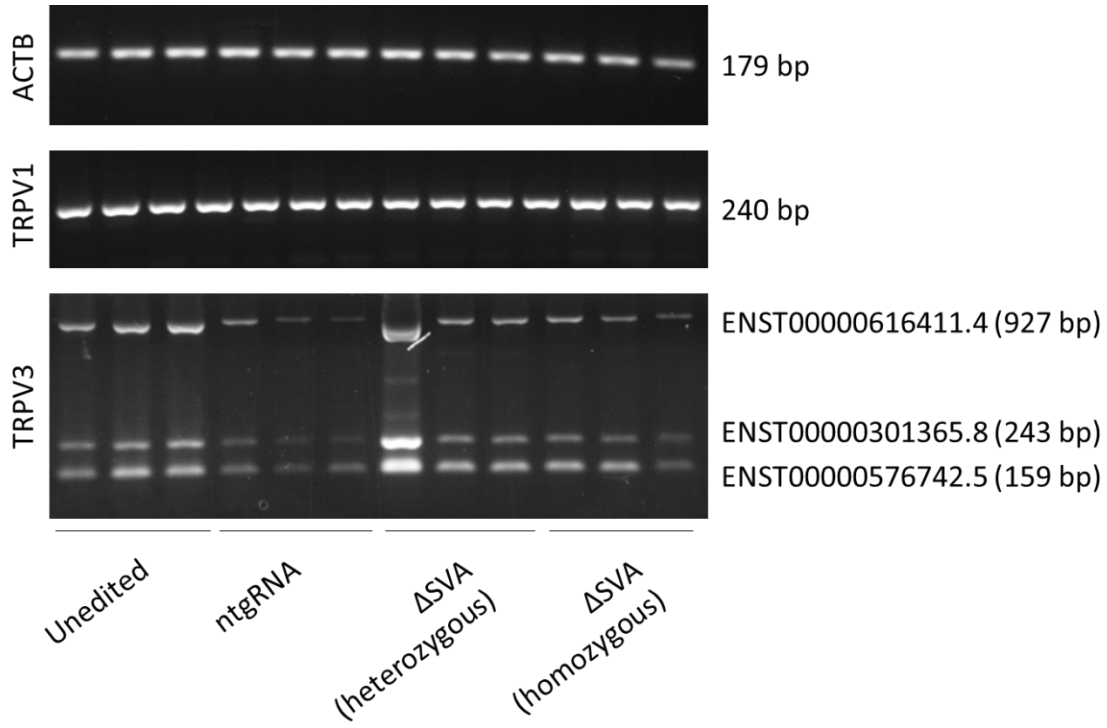


Figure 3.29. RT-PCR revealed no difference in TRPV1 or TRPV3 isoforms in CRISPR modified HEK293 cell lines. Multiplex PCR results to assay different isoforms of TRPV1 and TRPV3. All data was compared against ACTB.

3.4 Discussion

At present, genetic elements that contribute to human specific expression of *TRPV3* have not been identified. We identified a human specific SVA insertion at the intergenic region between *TRPV1* and *TRPV3* and hypothesised that it exerted *cis*-regulatory effects, potentially contributing to the human specific expression of *TRPV1* and *TRPV3*. Identification of human specific regulatory elements in genes implicated in pain will open an alternative avenue to explore the dysregulation of gene expression associated with chronic pain. This may help address the lack of translational research currently affecting the development of effective therapeutics.

Using *in vitro* reporter gene assays, we first demonstrated the SVA was functional as a transcriptional regulator in HEK293 cells (Figure 3.6). This was consistent with previous publications utilising reporter gene assays to demonstrate SVAs function as transcriptional regulatory domains including; the SVA at Amyotrophic Lateral Sclerosis (ALS) associated gene *FUS*[53] and the SVA at Parkinson's Disease associated gene *PARK7*[215]. We did not obtain any strong evidence to suggest the ECR adjacent to the SVA was functional as a transcriptional regulatory domain in HEK293, however this may be a cell dependent effect and should be tested in additional cell lines. In this cell model, the reporter genes assay data demonstrated strong transcriptional effects of the SVA compared to the ECR (Figure 3.6). Testing the same SVA reporter construct in mouse DRG primary cell cultures also confirmed that the SVA had the potential to function as a transcriptional regulator in a neuronal context (Figure 3.8). These results indicated that the SVA has the capacity to bind transcription factors present in mouse DRG and function as a regulatory element or affect DNA structure affecting transcription factor binding at the SV40 promoter within the vector. Human DRG were not obtainable during this study, therefore mouse DRG were used as a model. A recent RNA-seq study conducted by Ray *et al.* identified that transcription factor expression profiles in human and mice DRG were largely conserved, therefore it could be hypothesised that the SVA would

also exert regulatory function in human DRG[6]. Is important to note that whilst the expression of transcription factors in human and mouse DRG were largely conserved, other genes expressed in DRG did show species specificity (i.e. expressed in one species DRG, but not the other) – highlighting transcriptomic differences between the tissues of both species. Ray *et al.* only compared protein coding transcripts, thus the expression conservation of other factors, e.g. lncRNAs, which may also impact transcription in DRG is currently unknown[6,229]. Therefore, the results from the reporter gene assay conducted in mice DRG cannot be extrapolated to human DRG and further studies are required (discussed later).

Whilst reporter gene assays are useful to assess potential regulatory activity, they do not prove that the sequence of interest will exert *cis*-regulatory effects on a gene of interest. Previous studies have also demonstrated that regulatory activity is variable in reporter gene assays dependent on the cell line used[53,215]. Therefore, we deleted the SVA from human cell lines using CRISPR to directly measure the relationship between SVA presence/absence and *TRPV1* and *TRPV3* expression and found that *TRPV3* mRNA expression was mainly affected (Figure 3.26). We successfully developed a CRISPR protocol utilising a dual gRNA targeting approach to generate modified HEK293 cell lines containing SVA deletions (Figure 3.22). We also attempted to generate cell lines containing ECR deletions however we were unsuccessful (Figure 3.23). Sequencing of modified clones containing SVA deletions identified a high level of modification accuracy using this system. The repair of double strand breaks is mediated through non-homologous end joining (NHEJ), which is considered error prone in a mutagenic context. A study of NHEJ following CRISPR has reported at least 50% accuracy in repair of sites less than 150 bp apart[230]. The generation of large-scale deletions using CRISPR has been demonstrated by others (e.g. deletion of lncRNAs and even generating haploid cell lines) however the DNA repair efficiency following larger deletions is not well characterised[221,231]. This study indicated that NHEJ accuracy is maintained when performing larger deletions.

A recent study used CRISPR to delete an intronic SVA insertion in the *TAF1* gene (causative of X-linked Dystonia Parkinsonism), and also observed an increase in gene expression of *TAF1* in modified cell lines (<2-fold). The trend was consistent with the data presented in this chapter however the fold change observed in *TRPV3* mRNA expression is our study was much greater (up to 10-fold in one cell line) (Figure 3.26). The SVA in the *TAF1* study is a retrotransposon insertion polymorphism that affects a small population in the Phillipines[54,232,233]. The reported modification efficiency in that study was 30%, however it is important to note that this was performed in patient derived cell lines that contained a single copy SVA insertion[232]. The modification efficiency achieved in the work presented in this chapter was 10% for a heterozygous deletion and 1% for a homozygous deletion, highlighting the difficulty in using CRISPR to achieve the necessary modifications, especially deletion of more than a single copy. Low modification efficiencies have been discussed at local CRISPR workshops held within the Institute of Translational Medicine and are common to other researchers within the institute when using cell lines. There are currently no other published reports of SVA function being assessed using CRISPR (at the time of writing).

Using qPCR, we identified that deletion of some SVA copies resulted in an increase in total *TRPV3* mRNA expression and deletion of all SVA copies resulted in a decrease in total *TRPV3* mRNA expression (Figure 3.26)- highlighting the SVA as a transcriptional regulator of *TRPV3*. The literature describes four copies of chromosome 17, therefore there are potentially four copies of the SVA, however to confirm if HEK293 were tetraploid at chr17p13.2, FISH would need to conducted[227]. There is conflicting evidence regarding the effect of long-term cultivation of HEK293 on karyotype changes therefore the results should be interpreted with caution as we did not karyotype our cell line during this study[224,226]. HEK293 cell lines carrying heterozygous modifications presented the largest variation in *TRPV3* expression compared to unmodified and homozygous cell lines. It was not possible to fully characterised whether the SVA deletion was monoallelic, biallelic or triallelic in

heterozygous cell lines, however we propose the variability in expression in heterozygous cell lines may be due to different number of deleted SVA copies across the four copies of chromosome 17. Sequencing revealed that heterozygous cell lines had different repair breakpoints following NHEJ (Figure 3.21). NHEJ DNA repair can induce methylation changes, which could potentially have contributed to differences in mRNA expression between clones with different DNA repair breakpoints[234]. We did not obtain any evidence from the CRISPR experiment to suggest the deletion of the SVA had any major effect on *TRPV1* mRNA expression however these experiments were only conducted under basal conditions. We cannot discount it as a potential regulator in response to stimuli until further experiments have been completed. The SVA is located approximately 400 bp downstream of the TRPV1 3' UTR, which could potentially affect mRNA stability. Therefore, TRPV1 protein levels should be assayed to assess if the SVA affects mRNA stability and consequently protein levels.

The changes exerted to this genomic region via the deletion of the SVA may have resulted in epigenetic or structural changes that ultimately altered the expression of *TRPV3*. DNA methylation against transposable elements is a defence mechanism evolved by the host to silence retrotransposon activity[38]. Studies have reported that SVAs are typically methylated in the human genome due to their high GC content within the VNTR region[235,236] (typically thought of as mobile CpG islands). In chapter 4, we present data which highlights the SVA (specifically the *Alu*-like sequence) as a putative CpG island within the intergenic region (Figure 4.16). Thereby we propose that deletion of the SVA may alter the epigenetic status of the region, affecting the capacity of the ECR to bind transcription factors that could lead to changes in expression of *TRPV3*. Upon reflection, profiling the methylation status of the intact SVA in modified and unmodified cells would have strengthened this interpretation greatly. Follow up studies should repeat the reporter gene assays and measure *TRPV1* and *TRPV3* mRNA expression in CRISPR modified cell lines in response to a relevant stimulus, e.g. capsaicin or camphor, to give insight into the capacity of the SVA as an epigenetic modulator

of gene expression. To summarise, the data from this CRISPR assay and subsequent qPCR gave evidence to support the role of the SVA as a *cis*-regulatory element at the *TRPV1* and *TRPV3* locus.

Whilst trends were observed in *TRPV3* expression, statistical significance was not reached. A limiting factor in this study is the low number of modified cell lines generated, reducing sample number available for qPCR assays. As gene expression can be variable between independently grown cell cultures, this study would be greatly improved by the generation of additional modified cell lines. A total of 23 heterozygous ΔSVA cell lines were identified, however only 3 were utilised in initial qPCR experiments. All available cell lines should be assayed to increase sample size and further explore the relationship between specific allelic deletions and effects of *TRPV3* expression. PacBio long read sequencing has shown that aside from simple insertions and deletions, large complex rearrangements can result from repair of DSB introduced by CRISPR, which cannot be detected by short range PCR assays[237]. Profiling off-target effects was not conducted during this study however it is recommended that this is addressed in future studies via PCR amplification of predicted off target sites, followed by sequencing to verify any indels, as described by Hay *et al*[238]. Further recommendations include profiling the methylation status of the remaining alleles in cell lines carrying a heterozygous modification. This may address if some of the variability in gene expression is due to the epigenetic status of the remaining intact locus, which in turn may rule out the impact of off-targets effects on target gene expression.

The simplistic hypothesis in this chapter focused on whether the SVA could function as a transcriptional regulator. The wider hypothesis was that the SVA insertion contributed to human specific expression of *TRPV3* in the DRG, as documented in the literature. Reflecting on the current data presented in this chapter, we have given evidence to support the role of the SVA as a transcriptional regulatory domain. If we consider that the mouse

genome does not contain SVAs, it could be hypothesised that presence of SVAs in the human genome could influence transcriptional changes in the human DRG that are not observed in the mouse – consistent with our original hypothesis that it functions as a human specific regulator at the *TRPV1* and *TRPV3* locus and could potentially contribute to the expression of *TRPV3* in human DRG but not in the mouse. A recent study identified differential histone methylation and acetylation profiles across SVA subclasses in different tissues during embryonic gene activation and in adulthood, supporting the role of SVAs as tissue specific enhancers[217,239,240]. This could therefore potentially play a role in tissue specific regulation at this locus. To address the wider hypothesis properly, the CRISPR protocol described here should be repeated in a more appropriate *in vitro* model. This could be achieved by differentiating human iPSCs into peripheral sensory neurones[241].

Previous publications and the data presented in this chapter gives evidence that SVAs act as transcriptional regulators in the human genome. The work presented here is the first reported example of a reference SVA being deleted using CRISPR and is also the first report of an SVA affecting gene expression in a non-disease model. We demonstrated *cis*-regulatory effects on *TRPV3* expression, which may contribute to the lack of translational research using preclinical mouse models. In addition, we suggest the epigenetic status of this regulatory element should be profiled to further explore mechanisms contributing to *TRPV3* and *TRPV1* expression, which is currently poorly understood in a molecular pain context.

Chapter 4.

Genotyping the SVA at the *TRPV1* & *TRPV3* locus and associations with pain phenotypes

4.1 Introduction

Transposable elements offer a large source of genetic variation that is currently under-characterised in the human genome, due to difficulties in mapping repetitive DNA sequences from short read sequencing. In addition to their functional role, SVAs can be polymorphic in the general population; either as a sequence length polymorphism and/or as a RIP[53,242]. Recent work identified a RIP SVA insertion in *TAF1* intron 32 was causative of X-linked Dystonia Parkinsonism (XDP)[232,233]. Furthermore, this SVA RIP also displayed sequence length polymorphism within the TCTCCC hexamer repeat domain (copy number range 35 – 52), and that increasing hexamer copy numbers associated with reduced age of onset of XDP (range 35 – 55 years), highlighting the relevance of SVA polymorphism in relation to phenotype[54]. This is a similar notion to the well-known CAG repeat expansion in the Huntingtin gene HTT, whose expanded alleles affect age of onset of Huntington disease[243]. The *TAF1* SVA was later shown to affecting the expression of *TAF1* mRNA however the effects of SVA sequence length polymorphism on *TAF1* mRNA expression are unknown. Nevertheless, the effect of SVA polymorphism on gene expression have been proposed to contribute to individual differences and potentially disease risk[244]. The *TAF1* SVA demonstrates an example of a RIP being a dominant cause of a monogenic disease.

At present, genetic variation at the *TRPV1* and *TRPV3* locus has only been partially characterised by SNPs associated with sensory phenotypes. *TRPV1* SNP rs222747 (coding missense, C=0.26/G=0.73, 1000 Genomes) is associated with cold hyperalgesia in neuropathic pain patients[245]. *TRPV1* SNP rs222741 (intronic, G=0.25/A=0.75 1000 Genomes) with migraine[246]. *TRPV1* SNP rs8065080 (coding missense, T=0.68/C=0.31, 1000 Genomes) is associated with increased sensitivity to capsaicin[247,248]. *TRPV1* SNPs rs4790521 (3' UTR, C=0.68/C=0.31, 1000 Genomes) and rs4790522 (3' UTR, A=0.43/C=0.56, 1000 Genomes) have been associated with TRPV1 mRNA expression changes in a study of asthma, however the functional mechanism has not been identified[249]. Interestingly, these SNPs are both in

the 3'UTR of *TRPV1*, and are both in close proximity to the SVA (Supplementary figure 2). SNPs associated with disease essentially function as a tagging SNP, indicating a causal variant that is in linkage disequilibrium (LD) with said SNP. LD is the non-random association of alleles at two or more loci in the general population, therefore tagging SNPs are a useful tool to indicate presence of a causal variant if LD can be established[250]. Therefore, despite numerous TRPV1 and TRPV3 SNPs identified in human cohorts to date, which associate with either changes in TRPV1/TRPV3 mRNA expression or pain phenotypes, the causal variants underlying these phenotypes (molecular or clinical) are yet to be identified in some examples.

SVAs are hypothesised to function as CREs and can modulate gene x environment interactions via epigenetic modification[251]. Due to their high GC content, SVAs are described as mobile CpG islands and DNA methylation is one mechanism by which gene expression can be mediated via proximal SVA insertions[26]. Hypermethylation of the SVA insertion at the *TAF1* resulted in decreased *TAF1* mRNA expression, causative of XDP[252]. Differential methylation of *TRPV1* (intron chr17:3425000–3426000, hg19) and *TRPV3* (intron chr17:3398500–3399500, hg19) has been associated with heat pain-sensitivity in a study of monozygotic twins (median age 62 years)[253]. This implicates that methylation status at the *TRPV1* and *TRPV3* locus is a potential driver of pain in non-pathological ageing. No study has yet looked at the methylation status in association with genetic variants at this locus. Direct links between methylation status and *TRPV1*/*TRPV3* mRNA expression have not been explored either.

In chapter 3, we present data which suggests that the SVA insertion at the *TRPV1* and *TRPV3* locus functioned as a CRE *in vitro.* If shown to be polymorphic, this could lead to gain or loss of transcription factor binding sites or differential modulation by epigenetic mechanisms such as DNA methylation. Ultimately, this could carry functional implications for *TRPV1* and *TRPV3* mRNA regulation between individuals and in response to stimuli.

To address potential associations between SVA polymorphism and pain related phenotypes, we obtained access to two distinct groups. As pain is associated with ageing, the first group we genotyped was the Dyne-Steele cohort; a cohort of healthy aged individuals who were given a pain questionnaire. The second group we genotyped was a cohort of osteoporosis patients that were taking part in a trial that measured adverse drug responses to treatment with tramadol.

Osteoporosis is an age-associated disease characterised by decreased bone mass, increased risk of bone fracture and chronic pain[254]. This cohort had corresponding pain data from a number of commonly used psychometric pain scales, including the visual analogue scale[255] and the McGill pain questionnaire[256], thereby providing an opportunity to associate genetic polymorphism with phenotypic pain data. Tramadol is a synthetic opioid used to treat moderate to severe pain across a wide range of conditions including bone pain in osteoporosis[257]. The primary mechanism of action of tramadol is via metabolism in the liver, by the enzyme CYP2D6, which produces the metabolite O-desmethyltramadol (M1), an agonist of the μ-opioid receptor that blocks ascending pain pathways[257].

In vitro studies in Chinese hamster ovary (CHO) cells, have shown that Tramadol also functions as an agonist of TRPV1[258]. This off-target effect has been proposed as a potential mechanism that may contribute to adverse drug reactions (ADRs), often experienced when taking tramadol[257,258]. Several studies have investigated correlations between tramadol efficacy and genetic polymorphisms in its metabolising enzyme *CYP2D6*[259], however no study has investigated polymorphisms associated between ADRs and the *TRPV1* and *TRPV3* locus. To further this point, TRPV1 is expressed in osteoclasts (bone cells) and has been shown to contribute directly to bone pain[260]. TRPV1 is also implicated in bone resorption in osteoporosis mouse models and immunoreactivity of TRPV1 is increased in the DRG in osteoporosis rat models[261-263]. Therefore, TRPV1 also plays a role in the pathogenicity of the

osteoporosis itself, alongside mediating pain. The osteoporosis cohort used in this chapter was therefore suitable to explore associations between genetic polymorphisms encoded within the SVA at the *TRPV1* and *TRPV3* locus, with pain phenotypes and adverse drugs reactions in response to tramadol. A recent study in the USA highlighted that people receiving tramadol are at a greater risk of developing long term opioid addiction compared to other opioids and more than 10% of people experience adverse drug responses [257]. This demonstrates the need to explore genetic variants that may be able to predict the risk of ADRs in response to tramadol, in an attempt to address the poor management of opioid use.

In this chapter, we screened the SVA at *TRPV1* and *TRPV3* for polymorphism and methylation status. Development of a simple PCR protocol enabled us to identify SVA sequence length polymorphism present within the CT-rich domain. DNA sequencing validated our findings and identified an unexpected repeat sequence contributing to length polymorphism, in addition to the previously published "hexameric" copy number. A specific heterozygous SVA genotype was shown to significantly associate with self-reported pain in non-pathological age-associated pain and similar trends were also observed with responses to tramadol in osteoporosis patients. Finally, we demonstrated differential methylation of the SVA between individuals and highlight further studies to explore the relationship between SVA genotype, epigenetic modification and phenotypic data.

4.2 Hypotheses and aims

We hypothesised that the reference SVA present at the *TRPV1* and *TRPV3* locus may be polymorphic (based on sequence length) in the general population. We further hypothesised that any potential polymorphism could be utilised in genetic association studies and be tested for associations with complex disease phenotypes in a similar manner to SNPs.

The aim of this work was to characterise genetic and epigenetic variation of the SVA encoded at the intergenic region between *TRPV1* and *TRPV3*, and also identify potential associations with phenotypes relevant to pain, by completing the following:

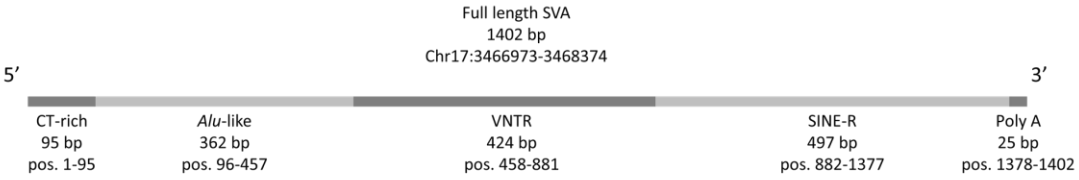
1. Characterise SVA genetic and epigenetic variation:
    a. Identify if the SVA is polymorphic in sequence length in the general population
    b. Characterise the polymorphism within composite sequences of the SVA
    c. Assess methylation of the SVA in blood and brain from the same individuals
2. Identify potential associations between SVA polymorphisms and relevant pain phenotypes:
    a. Genotype the SVA in non-pathological ageing cohort
        i. Identify associations between SVA genotype and self-reported pain measures
        ii. Identify associations between SVA genotype and sex
    b. Genotype SVA in osteoporosis cohort (with bone pain) treated with tramadol
        i. Identify associations between SVA genotype and adverse drug responses to tramadol
        ii. Identify associations between SVA genotype and analgesic efficacy of tramadol
        iii. Identify associations between SVA genotype and sex

4.3 Results

4.3.1 Characterising genetic variation in the SVA at *TRPV1* and *TRPV3*

4.3.1.1 Annotation of composite sequences within SVA

We defined the composite sequences that make up the full length SVA (Figure 4.1). The reference allele contains a 95 bp CT-rich domain at the 5' end (position 1–95), which consists of 8 hexamer repeats and additional bases CATGG between the 5th and 6th hexamer repeats, and TCTCTTTCCACGG between the 7th and 8th hexamer repeat. Adjacent to the CT-rich domain is a 362 bp *Alu*-like sequence (position 96–457). There is a central 424 bp Variable Number Tandem Repeat (VNTR) region (position 458–881). The VNTR sequences are variable in size 36-41 bp. There are 10 copies of the VNTR. Adjacent to the VNTR region towards the 3' end is a 497 bp SINE-R region (position 882–1377). At the 3' end of the SVA there is a 25 bp poly A region (position 1378–1402). Figure 4.1 provides a full overview of this annotation.



CT-rich

TCTCCC TCTCCC TCTCCC TCTCCC TCTCCC TCTCCC TCTCCC CATGG TCTCCC TCTCCC GATGG TCTCCC TCTCCC TCTCTTTCCACGG TCTCCC

Alu-like

TCTCATGCCGAGCCGAAGCTGGACTGTACTGCTGCCATCTCGGCTCACTGCAACCTCCCTGCCTGATTCTCCTGCCTCAGCCTGCCCAGTGCCTGCGATTGCAGGCG
CGCGCCGCCACGCCTGACTGGTTTTCGTATTTTTTTGGTGGAGACGGGTTTCGCTGTGTTGGCCGGGCTGGTCTCCAGCTCCTAACCGCGAGTGATCCGCCAGCCT
CGGCCTCCCGAGGTGCCGGGATTGCAGACGGTGTCTGGTTCACTCAGTGCTCAATGGTGCCCAGGCTGGAGTGCAGTGGCGTGATCTCGGCTCACTACAACCTCC
ACCTCCCAGCCGCCTGCCTTGGCCTCCCAAAGTGCAAAGCTTGCAGCCTCTGCCTGGCCGCCCATCGTCTGGGATGTGAGG

VNTR

AGCCCCTCTGCCTGGCTGCCCCGTCTGGAAAGTGAGG
ACTGTCTCCGCCCGGCCGCCCATCGTCTGAGATGTGGGG
AGCGCCTCTGCCCCGCCGCCCCATCTGGGATGTGAGG
AGCGCCTCTGCCCGGCCACGACCCCGTCTGGGAGGTGAGG
AGCGTCTCTGCCCAGCCGCCCCGTCTGAGAAGTGAGG
AGCCCCTCCACCCGGCAGCTGCCCCTACTGGGAAGTGAGG
AGCGTCTCCGCCCGGCAGCCACCCCGTCCAGGAAGG
AGGTGGGGGTCACCCACCGCCAGGCCAGCCGCCCCGTCCGGG
*AGG*TGAGGGGCGCCTCTGCCTGGCCGCCCCTACTGGGAAGTGAGG
AGCCCCTCTGCCCGGCCACCACCCCGTCTGGGAG

SINE-R

GTGTGCCCAACAGCTCATTGAGAACGGGCCATGATGACAATGGCGGTTTTGTGGAATAGAAAAGGGGGAAAGGTGGGGAAAAGATTGAGAAATCGGATGGTTG
CCGTGTCTGCGTAGAAAGAAGTAGACATGGGAGACTTCATTTTGTTCTGTACTAAGAAAAATTCTTCTGCCTTGGGATCCTGTTGATCTGTGACCTTACCCCCAACCC
TGTGCTCTCTGAAACATGTGCTGTGTCCACTCAGGGTTAAATGGATTAAGGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCGTTAAG
AGTCATCACCACTCCCTAATCTCAAGTACCCAGAGACACAAACACTGCGGAAGGCCGCAGGGTCCTCTGCCTAGGAAAACCAGAGACCTTTGTTCACTTGTTTATCT
GCTGACCTTCCCTCCACTATCGTCCTATGACCCTGCCAAATCCCCCTCTGTGAGAAACACCCAAGAATGATC

poly A

AATAAAAAAATAAATTAAAAAAAAA

Figure 4.1. Annotation of composite sequences within SVA. (Above) Schematic representing the full length SVA and its composite sequences from 5' to 3' including a CT-rich domain, an Alu-like sequence, a central Variable number Tandem Repeat (VNTR), a SINE-R region and a poly A region. This was drawn to scale. (Below) Annotation of composite sequences using the complete SVA DNA sequence obtained from UCSC genome browser chr17:3466973-3468374 (hg19).

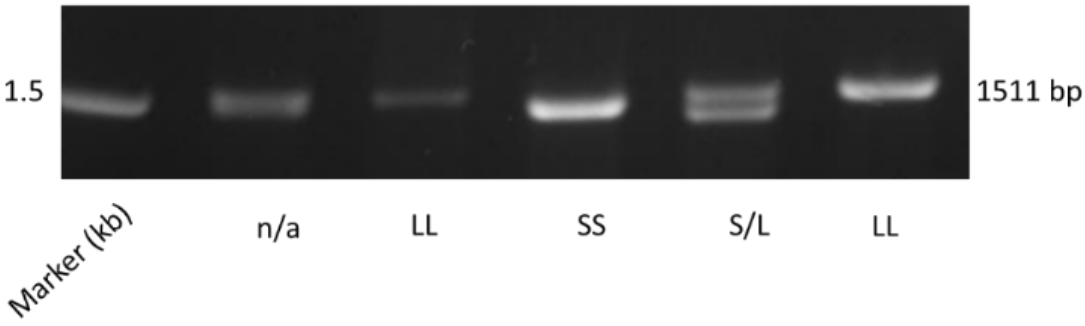
4.3.1.2 SVA is polymorphic based on sequence length

To screen for sequence length polymorphisms, we amplified the full length SVA (estimated product size 1511 bp) using PCR (methods section 2.2.2.1). Primer sequences and protocol are listed in the appendix A, entry 1. We used control samples (n=48) that were individuals from the general population and free from neurological disorders (refer to materials section 2.1.5.1). During this analysis, we observed at least two different sized products at the expected 1.5 kb size marker, named long (L) and short (S), estimated to be approximately 30 bp apart (Figure 4.2A). Due to the large sizes of the PCR products and the small difference observed, it was difficult to fully resolve them using agarose gel electrophoresis, and accurately identify the genotype. 19% (n=9) of samples could not be accurately identified, suggesting there were potentially more than two variants, however in those that were easily distinguished, we observed 23% (n=11) of individuals were homozygous short (SS), 33% (n=16) were homozygous long (LL) and 13% (n=6) were heterozygous short/long (S/L) (Figure 4.2B and C). These frequencies indicated that this SVA polymorphism was a common genetic variant (>1% frequency).

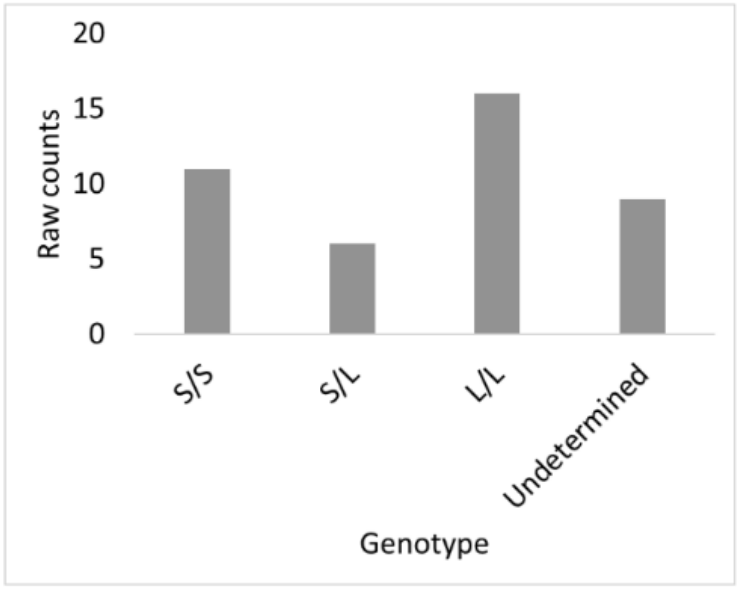
We then designed a PCR strategy to amplify composite sequences including the 5' CT-rich domain, the central VNTR and the 3' poly A region (Figure 4.3). All primer sequences and thermal cycles are listed in appendix A, entries 3 – 5. We amplified each region using the same control samples (n=48) used for the full length SVA and screened them using capillary electrophoresis to achieve a higher resolution than standard agarose gel electrophoresis. We identified a minimum of 4 distinct PCR products for the 5' CT-rich domain (expected product size 424 bp) at sizes 405 bp, 427 bp, 453 bp and 471 bp (Figure 4.4). We identified a single PCR product for the central VNTR (478 bp) and the 3' poly A region (312 bp) therefore we obtained no evidence to suggest these regions were polymorphic in size (Figure 4.4). As we observed polymorphism in the full length SVA but found no evidence to support

polymorphism in the VNTR or poly A in the same individuals, we focused on the CT-rich domain as the main source of polymorphism within this SVA.

A



B



C

Genotype	Count	Frequency
S/S	11	26%
S/L	6	14%
L/L	16	38%
Undetermined	9	21%
Total	42	n/a

Figure 4.2. Common sequence length polymorphism identified in SVA at *TRPV1* and *TRPV3*. (A) Agarose gel electrophoresis results showing examples of different sized products identified at expected product size for full length SVA. Lane 2 (denoted as n/a) indicates undetermined sample, due to poor resolution. LL indicates homozygous long genotype. SS indicates homozygous short genotype. S/L indicates heterozygous short/long genotype. (B) Raw genotype counts observed in small sample cohort (n=48). (C) Summary table of raw counts and frequencies.

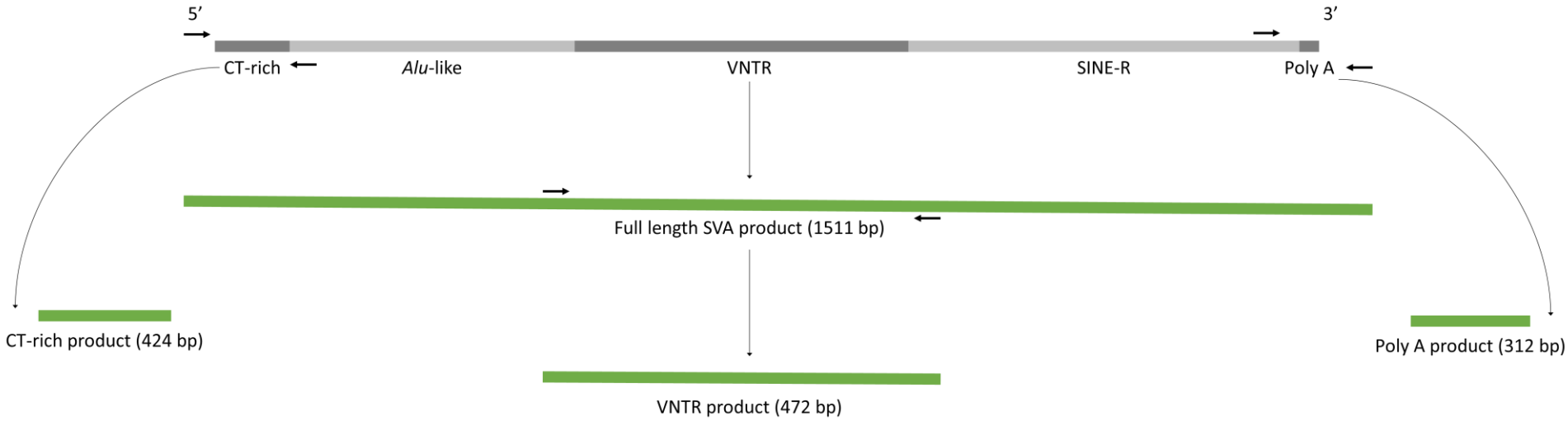


Figure 4.3. PCR strategy developed to amplify composite sequences within SVA. Primers were designed flanking the SVA at the 5’ and 3’ end. Internal primers were then designed in the *Alu*-like region to amplify the 5’ CT-rich domain and in the SINE-R region to amplify the 3’ poly A region. To amplify the central VNTR, the full length SVA was amplified first and then used as a template to amplify the VNTR using primers designed in the *Alu*-like and SINE-R region. Drawn to scale.

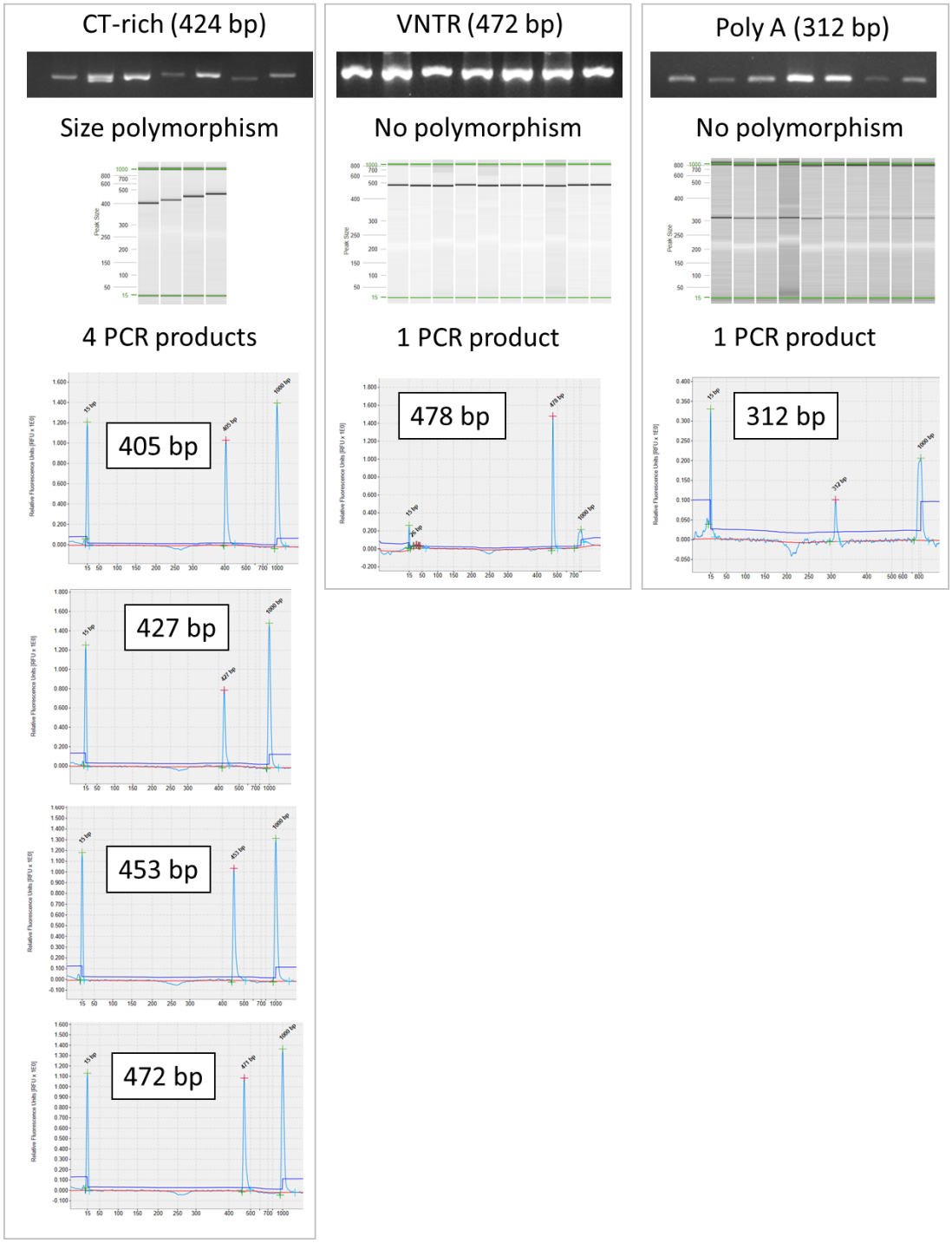


Figure 4.4. Screening of composite sequences revealed size polymorphism in CT-rich domain. PCR products identified for CT-rich domain (expected size 424 bp), VNTR (expected size 472 bp), and poly A region (expected size 312 bp). (Above) Examples of agarose gels which show relatively poor resolution. (Central) Digital gels from capillary electrophoresis show different size PCR products for CT-rich domain and a single product for both VNTR and poly A. (Below) Examples of raw data from capillary electrophoresis.

4.3.1.3 Sequencing polymorphic variants revealed a novel 17mer repeat in the CT-rich domain

To validate that the polymorphism observed using electrophoresis was due to the CT-rich sequence itself, and not due to variation within the flanking regions of the PCR product, we sequenced PCR products corresponding to different alleles. We named each PCR product based on increasing product size, a (404 bp), b (424 bp), c (453 bp) and d (478 bp). The sequence for allele b was taken from the reference genome as this was the expected size based on *in silico* PCR (hg38). The sequence for alleles a and c were generated by sanger sequencing (refer to methods section 0). Allele d could not be sequenced, perhaps due to complexity in secondary structure that could not be resolved. Sequences were aligned to visualise differences (Figure 4.5). We identified that the reference allele b was 95 bp in length and contained 12 TCTCCC hexamer repeats, alongside 2 interspersed C/GATGG repeats. The shorter allele a was 66 bp, contained only 8 TCTCCC hexamer repeats and contained a single C/GATGG repeat. The longer c allele was 111 bp, contained 13 TCTCCC hexamer repeats and 4 interspersed C/GATGG repeats. The size differences from the shortest to largest allele were as follows; 20 bp, 29 bp and 25 bp). This was in line with the approximate 30 bp differences observed in the full length SVA which were difficult to resolve using agarose gels.

After aligning the sequences, it was evident that much of the polymorphism was not due to differences in copy number of TCTCCC hexamers alone, but also due to a novel 17mer repeat composed of C/GATGG followed by 2 TCTCCC hexamers (Figure 4.5). The shortest allele a contained a single 17mer repeat and 8 hexamers. The reference allele b contained 2 17mer repeats and 12 hexamers. The larger allele c contained 4 17mer repeats and 13 hexamers. In the context of this SVA example, most of the polymorphism between alleles was supplied to the allele via differences in the copy number of 17mer repeats, not just hexamer repeats.

To assess if the 17mer repeat was a novel sequence in this particular SVA or if it occurred in others, we took a 34 bp sequence corresponding to 2 17mer repeats and used BLAT to identify other sequences in the genome (hg38). It was not possible to detect a single repeat using BLAT as it required a minimum input of 20 nt. We identified 55 SVAs in the human genome that contained at least 2 copies of this 17mer repeat (Figure 4.6). We observed that most 17mer repeats were present in SVA D. Relatively few copies were identified in the evolutionarily older subclasses SVA A, B and C. No SVA E were found to contain this 17mer repeat. To assess if this repeat was specific to human SVAs, we also screened the chimpanzee (panTro6) and gorilla (gorGor5) genomes (Figure 4.6). We identified 15 SVAs in the chimpanzee genome and 5 SVAs in the gorilla genome that contained this repeat, therefore it is not human specific. Across all SVA D in the human genome, 32.5% are conserved with the chimpanzee[15], however out of all human SVA D containing the 17mer repeat (n=46), only 13% were conserved with the chimpanzee. This highlighted that the 17mer repeat was overrepresented in human specific SVA D.

To summarise, sequencing data confirmed the polymorphism observed in the PCR product was present in the CT-rich domain and not the flanking sequences. As a further validation step, we resolved the characterised alleles using agarose gel electrophoresis, for visual comparison against capillary electrophoresis digital gel images (Figure 4.7A and B). An overview of the variation including PCR product sizes, CT-rich domain length, hexamer repeat content and 17mer repeat content is given in the table in Figure 4.7C. Here we demonstrated a minimum of 4 distinct alleles of this SVA at *TRPV1* and *TRPV3* that are present in the general population, which we used as a basis for genotyping studies. A schematic is shown in Figure 4.7D.

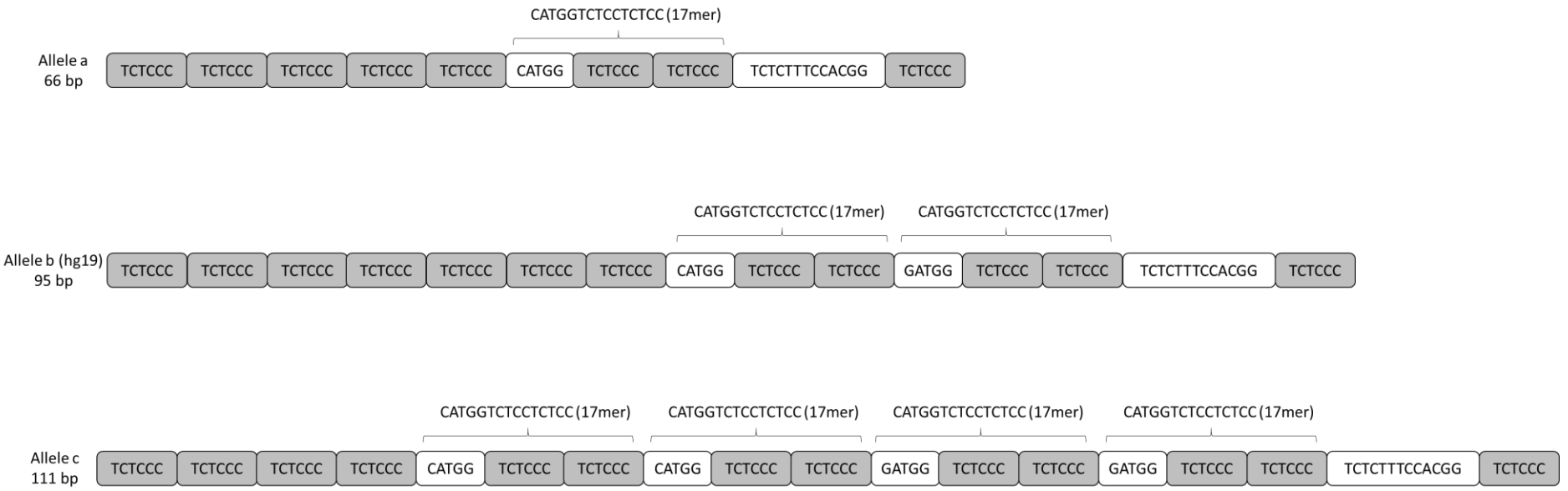
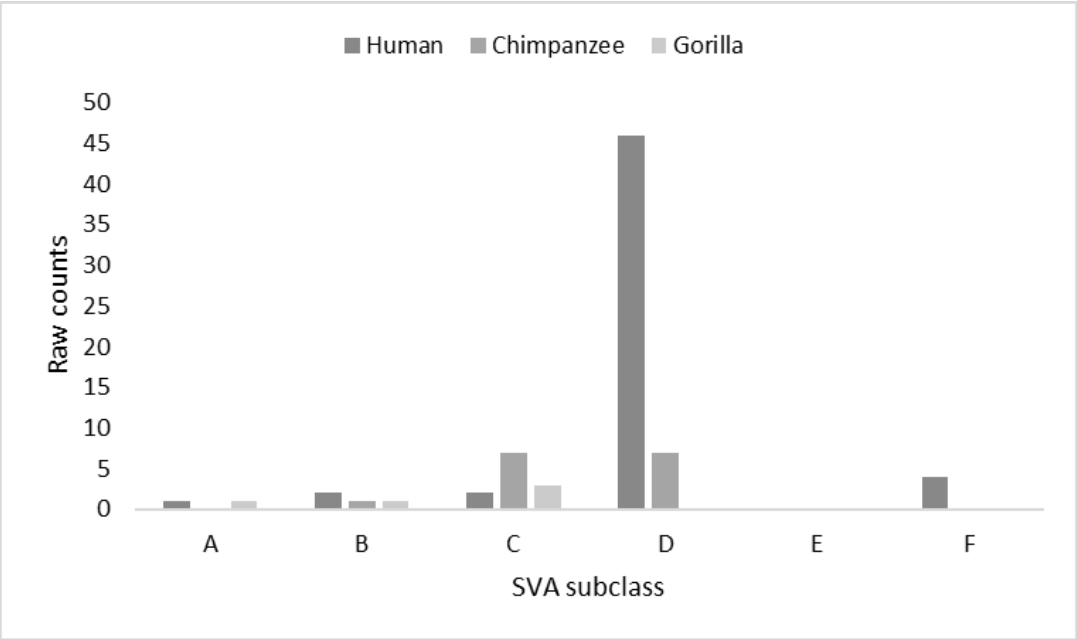
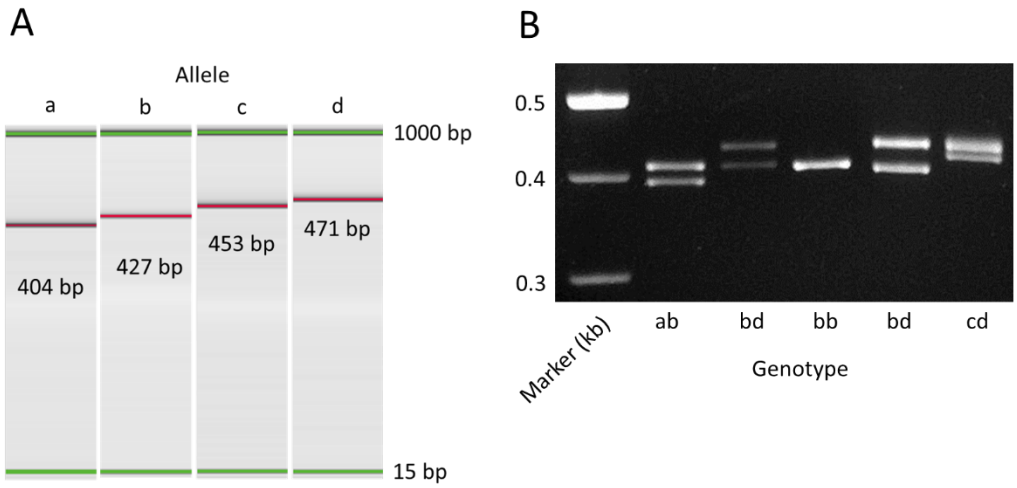


Figure 4.5. Sequence variation within CT-rich repeats identified 17mer repeat contributing to polymorphism. The canonical repeat in the CT-rich domain has been extensively described as a hexamer CCCTCT repeat. Here we demonstrate that the polymorphism supplied to the SVA is a combination of hexamer repeat variation however have also identified a 17mer repeat that contributes to the polymorphism across all sequenced alleles.



SVA	Human	Chimpanzee	Gorilla
A	1	0	1
B	2	1	1
C	2	7	3
D	46	7	0
E	0	0	0
F	4	0	0
Total	55	15	5

Figure 4.6. Number of SVAs containing 17mer repeat in CT-rich domain across primate genomes. Raw counts of SVAs containing at least two copies of the 17mer repeat in the human genome (hg38), chimpanzee genome (panTro6) and gorilla genome (gorGor5).



C

Allele	PCR product size (bp)	CT-rich size (bp)	Hexamer repeat number	17mer repeat number
a	404	66	8	1
b	427	95	12	2
c	453	111	13	4
d	471	n/a	n/a	n/a

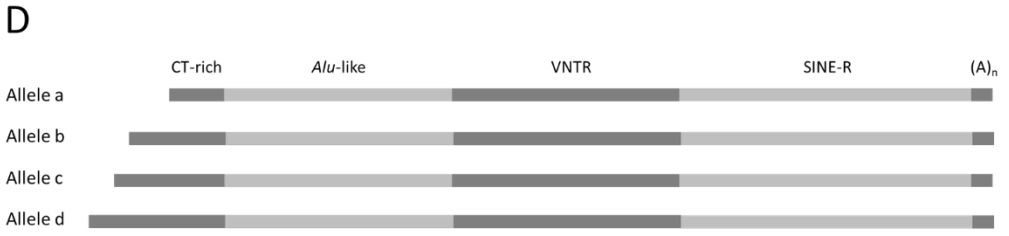


Figure 4.7. Summary of polymorphic variants identified of the SVA at TRPV1 and TRPV3. (A) Summary of four alleles identified based on capillary electrophoresis. (B) Homozygous and heterozygous genotypes identified containing the four alleles. (C) Summary table of PCR and sequencing data for each allele. (D) Schematic representing the four distinct alleles containing variable length CT-rich domains at the 5' end of the SVA.

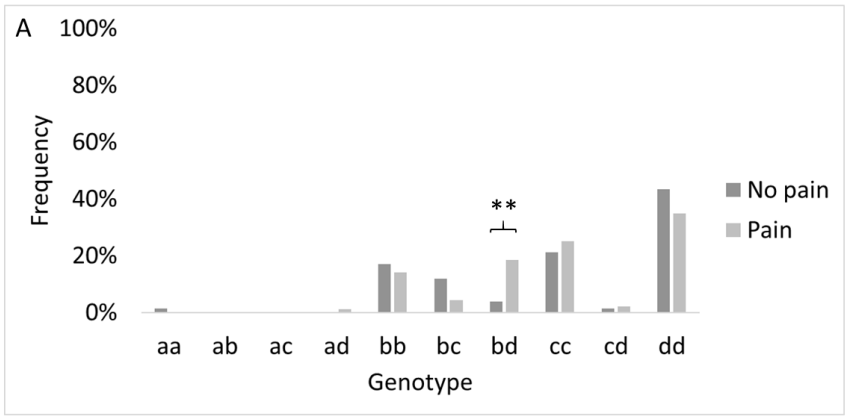
4.3.2 Genotyping studies

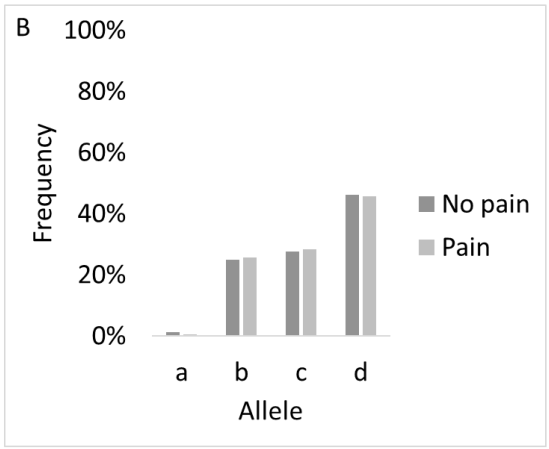
4.3.2.1 SVA genotype associates with self-reported pain measures

The CT-rich domain in individuals from the Dyne-Steele healthy ageing cohort (>50 years, mean = 66 years) was genotyped. Individuals (n=336) were subject to a pain questionnaire which assessed whether they were experiencing pain or free from pain. They were asked two questions: 1) if they had experienced pain in the last month, 2) had they experienced pain today. If participants answered “yes” to either question they were placed in the self-reported pain group. If participants answered “no” to both questions they were placed in the pain free group.

We assessed if there were any associations between SVA genotype at the *TRPV1* and *TRPV3* locus and self-reported pain. A total of 184 participants (55%) reported they experienced pain, and 152 (45%) did not. No differences in allele frequencies were observed between either group (Figure 4.8B) however we identified a significant increase in the frequency of the heterozygous bd genotype in individuals that self-reported pain (18%, n=17) compared to the absence of self-reported pain (4%, n=3) (P=0.004) (Figure 4.8A). The sample size here is low, therefore the likelihood of a type II error (i.e. falsely rejecting the null hypothesis) resulting from low statistical power is increased. However, based on the available data, the combination of b and d alleles together is the variant of interest and not each distinct allele. The most frequent allele was d (46%), followed by c (28%), b (26%) and a (1%). This reflected allele frequencies observed in the initial genotyping study using healthy controls; d (43%), c (18%), b (35%) and a (4%).

We stratified genotypic data based on sex as an independent variable and found no association between sex and any genotype or allele in this cohort (female n=282, male n=54, P>0.05) (Figure 4.9). These results suggested that in this cohort, the bd genotype is more likely to be associated with the occurrence of self-reported pain measures in aged individuals.

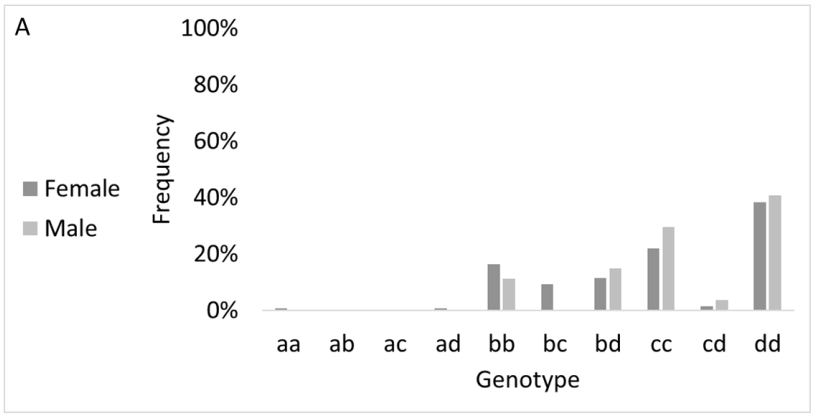


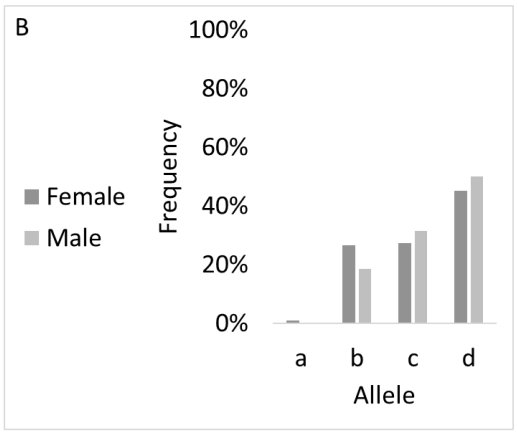


C

	No pain		Pain		
Genotype	Count	Frequency	Count	Frequency	P-value
aa	1	1%	0	0%	0.452
ab	0	0%	0	0%	1.000
ac	0	0%	0	0%	1.000
ad	0	0%	1	1%	1.000
bb	13	17%	13	14%	0.670
bc	9	12%	4	4%	0.086
bd	3	4%	17	18%	0.004
cc	16	21%	23	25%	0.586
cd	1	1%	2	2%	1.000
dd	33	43%	32	35%	0.269
Total	76	n/a	92	n/a	n/a
Allele	Count	Frequency	Count	Frequency	P-value
a	2	1%	1	1%	0.592
b	38	25%	47	26%	1.000
c	42	28%	52	28%	0.904
d	70	46%	84	46%	1.000
Total	152	n/a	184	n/a	n/a

Figure 4.8. Genotyping results showing association of SVA with self-reported pain measures. (A) Frequency of genotypes in Dyne-Steele cohort of aged individuals with self-reported pain measures; no pain (n=76) and pain (n=92). (B) Allele frequencies. (C) Summary tables showing exact counts and results from statistical analysis. Significant p-values indicated as <0.05*, <0.01**, <0.001*** (Fisher's exact two-tailed test).





C

	Female		Male		
Genotype	Count	Frequency	Count	Frequency	P-value
aa	1	1%	0	0%	1
ab	0	0%	0	0%	1
ac	0	0%	0	0%	1
ad	1	1%	0	0%	1
bb	23	16%	3	11%	0.7712
bc	13	9%	0	0%	0.1312
bd	16	11%	4	15%	0.5339
cc	31	22%	8	30%	0.4556
cd	2	1%	1	4%	0.4108
dd	54	38%	11	41%	0.8317
Total	141	n/a	27	n/a	n/a
Allele	Count	Female	Count	Male	P-value
a	3	1%	0	0%	1
b	75	27%	10	19%	0.2359
c	77	27%	17	31%	0.5133
d	127	45%	27	50%	0.5522
Total	282	n/a	54	n/a	n/a

Figure 4.9. Associations of SVA genotypes with sex in Dyne-Steele cohort. (A) Genotype frequencies in females (n=141) and males (n=27). (B) Allele frequencies. (C) Summary tables showing exact counts and results from statistical analysis. No statistical significance found (Fisher's exact two-tailed test).

4.3.2.2 SVA allele associates with males with osteoporosis

We genotyped 74 individuals with osteoporosis that were experiencing bone pain (54 female and 20 male). At the time of recruitment, patients were placed into group A (currently treated with tramadol for bone pain, n=42) or group B (previous adverse drug reaction to tramadol and therefore not treated, n=32). Adverse drug responses (ADRs) were monitored in group A and previous ADRs were listed in group B (Figure 4.11). In total, 48 participants (65%) experienced ADRs and 26 (35%) were tolerant to tramadol.

We first looked for associations between genotype and ADRs. We identified an increased frequency of the heterozygous bd genotype in individuals experiencing ADRs (25%) compared to tolerant individuals (8%) (P=0.118) (Figure 4.12). This was not significant, likely due to small sample size, but it is important to note that this genotype was significantly associated with the incidence of self-reported pain in the healthy ageing cohort detailed in the previous section 4.3.2.1 (P=0.004). There were no significant differences in allele frequencies between groups. Sample size was too small to address specific ADRs.

We also addressed associations between SVA genotype at the *TRPV1* and *TRPV3* locus and analgesic efficacy of tramadol (i.e. an improvement in pain scores following treatment). In group A (n=42), pain measures were assessed in response to tramadol using the visual analogue scale (VAS) (Figure 4.10). The difference in scores prior to and 3 hours post treatment were used as the criteria to test for associations between genotype and effective analgesia. 27 patients (64%) had a VAS difference ≥0 (increase in pain) after 3 hours, indicating that pain intensity had increased, or did not improve, following treatment, therefore tramadol did not produce the desired analgesic effect. 15 patients (36%) had a VAS difference <0 (decrease in pain), indicating that pain intensity had decreased following treatment, therefore tramadol had the desired analgesic effect. We observed an increase in the frequency of heterozygous bd genotype in individuals with a VAS difference <0 (good response, 26%) compared to individuals with a VAS difference ≥0 (poor response, 13%) but

this was not significant. This means that the bd genotype was more frequent in individuals that had no improvement in pain score and therefore responded poorly to tramadol, as it did not exert the desired analgesic effect. On the contrary, we observed an increase in the frequency of homozygous dd genotype in individuals with a VAS difference <0 (good response, 47%) compared to those with a VAS difference ≥0 (poor response, 26%) but again this was not significant. To summarise, whilst not significant, we observed a trend of an increased frequency in this particular heterozygous genotype in individuals that had a reduced pain score (i.e. desired analgesic effect/good response) following treatment with tramadol. This observation is purely preliminary, as the small sample size is not sufficient to discern any true biological relevance.

Finally, we searched for associations based on sex as an independent variable. We identified no associations between genotype and sex (female n=108, male n=40, P>0.05). However, we did identify an increased frequency in allele c with males in the osteoporosis cohort (n=13, 33%, P=0.02) (Figure 4.14B). It must be reiterated that any results in this chapter which reached statistical significance should be replicated in a larger cohort. This is necessary to achieve sufficient power, in order to minimise the effects of a type II error (i.e. falsely rejecting the null hypothesis).

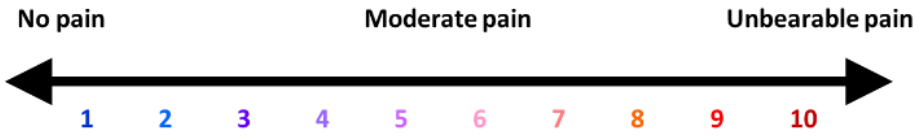


Figure 4.10. Example of Visual Analogue Scale used to calculate VAS difference.

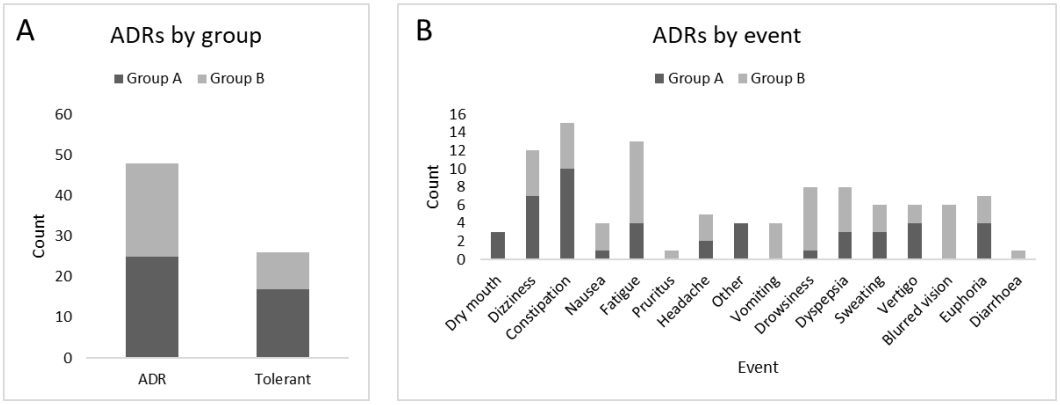
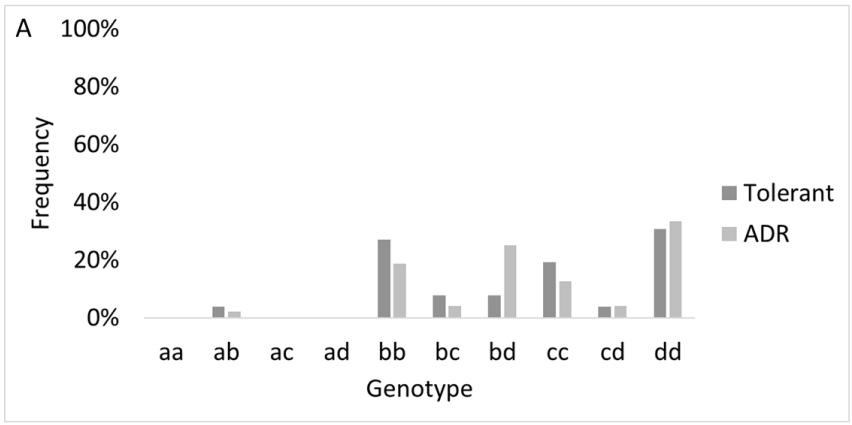
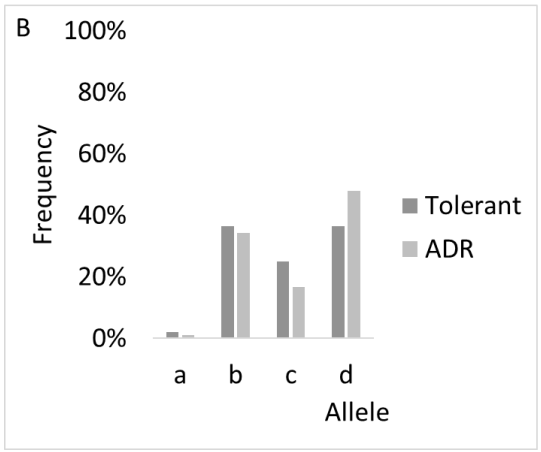


Figure 4.11. Adverse drug reactions (ADR) observed during tramadol study. Combined ADR frequencies from group A (actively taking tramadol) and group B (discontinued use due to prior ADRs). Frequency refers to incidence compared to all reported ADR events.

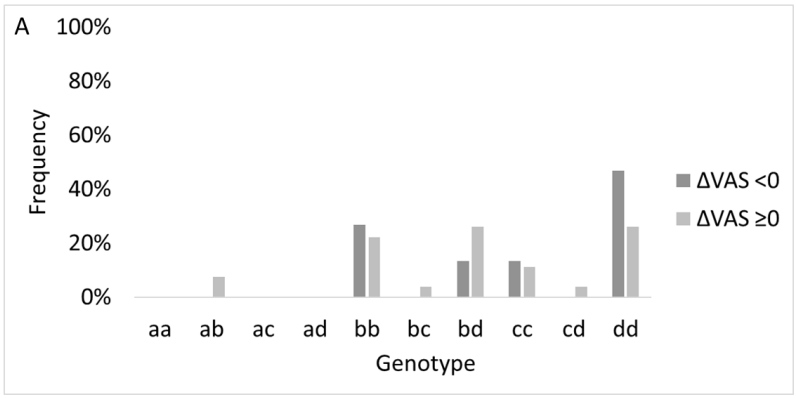


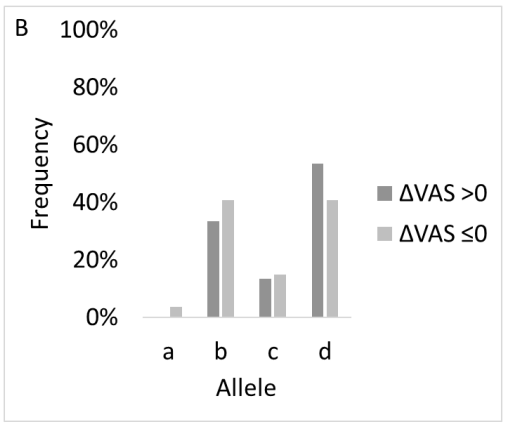


C

	Tolerant		ADR		
Genotype	Count	Frequency	Count	Frequency	P-value
aa	0	0%	0	0%	1.000
ab	1	4%	1	2%	1.000
ac	0	0%	0	0%	1.000
ad	0	0%	0	0%	1.000
bb	7	27%	9	19%	0.555
bc	2	8%	2	4%	0.609
bd	2	8%	12	25%	0.118
cc	5	19%	6	13%	0.502
cd	1	4%	2	4%	1.000
dd	8	31%	16	33%	1.000
Total	26	n/a	48	n/a	
Allele	Count	%	Count	%	P-value
a	1	2%	1	1%	1.000
b	19	37%	33	34%	0.249
c	13	25%	16	17%	1.000
d	19	37%	46	48%	0.225
Total	94	n/a	180	n/a	

Figure 4.12. Genotyping results show no association in response to tramadol in osteoporosis patients. (A) Genotype frequencies in osteoporosis patients that are tolerant to tramadol (n=26) or reported adverse drug response (n=48). (B) Allele frequencies. (C) Summary tables showing exact counts and results from statistical analysis. No statistical significance found (Fisher's exact two-tailed test).

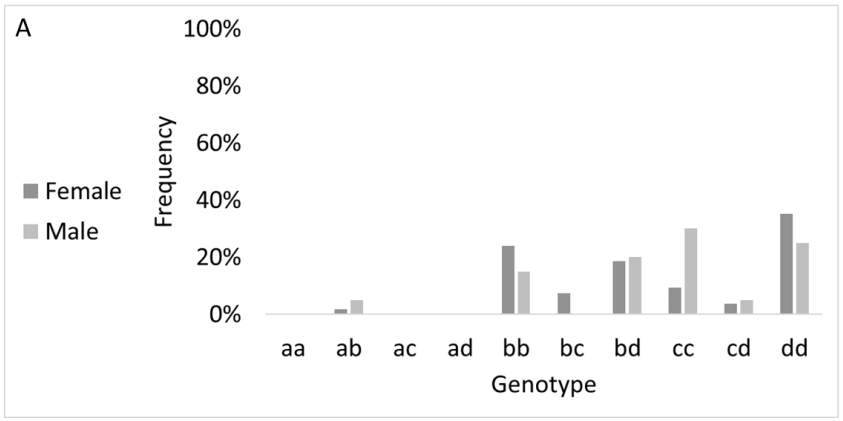


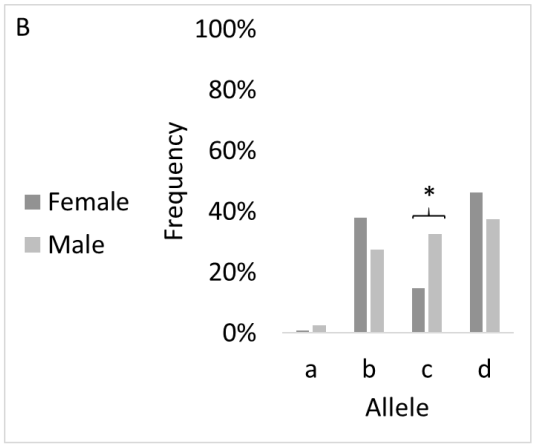


C

	ΔVAS <0		ΔVAS ≥0		
Genotype	Count	Frequency	Count	Frequency	P-value
aa	0	0%	0	0%	1.000
ab	0	0%	2	7%	0.524
ac	0	0%	0	0%	1.000
ad	0	0%	0	0%	1.000
bb	4	27%	6	22%	1.000
bc	0	0%	1	4%	1.000
bd	2	13%	7	26%	0.450
cc	2	13%	3	11%	1.000
cd	0	0%	1	4%	1.000
dd	7	47%	7	26%	0.322
Total	15	n/a	27	n/a	n/a
Allele	Count	Frequency	Count	Frequency	P-value
a	0	0%	2	4%	0.526
b	10	33%	22	41%	1.000
c	4	13%	8	15%	1.000
d	16	53%	22	41%	0.814
Total	30	n/a	54	n/a	n/a

Figure 4.13. Genotyping results show no association with pain scores in response to tramadol in osteoporosis patients. (A) Genotype frequencies in osteoporosis patients that showed an improvement in pain scores in response to tramadol (negative difference in pain score; ΔVAS<0) (n=15) and those that did not show an improvement in pain score (no difference/positive difference in pain score; ΔVAS≥0) (n=27). (B) Allele frequencies. (C) Summary tables showing exact counts and results from statistical analysis. No statistical significance found (Fisher's exact two-tailed test).





C

	Female		Male		
Genotype	Count	Frequency	Count	Frequency	P-value
aa	0	0%	0	0%	1.000
ab	1	2%	1	5%	0.470
ac	0	0%	0	0%	1.000
ad	0	0%	0	0%	1.000
bb	13	24%	3	15%	0.532
bc	4	7%	0	0%	1.000
bd	10	19%	4	20%	1.000
cc	5	9%	6	30%	0.059
cd	2	4%	1	5%	1.000
dd	19	35%	5	25%	0.557
Total	54	n/a	20	n/a	n/a
Allele	Count	Frequency	Count	Frequency	P-value
a	1	1%	1	3%	0.4688
b	41	38%	11	28%	0.2527
c	16	15%	13	33%	0.0208
d	50	46%	15	38%	0.358
Total	108	n/a	40	n/a	n/a

Figure 4.14. Genotyping results show association between SVA allele in males with osteoporosis. (A) Genotype frequencies in females (n=54) and males (n=20) with osteoporosis. (B) Allele frequencies. (C) Summary tables showing exact counts and results from statistical analysis. Significant p-values indicated as <0.05*, <0.01**, <0.001*** (Fisher's exact two-tailed test).

4.3.1 Characterising epigenetic status of the SVA at TRPV1 and TRPV3

4.3.1.1 G4 prediction is variable between SVA CT-rich alleles

G-quadruplexes are alternate secondary DNA structures that form from multiple runs of G-rich sequences[264]. Prediction tools have identified that they are enriched in the genome at *cis*-regulatory elements and they are predicted to play important roles in transcriptional and translational regulation at the DNA and RNA level[265]. In chapter 3, we gave evidence to support the role of the SVA at *TRPV1* and *TRPV3* as a cis-regulatory element, therefore we aimed to predict its potential to form G4 structures, and to further assess if the sequence length polymorphism identified in this chapter may potentially affect G4 formation.

Bragg et al[54] previously reported that sequence length polymorphism in the CT-rich domain results do not affect G4 formation. This analysis was conducted by DNA sequences generated *in silico* which contained variable numbers of hexamer repeats – however this data was inferred from PCR product size only and not DNA sequencing. They reported that an increasing number of TCTCCC hexamer repeats resulted in maintained G4 potential on the complimentary DNA strand containing AGAGGG hexamer repeats. We demonstrated in the previous section that a source of polymorphism within this composite sequence is from a 17mer repeat, not simply hexamer repeats alone (Figure 4.5). Therefore, we aimed to assess if the increase in size of the CT alleles due to the 17mer, maintained predicted G4 formation in the same manner as simple hexamer repeat variation.

We first assessed G4 prediction across the full length SVA (reference SVA listed in hg19, allele b). The programme we used, QGRS mapper (available at http://bioinformatics.ramapo.edu/QGRS/index.php), identifies Quadruplex forming G-Rich Sequences (QGRS) and assigns each QGRS a score from 0-100 indicating potential G4 formation. We identified that the SVA sequence itself has few QGRS and relatively low G-scores (Figure 4.15A). We then used the complimentary sequence to the SVA and identified many more QGRS with, indicating there is more G4 potential on the complimentary strand

of DNA (Figure 4.15B). The greatest potential for G4 corresponded was identified in the complimentary sequence corresponding to the CT-rich domain (AGAGGG repeats) (Figure 4.15B).

To assess difference between the alleles, we used the DNA sequences previously obtained for allele a, b and c. We did not include allele d as we could not simply infer the variation based on size of the PCR product alone, as variation is supplied by both hexamer and 17mer repeats. Based on findings from the full length SVA, we used the complimentary strand sequences only (AGAGGG) as no G-scores were identified for actual CT sequence (TCTCCC). The complimentary allele a sequence contained 2 QGRS, starting at position 4 (length 21 nt, G-score: 72) and position 29 (length 38 nt, G-score: 59) (Figure 4.15C). The complimentary allele b sequence contained 3 QGRS, starting at position 10 (length 21 nt, G-score: 72), position 34 (length 26 nt, G-score: 68) and position 68 (length 27 nt, G-score: 34) (Figure 4.15D). The complimentary allele c sequence contained 3 QGRS, starting at position 4 (length 21 nt, G-score:72), 33 (length 26 nt, G-score:68) and 67 (length 26 nt, G-score: 67) (Figure 4.15E).

All three alleles contain the same QGRS at the 5’ end, due to shared sequence similarity, supplied by the TCTCCC hexamer repeats (Figure 4.5). Towards the 3’ end of the CT-rich sequence, differences emerge between alleles, mainly from the 17mer repeat described in the previous section. Allele b is predicted to contain an additional QGRS towards the centre of the domain, compared to allele a. Allele c is predicted to contain the same number of QGRS as allele b however the score of the third QGRS in allele c is much greater than that in allele b, suggesting the likelihood of quadruplex formation is much greater.

We concluded G4 potential is still present across the length of the CT-rich domain but is subject to breaks and weaker G-scores at the region corresponding to the 17mer

repeats. This indicates G4 potential is not maintained in the same manner across the 17mer repeats, compared to simple hexamer repeats, and this is further affected by polymorphism.

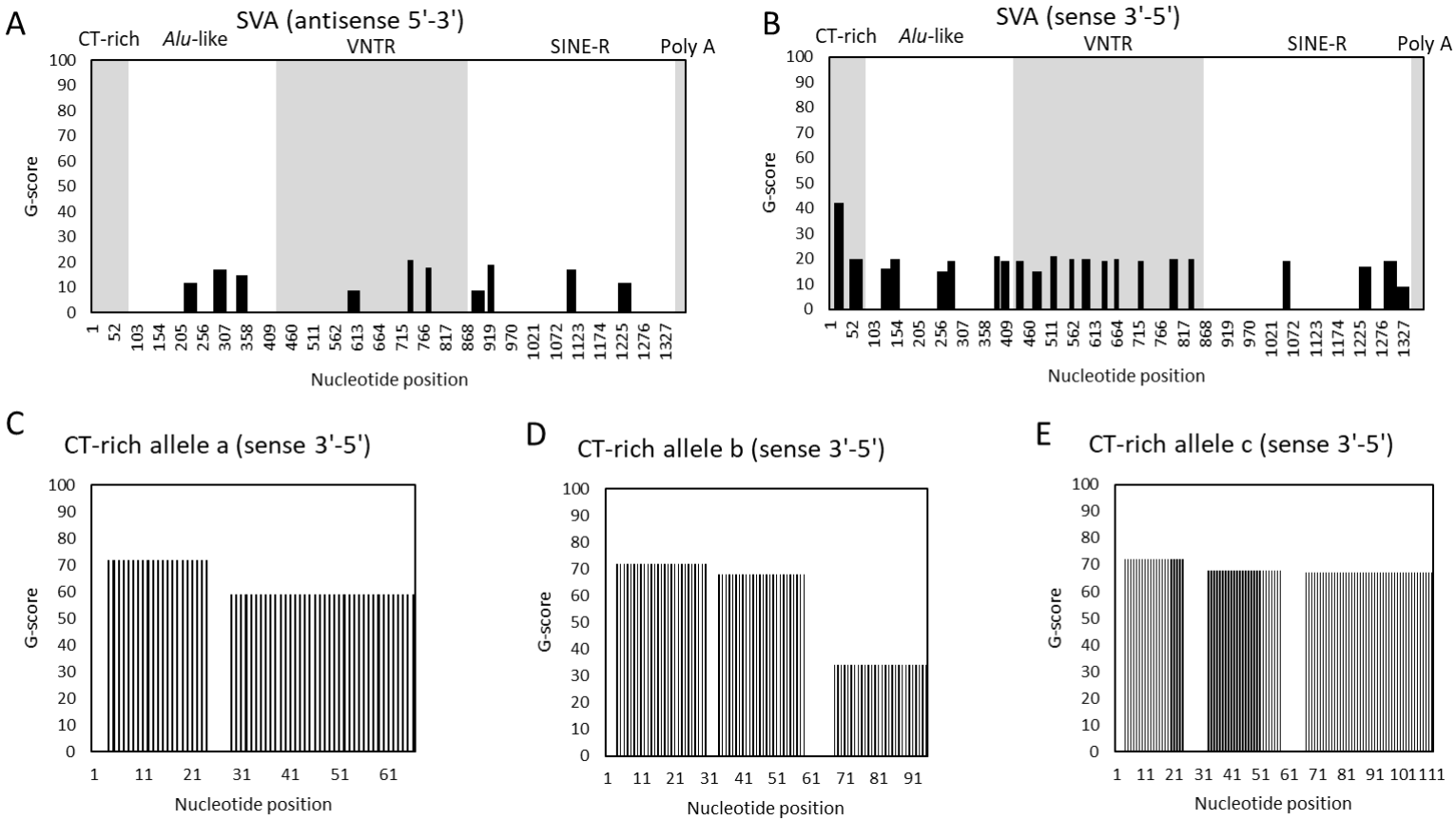


Figure 4.15. Predicted G4 formation across full length SVA and in different CT-rich alleles. QGRS mapper predicted scores (0-100) indicate likelihood of G4 formation. (A) SVA sequence on antisense strand 5'-3' shows weak G-scores <30 across entire length. (B) Complimentary sequence to SVA on sense strand shows increased G-scores and strongest signal generated from region corresponding to CT-rich region. Composite sequence boundaries are annotated. (C) Complimentary strand to CT-rich allele a. (D) Complementary strand to CT-rich allele b. (E) Complementary strand to CT-rich allele c.

4.3.1.2 Analysis of CpG islands at intergenic region between *TRPV1* and *TRPV3*

We first identified putative CpG islands across the intergenic sequence between *TRPV1* and *TRPV3* (chr17:3461250-3468806, hg19) using Cpgplot software (https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/). This software calculates the ratio of observed to expected CG dinucleotides within a sequence. Based on a 100 bp sliding window approach, the algorithm then searches for sequences at least 200 bp in length with a GC content of at least 50%. The algorithm identifies a CpG island if these criteria are met across a minimum of 10 windows. The antisense sequence was used as it encodes the SVA and both genes. We identified two putative CpG islands across the entire intergenic sequence (Figure 4.16C). The SVA was at position 433 – 1836 and the GC content was 59% overall (A, 262; T, 308; G, 375; C, 457), supplying the greatest GC content to this intergenic region (Figure 4.16B). The first CpG island was 208 nt in length and was located at position 582 – 789 and this corresponded to the *Alu*-like region of the SVA (position 150 – 375 in the SVA sequence) (Figure 4.16D). The second CpG island was 227 nt in length and was located at position 4109 – 4335 which corresponded to Alu repeats (Figure 4.16D).

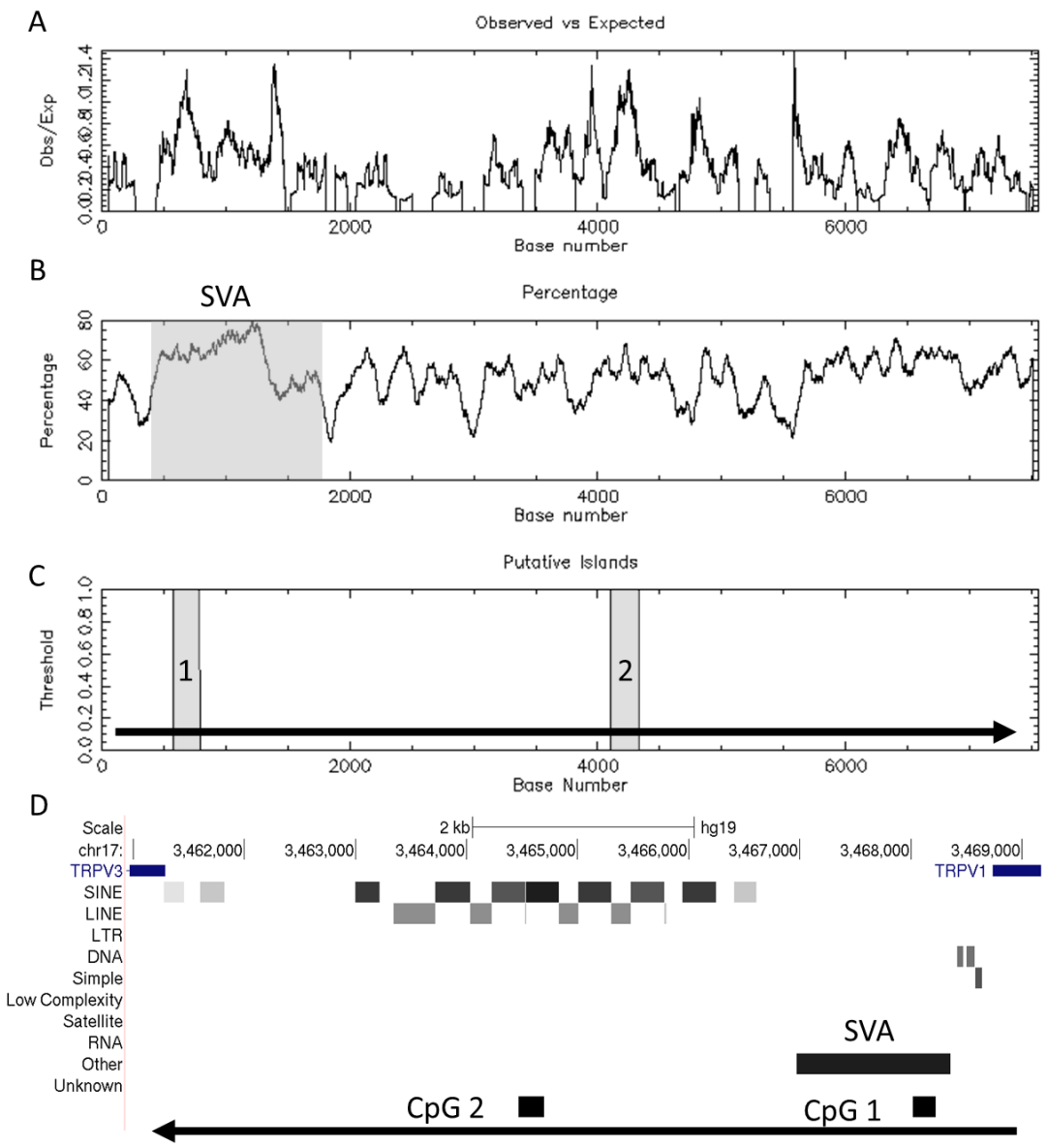


Figure 4.16. Identification of putative CpG islands across intergenic sequence between TRPV1 and TRPV3. (A) Observed vs expected CpG dinucleotides. (B) GC content presented as a percentage. (C) Putative CpG islands. (D) Schematic showing where putative CpG islands reside in the intergenic sequence. 1 corresponds to Alu repeats, 2 corresponds to the Alu-like region at the 5' end of the SVA.

4.3.1.3 Methylation status of SVA is variable between individuals

To assess the methylation status of the SVA, we used an MBD pulldown to enrich for methylated DNA from whole genomic DNA taken from the temporal cortex and whole blood (refer to methods section 2.2.10). Methylated and unmethylated fractions of DNA were then used as templates for PCR to amplify the SVA and compare relative product intensity as an estimate for the amount of methylation present across the amplified region. These tissues were matched in the same individuals (n=12) (Figure 4.17). The genomic DNA samples were from the Dyne-Steele cohort and had previously been genotyped using the sequence length of the 5' CT-rich domain (results section 4.3.2). We amplified the 5' CT-rich domain (approx. 400 – 460 bp) and the 3' poly A region (312 bp) in the methylated and unmethylated DNA fractions using PCR. The VNTR (472 bp) was not included in this analysis as it required a nested PCR of the full length SVA (1151 bp) to be used as a template and ensure specific primer binding but and the DNA had been sonicated to <500 bp.

Amplification of the 5' CT-rich domain (400 – 460 bp) failed when using sonicated DNA as a template however amplification of the 3' poly A region (312 bp) was successful and PCR products were subsequently confirmed for all unmethylated samples in both tissues. It was possible that the sonicated DNA fragments (<500 bp) were too short in this case to amplify the CT-rich target, therefore analysis of the poly A region was used to infer the methylation status of the SVA (Figure 4.18). It was not possible to identify the relative amount of methylated SVA copies using qPCR, as the product size for the poly A region was too large (>200 bp), therefore we used agarose gels and densitometry. The percentage of methylated SVA copies were calculated by comparison with the combined product intensities of methylated and unmethylated DNA fractions and was used for all statistical analysis (two-sample t-test). Sex, genotype and self-reported pain measures were available for these samples, but we only stratified methylation status based on sex due to the low sample size (n=12) (genotypes and pain measures are presented for future reference).

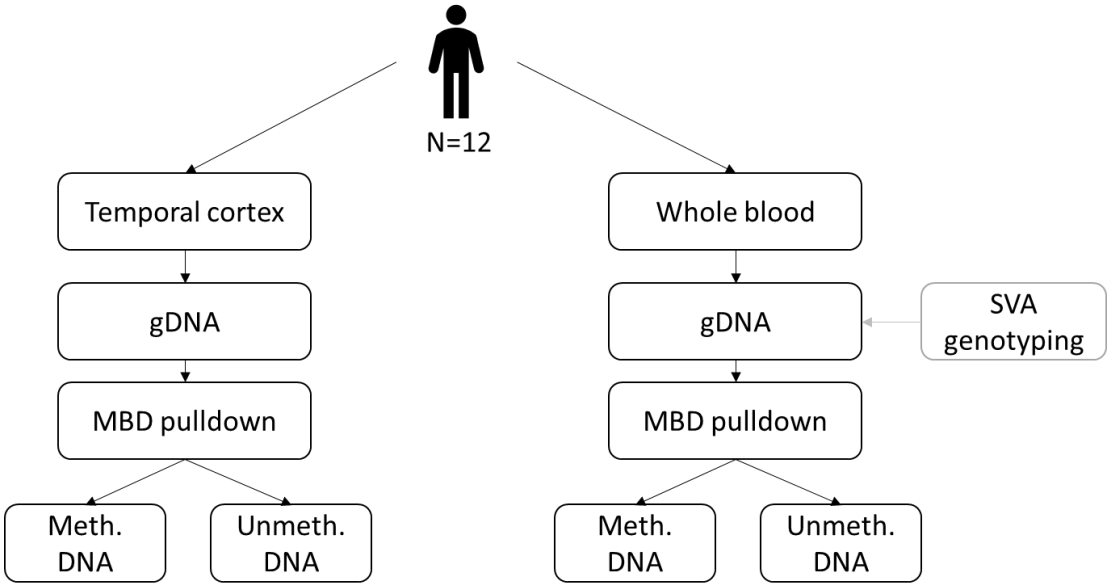
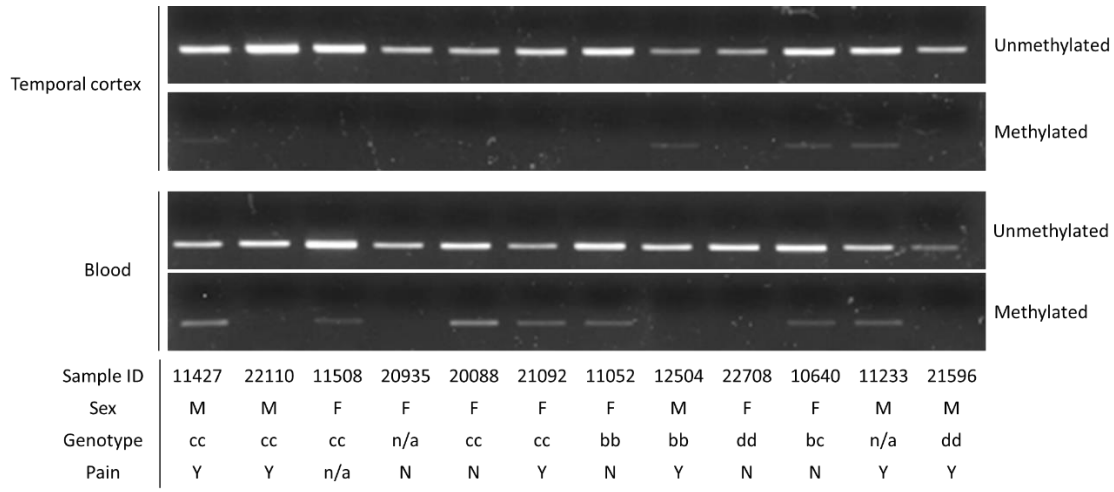


Figure 4.17. Overview of unmethylated and methylated DNA fractions generated from genomic DNA. Temporal cortex and whole blood were taken from each individual. Genomic DNA was prepared and used in the MBD pulldown to isolate unmethylated and methylated fractions (methods section 2.2.10). These samples were used in PCR to assess methylation status of DNA. Some genomic DNA samples from whole blood were used in the genotyping study conducted in section 4.3.2.



Sample ID	11427	22110	11508	20935	20088	21092	11052	12504	22708	10640	11233	21596
Sex	M	M	F	F	F	F	F	M	F	F	M	M
Genotype	cc	cc	cc	n/a	cc	cc	bb	bb	dd	bc	n/a	dd
Pain	Y	Y	n/a	N	N	Y	N	Y	N	N	Y	Y

Figure 4.18. Amplification of SVA 3' poly A region in unmethylated and methylated DNA from temporal cortex and blood. Matched DNA samples were taken from temporal cortex (t. cortex) and blood from the same individuals (n=12). Samples were from the Dyne-Steele cohort; ID numbers specified when possible. Males (M) (n=5) and females (F) (n=7) were used. SVA genotypes based on the length of CT-rich domain are specified. Self-reported pain measures (i.e. presence/absence of pain) are also specified; yes (Y) or no (N).

In temporal cortex, amplification of the SVA was significantly greater from the unmethylated fraction (90%) compared to the methylated fraction (10%) across all samples (n=12) (P<0.001) (Figure 4.19A). When stratified based on sex, this trend was still observed in males (unmethylated=76%, methylated=24%, n=5, P=0.001) (Figure 4.19A). No amplification was detected from the methylated fraction in female temporal cortex (n=7) (Figure 4.19A). As sample numbers were low, we were reluctant to infer any conclusions from this sex difference. Overall this data suggests that copies of this SVA are more frequently unmethylated than methylated in the temporal cortex, however it is undetermined if there are any sex-specific differences regarding the level of methylated copies.

In whole blood, across all samples (n=12), amplification was significantly greater from the unmethylated fraction (76%) compared to the methylated fraction (24%) across all samples (P<0.001) (Figure 4.19B). When stratified based on sex, this trend was still observed in males (unmethylated=71%, methylated=29%, n=4, P=0.046) and females (unmethylated=80%, methylated=20%, n=4, P<0.001) (Figure 4.19B). Three females and two males did not amplify any product from the methylated fraction. Similarly, like temporal cortex, this data suggests this SVA is more frequently unmethylated in whole blood in both sexes.

We then compared the levels of amplification from methylated fractions between temporal cortex and whole blood to assess for tissue specific differences. Across all samples (n=12), we observed a trend that showed an increase in methylation in the blood compared to the brain (temporal cortex=10%, whole blood=23%, P=0.10) (Figure 4.19C). However when we stratified this data based on sex, this trend disappeared in males, there was also no significant difference (temporal cortex=24%, whole blood=76%, P=0.726) (Figure 4.19C). As previously stated, in females, amplification was only detected in the methylated fraction from whole blood and none from temporal cortex. This data suggested that the frequency of

methylated SVA copies in temporal cortex and blood are similar in males, but differences may occur in females, however this requires further investigation. Whilst trends were evident, this data also indicated that methylation status was variable between individuals, as most individuals lacked methylation in both tissues (n=6, 50%), some had methylation in both tissues (n=3, 25%), and others had methylation in one tissue but not in the other (n=3, 25%) (Figure 4.19D and E). Absence of methylation was more frequent overall in the temporal cortex than blood (temporal cortex n=8, whole blood n=5).

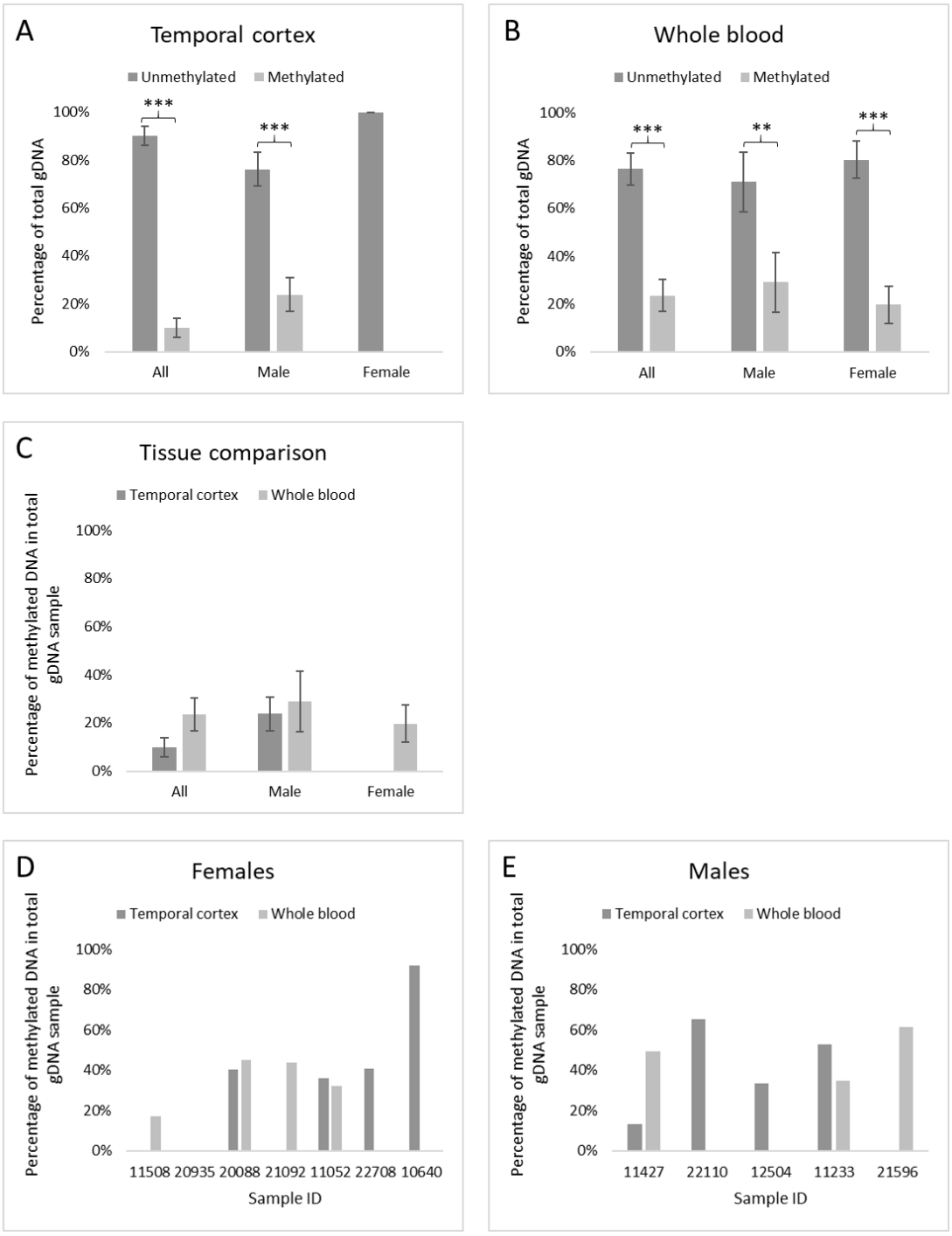


Figure 4.19. Methylation status of the SVA 3' poly A region. (Above) Percentage of amplification from methylated and unmethylated DNA fractions across all samples (n=12), males (n=5) and females (n=7) in temporal cortex (A) and whole blood (B). (Central) Percentage of amplification from methylated fraction compared between temporal cortex and whole blood across all samples, males and females (C). (Below) Percentage of amplification from methylated fraction of each individual; females (D) and males (E). Significant p-values indicated as <0.05*, <0.01**, <0.001*** (two-sample t-test).

4.4 Discussion

In this chapter, we have identified sequence length polymorphism in the SVA encoded at the intergenic region between *TRPV1* and *TRPV3.* The polymorphic nature of the SVA is variable in sequence length, which can at least in part be attributed to repeat copy number differences within the CT-rich domain. Previous studies have referred to this domain as hexameric[53,54,215]. We identified using sequencing that the CT-rich domain is polymorphic based on the copy number of 2 distinct repeats; the first a TCTCCC hexamer repeat, the second a 17mer repeat (C/GATGGTCTCCCTCTCCC) (Figure 4.5). This is the first description of the CT-rich domain within SVAs to contain a 17mer repeat. Our sequencing data also verified the PCR and capillary electrophoresis analyses used to genotype the CT-rich domain was robust and accurate at identifying the four distinct alleles (variable by sequence length) observed across the various cohorts. The frequency of all four alleles were >1% therefore can be classed as common genetic variants and this was the basis for focusing our study on this domain. We did not find any evidence to suggest the VNTR or poly A region were common genetic variants, however we cannot exclude the possibility that rare genetic variants (frequency <1%) may exist. This is also true for the CT-rich domain.

Further analysis identified the 17mer repeat as being enriched in human specific SVA subclass D, when compared to the other SVA subclasses. However this repeat was not found to be human specific as some SVAs containing this repeat were also identified in the chimpanzee and gorilla genome (Figure 4.6). Surprisingly we observed no 17mer repeats in subclass E. During this analysis, we used supplementary data from Tang *et al*[15], who published an analysis of human specific mobile elements (hg38) based on conservation with other primates. SVA subclasses A, B, C and D are present in the human genome and in other primate genomes, however subclasses E and F emerged in the human lineage after divergence with the last common ancestor shared with the chimpanzee, therefore E and F are human specific[25]. We observed in the supplementary data that only 26.2% of SVA E and 50.9% of

SVA F were listed as human specific. Therefore, according to this dataset, 73.8% of SVA E and 49.1% of SVA F were identified in other primate genomes. As SVA E and SVA F subclasses are human specific elements, this data was incorrect and highlighted a large error rate in the current genome build (hg38) for SVA annotation. This may explain why we observed no 17mer repeats in subclass E, which had the greatest annotation error rate based on annotated conservation with other primates. Communication with the author informed us that full length SVA E and F are mapped correctly. However, SVA copies (from any subclass) can carry 5' truncations or internal rearrangements, which are a hallmark of L1 mediated mobilisation, and can be incorrectly classified[26,266]. This results in some SVA sequences that are difficult to map being incorrectly classified as subclass E and F. This then results in SVA copies that belong to subclass A, B, C or D incorrectly being identified as SVA E or F. This explains how a large proportion of SVA E and F currently listed in the human genome also appear in other primate genomes. Another interesting observation in the data published by Tang *et al.* was that there are human specific occurrences of SVA A (1.9%), SVA B (1.6%) and SVA C (17%). It is typically thought that new SVA subclasses emerge via mutation to escape suppression by KRAB zinc finger nuclease proteins, therefore older subclasses are silenced and no longer transpose[38]. This data suggests that some have escaped silencing in the genome after divergence with the last common ancestor and created new insertions. Ancient *Alu* element subclasses *AluJ* and *AluS* have been reported to escape suppression in humans and generate polymorphism[267,268], therefore it is plausible to hypothesise some ancient SVAs can generate human specific polymorphisms too. These findings highlight some of the current issues with SVA annotation in the current human genome build (hg38) and demonstrate that further work is needed to accurately classify SVAs.

In chapter 3, we identified the SVA as a transcriptional regulatory domain. Here, we explored its role as a potential marker of relevant pain phenotypes. Previous work has identified an association between polymorphism in the CT-rich domain of an SVA and the

age-of-onset of X-linked dystonia parkinsonism[54]. The most well characterised association of repeat polymorphisms and disease is the CAG-repeat copy number and age of onset in Huntington disease[269]. *TRPV1* mRNA has been shown to increase 2.4-fold in human aged skin (mean 74 vs 26 years), highlighting age associated changes in gene regulation[270,271]. *TRPV3* has not been characterised in the context of ageing. Genotyping the SVA revealed a significant increase in the heterozygous bd genotype associated with an increase of self-reported pain in healthy aged individuals (Figure 4.8).

Next we genotyped the SVA at the *TRPV1* and *TRPV3* locus in a cohort of osteoporosis patients, some of which were being treated with the synthetic opioid tramadol, and many reported adverse drug responses (Figure 4.11). *TRPV3* has not been implicated in osteoporosis or response to tramadol however TRPV1 has been implicated in both the pathogenicity and occurrence of bone pain in osteoporosis using animal models. Therapeutic action of tramadol is mediated by activation of μ-opioid receptors, however it has been identified as an agonist of TRPV1[258], therefore this cohort offered the potential to explore associations between SVA polymorphisms at this locus, occurrence of osteoporosis and ADRs associated with tramadol. The frequency of the heterozygous bd genotype was greater in osteoporosis patients who experienced an adverse drug response to tramadol (Figure 4.12) and reported a lack of analgesic efficacy after taking tramadol (Figure 4.13). We identified a minimum of 4 distinct alleles when screening the SVA for polymorphism, therefore upon reflection, the size of this cohort was very small (n=74) considering the aim was to identify associations between numerous genotypes and complex phenotypes. For example, when testing for associations between ADRs and genotypes, there were 10 possible genotypes, however each group contained less than 16 individuals. This resulted in statistical power that was too low to draw any relevant biological conclusions.

Sex differences are also well characterised in osteoporosis (extensively reviewed here[272]). Interestingly, we observed a significant increase in allele c in males with osteoporosis, but no associations between allelic frequencies and self-reported pain were found in healthy aged individuals from the Dyne-Steele cohort (Figure 4.14). These findings give some evidence to support SVA genotype in association with TRPV1-related phenotypes. Overall, the main limitation of the genotyping studies and methylation study was the low sample sizes available in each cohort. One study estimated that in order to test a single SNP as a marker of a specific trait, the minimum number of cases required was 248 achieve enough statistical power, estimated using the Genetic Power Calculator under various assumptions about genetic models (e.g. allelic, dominant, recessive, co-dominant)[273]. This was only a single estimate and the actual number for each study will depend upon the biological question being asked (e.g. number of variants and genetic model). The SVA investigated in this chapter has a minimum of four distinct alleles, therefore requiring a much larger sample size. It was not possible to estimate the required sample size for this study as calculators such as the Genetic Power Calculator assume only two variants based on SNPs[273]. One strategy to overcome this problem could be to group alleles into risk and non-risk. However, in order to identify what may be classed as a risk allele, more samples would need to be genotyped to provide more evidence for a risk allele (preferably in a separate cohort). Another limitation is the use of self-reported pain measures, which are subjective between individuals. At this stage, it would be a more appropriate study to associate SVA genotype with gene expression as a molecular phenotype instead of complex trait phenotypes such as self-reported pain measures and/or ADRs. To overcome this, work is ongoing to develop tagging SNPs (SNPs in linkage with SVA genotype), which will enable large scale genotyping in larger datasets such as 1000 genomes, for which there is whole genome sequencing data available for over 4000 individuals. This would provide a better insight into genotype and allele frequencies and could be stratified based on population and/or sex.

We also investigated the epigenetic status of the SVA by predicting G4 structure in polymorphic CT-rich variants and assessed the methylation status of the SVA in whole blood and temporal cortex. We showed that G4 formation may vary dependent upon the polymorphism identified in the CT-rich domain. G4 DNA structure is hypothesised to act as an epigenetic regulator of gene expression. DNA methylation can stabilise G4 structures, demonstrated in the *C9orf72* repeat expansion, which may also modulate epigenetic mechanisms[274,275]. We demonstrated that the SVA was more frequently unmethylated in both blood and temporal cortex, suggesting it may be active and unsilenced (Figure 4.19A and B). We identified a larger proportion of methylation in the blood however this was not significant (Figure 4.19C). Increased expression of TRPV1 mRNA and protein occurs in the temporal cortex of patients with epilepsy, implying a pathogenic role[276]. This disease model would be ideal to investigate SVA methylation and *TRPV1* mRNA expression, to explore the role of the SVA as a tissue specific epigenetic modulator of gene expression. We then demonstrated that the SVA methylation status was variable between individuals. However, there were not enough participants to correlate this with the SVA genotype (Figure 4.19D and E). This methylation analysis had some limitations. The blood samples were taken when the participants were alive and temporal cortex were sampled at post-mortem. Therefore the difference in methylation status observed may be due to the difference in the age of the individual at the time of blood sampling versus the age of the individual at time of death and thus not necessarily reflect tissue-specific methylation patterns[277]. Data relating to age of participants at time of blood sampling versus age at death was not available at the time of study. Post-mortem interval (PMI) can affect DNA integrity and methylation stability therefore results should be interpreted with caution[278,279]. Data relating to PMI was not available at time of study. Sample size was also very small (n=12) therefore we could not stratify genotype or phenotype data with methylation status. Comparison of alleles using bisulphite sequencing and DNA from whole blood taken from homozygous individuals would

provide some insight into whether methylation status is correlated with genotype and would overcome the lack of amplification following sonication of DNA, enabling full profiling of the SVAs methylation status. As we demonstrated in chapter 3, this SVA is functional as a transcriptional regulator (Figure 3.6) and evidence supports its role as a CRE (Figure 3.26). Considering these epigenetic factors, it presents a hypothetical mechanism in which DNA sequence variation of the SVA could affect the epigenetic mechanisms that influence gene regulation.

To summarise, this is the first report of SVA polymorphism associated with a common trait (i.e. non-rare disease e.g. XDP). If validated in larger cohorts, this finding could potentially be used as a biomarker to predict efficacy and adverse drug responses of tramadol. More work is required in this area to reduce the administration of ineffective opioids in a personalised manner. Increasing the sample size via the application of tagging SNPs in the cohorts suggested will greatly improve the utility of this information and potentially delineate functional epigenetic mechanisms between genotype, methylation status and gene expression.

Chapter 5.

LINE-1 retrotransposon expression in neuronal models of ageing and inflammation

5.1 Introduction

Neuroinflammation is a key driver of central and peripheral sensitisation which leads to the transition of acute to chronic pain, and can be caused by trauma (nerve injury), drugs treatments[184] (e.g. chemotherapy), autoimmune conditions[179,180,185]. It is also considered to be one of the hallmarks of ageing - the primary risk factor for the development of neurodegenerative disease and associated with chronic pain in humans[1,280]. Studies in rodent models have provided insight into this mechanism and demonstrated age associated changes in DRG morphology and sensory thresholds which contribute to changes in pain phenotype[198,281-284]. Examples include increased DRG innervation following nerve injury and lower sensory thresholds in aged rats, leading to increased mechanical allodynia and thermal hyperalgesia[281-283]. Changes in inflammatory markers (e.g. CCL2, IL6, CD45, TGFβ1, CD68 and IL1β) have been documented in rat DRG in ageing and in response to nerve injury[198,284]. Epigenetic mechanisms are also implicated in the transition of acute to chronic pain. Several studies have demonstrated global DNA methylation changes in response to nerve injury in rat and mouse models[192-194]. To our knowledge a longitudinal study of age-related DNA methylation changes in DRG has not yet been done.

Recent work by De Cecco *et al.* demonstrated that L1 mRNA expression increased in non-pathological ageing, which resulted in L1 cDNA accumulation in the cytoplasm and induction of an inflammatory response via cGAS and IFN[85,86]. Age associated L1 activity associated with age is due to loss of silencing mechanisms (e.g. DNA methylation) and is implicated in many types of cancer[62-71] and central nervous system neurological disorders[89] (discussed in introduction section 1.2.7). This data highlights TEs, and L1 activity, as a potential contributor in age associated neuroinflammation present in DRG, and thus may play a role in the development of chronic pain. However, there are no reports in the current literature (at time of writing) exploring L1 mRNA expression in the peripheral nervous system (of either rodent or humans) or in the context of pain. TDP-43, an RNA binding protein has

also been heavily implicated in regulating TE activity[90,91] (discussed in detail in introduction section 1.2.7). A large body of evidence suggests that deregulation of TDP-43 expression results in defective mechanisms that suppress L1, yet the majority of research is conducted in regard to ALS (97% of ALS cases characterised by TDP-43 proteinopathy in the form of cytoplasmic inclusions). TDP-43 proteinopathy has been described in dorsal root ganglia from patients presenting with sensory neuropathy and that mutant TDP-43 has been shown to affect sensory function in DRG cultures[285-287]. This area of research is very limited and warrants further investigation, especially regarding the relationship with L1 in the peripheral nervous system.

In addition to age related changes in neuroinflammation and its role in the development of chronic pain, inflammation can be modulated by diet. Studies in rodent models have demonstrated that consumption of a high fat diet causes an increase in cytokines and levels of brain inflammation[182]. A recent study showed that that obesity-associated neuroinflammation is exacerbated in the hippocampus with ageing, demonstrating the synergistic effects between these two variables[183]. Obesity and pain are commonly comorbid conditions in humans, demonstrating the impact of diet on pain prevalence[280,288,289]. Diet is therefore another factor to be considered in TE deregulation in the nervous system[290]. A group led by Prof. Margarett Harnett at the University of Strathclyde investigate effects of high fat diet (HFD) induced inflammation in mice. In addition, this HFD induced inflammation model was also being trialled for a novel anti-inflammatory treatment called ES-62. ES-62 is a secreted product from the filarial nematode *Acanthocheilonema viteae*. Previous work on ES-62 drug analogues 11a and 12b (referred to still as ES-62) has identified successful immunomodulatory properties in models of arthritis[291], asthma[292] and systemic lupus erythematosus[293]. ES-62 binds directly to the Toll/interleukin-1 receptor homology domain (TIR) of Myd88, a central signalling protein in the immune system, which triggers cascades involving NF-κB and MAPK. MyD88 has also been highlighted as a mediator

of neuroinflammation and pain sensory neurons[294]. This presented a novel opportunity to assess the L1 and TDP-43 mRNA in an obesity-induced model of inflammation and in response to a novel treatment impacting inflammatory pathways relevant to pain.

In this chapter we present a robust qPCR protocol developed to quantify L1 and TDP-43 mRNA expression in the mouse. This was applied first to an ageing DRG model in which we present an increase in L1 mRNA with increasing age in both male and female mice. To further this, we also observed an increase in TDP-43 mRNA associated with increasing age. To our knowledge this is the first study of L1 in the mammalian peripheral nervous system. We also applied this methodology to a high-fat diet induced model of inflammation and measured L1 mRNA expression in response to a combination of variables including age, high fat diet and treatment with ES-62.

5.2 Hypotheses and aims

There were several hypotheses formed within this chapter and are presented as follows:

1. *L1* mRNA expression increases in the ageing mouse peripheral nervous system, specifically the dorsal root ganglia.
2. *L1* mRNA increases in the mouse ageing brain and can be modulated in response to diet.
3. To further both points, we hypothesised that TDP-43 expression would play a role in *L1* mRNA regulation in a broader neuroinflammatory context and therefore *TDP-43* mRNA expression could be correlated with *L1* mRNA changes.

We tested these hypotheses by completing the following aims:

1. Develop a robust qPCR protocol to reliably measure L1 mRNA expression in mouse.
2. Measure L1 and TDP-43 mRNA expression in DRG sampled from mice of increasing age.
3. Measure L1 and TDP-43 mRNA expression in brains of mice from a model of HFD induced inflammation:
    a. Measure gene expression in young versus old mice.
    b. Measure gene expression in old mice fed HFD versus normal chow.
    c. Measure gene expression in old mice fed HFD and treated with anti-inflammatory molecule ES-62 versus PBS.

5.3 Results

5.3.1 Design and optimisation of L1 qPCR assay

5.3.1.1 Identification of suitable reference genes for use in mouse models

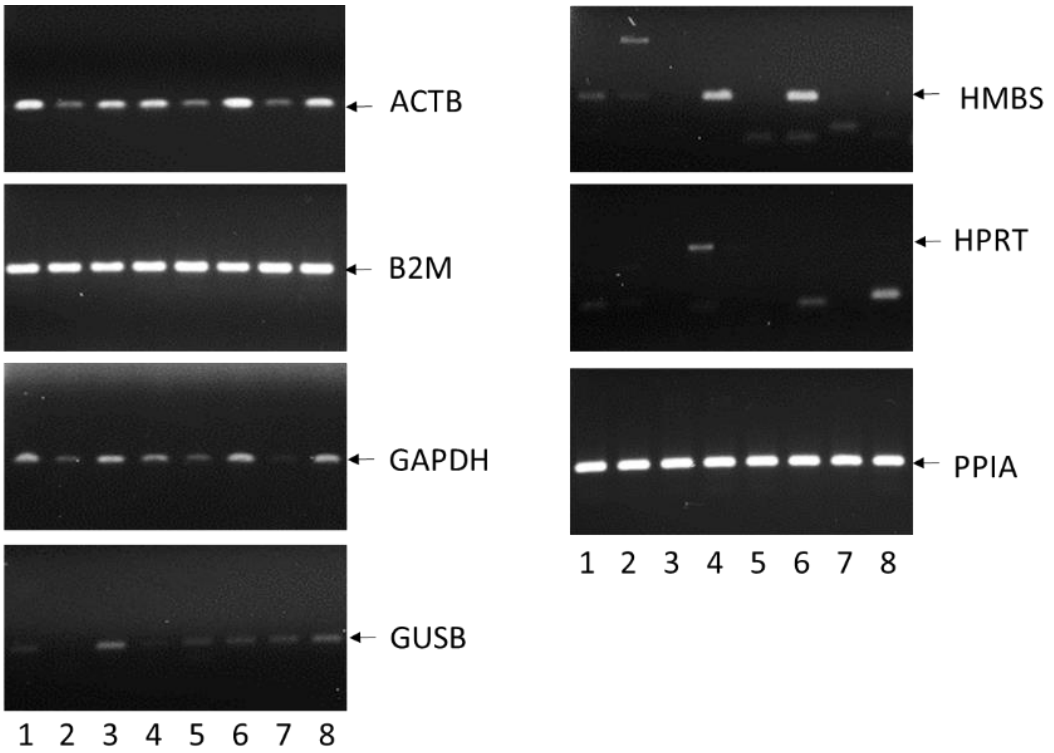
To assess gene expression reliably, it was necessary to identify two reference genes that were stably expressed in tissues across all biological groups within each distinct model, in accordance with MIQE guidelines[295]. A literature search was conducted to identify commonly used reference genes assayed in mouse models that included different sexes, ages and dietary conditions (Table 5.1). Primers were designed for ACTB, B2M, GAPDH, GUSB, HMBS, HPRT and PPIA. Primer sequences are given in appendix A, entry 12. RT-PCR reactions were set up for each primer sets and thermal cycles followed the touchdown PCR protocol described in section 2.2.2.1. Products were analysed using 2% agarose gels.

First, RT-PCR was used to assess expression in the ageing brain C57BL/6 model using a sample from each biological group which differed based on sex (male/female), diet (chow/high fat), age (8 weeks/71 weeks) and treatment (ES-62/PBS). Agarose gel electrophoresis showed that in brain samples, B2M and PPIA had robust and consistent expression across all samples, therefore were selected for use as candidate genes in the brain (Figure 5.1).

We then repeated the same RT-PCR assay in DRG samples taken from Thy1-YFP mice which differed based on sex (male/female) and age (ranging from 3 – 57 weeks). The previous results for GUSB, HMBS and HPRT showed evidence of non-specific products in brain samples which was attributed to poor primer design, therefore were not used in the screen for DRG samples. ACTB, B2M, GAPDH and PPIA were tested and showed that ACTB and PPIA were stably expressed across all ages in both sexes, therefore were used as reference genes in this model (Figure 5.2).

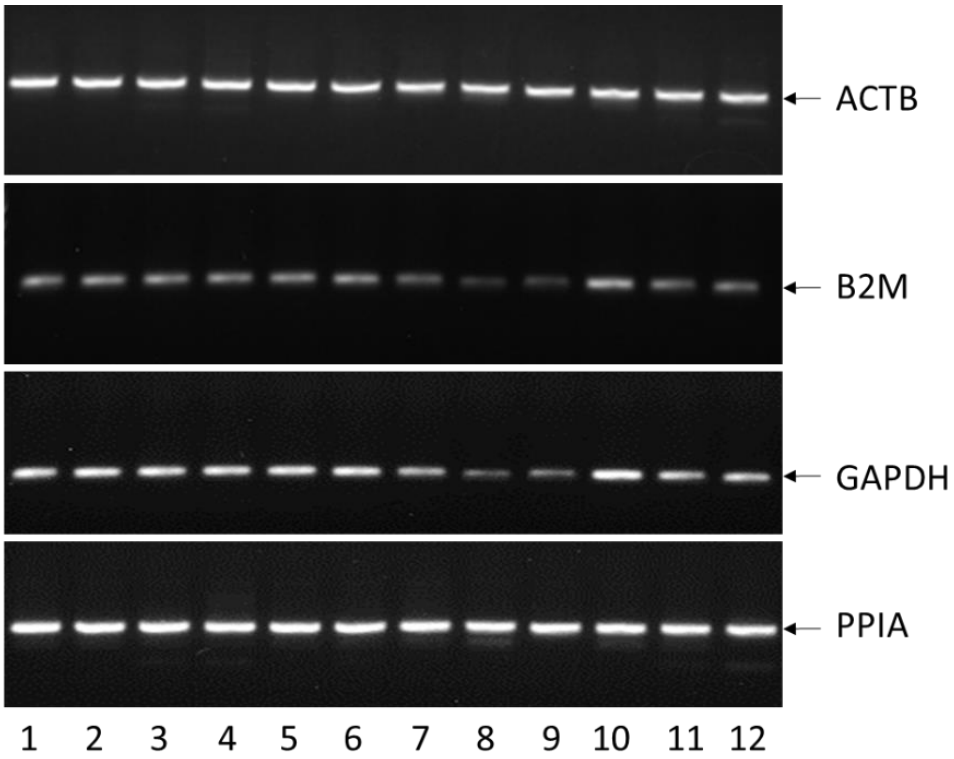
Table 5.1. Candidate reference genes chosen for screen based on literature search.

Variable	Attribute	Gene symbol	Gene name	Accession number	Function	Reference
Tissue	Brain	*PGK1*	Phosphoglycerate kinase 1	NM_008828.3	Carbohydrate metabolism.	[296]
Sex	Male	*SDHA*	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	NM_023281.1	Citric acid cycle and the respiratory chain.	[297]
		ACTB	Beta-actin	NM_007393.5	Formation of microfilaments in eukaryotic cells.	
	Female	*SDHA*	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	NM_023281.1	Citric acid cycle and the respiratory chain.	
		GAPDH	Glyceraldehyde 3-phosphate dehydrogenase	NM_008084.3	Catalyses sixth step of glycolysis.	
Age	2 months	*PPIA*	Peptidylprolyl isomerase A	NM_008907.1	Catalyzes cis-trans isomerization of proline imidic peptide bonds. Role in protein folding.	[298]
		HPRT	Hypoxanthine guanine phosphoribosyl transferase	NM_013556.2	Transferase that plays a role in the generation of purine nucleotide through the purine salvage pathway.	
	22 months	*PPIA*	Peptidylprolyl isomerase A	NM_008907.1	Catalyzes cis-trans isomerization of proline imidic peptide bonds. Role in protein folding.	
		GAPDH	Glyceraldehyde 3-phosphate dehydrogenase	NM_008084.3	Catalyses sixth step of glycolysis.	
Diet	Calorie restriction	*B2M*	Beta-2 microglobulin	NM_009735.3	Beta-chain of MHC class I molecules.	[299]
		PPIA	Peptidylprolyl isomerase A	NM_008907.1	Catalyzes cis-trans isomerization of proline imidic peptide bonds. Role in protein folding.	
		HMBS	Hydroxymethylbilane synthase	NM_013551.2	Heme synthesis and porphyrin metabolism.	



Lane	Sex	Age (weeks)	Diet	Treatment
1	Male	71	High fat	PBS
2	Male	71	High fat	ES-62
3	Male	71	Chow	n/a
4	Male	8	Chow	n/a
5	Female	71	High fat	PBS
6	Female	71	High fat	ES-62
7	Female	71	Chow	n/a
8	Female	8	Chow	n/a

Figure 5.1. B2M and PPIA identified as suitable reference genes in brain samples from C57BL/6 mouse model. Agarose gels showing reference gene products across biological conditions. Lanes 1-8 contain a different biological sample, details of which are given in the corresponding table. Samples differed based on sex (male or female), age (8 weeks or 71 weeks), diet (chow or high fat) and treatment (ES-62 or PBS). B2M and PPIA are robust in both sexes across all ages, diets and treatments. ACTB and GAPDH showed variable expression. GUSB expression was very low therefore not suitable for use. HMBS and HPRT amplified non-specific products therefore not suitable for use.



Lane	Sex	Age (weeks)
1	Male	3
2	Male	10
3	Male	12
4	Male	20
5	Male	35
6	Male	49
7	Female	3
8	Female	10
9	Female	12
10	Female	31
11	Female	42
12	Female	57

Figure 5.2. ACTB and PPIA identified as suitable reference genes in DRG samples from Thy1-YFP mouse model. Agarose gels show reference gene products across biological conditions. Lanes 1-12 contain a different biological sample, details of which are given in the corresponding table. Samples differed based on age (ranging from 3 - 57 weeks) and sex (male or female). ACTB and PPIA show robust expression across all ages in both males (lanes 1-6) and females (lanes 7-12). B2M and GAPDH are variable in females across ages.

5.3.1.2 qPCR primer specificity and amplification efficiencies

Following RT-PCR, primers for reference genes ACTB, B2M, PPIA and target genes L1 and TDP-43 were validated in an initial qPCR assay (Figure 5.3). Melt curves were used to identify at which temperature the PCR product separated into single stranded molecules, resulting in the dissociation of the fluorescent dye from the double stranded molecule. The decrease in fluorescence is measured and represented as a peak corresponding to that temperature, thus the presence of a single peak indicates dissociation at one specific temperature i.e. one specific product. The melt curves showed a clear distinct peak for each primer pair, indicating a single product was amplified. A smaller peak was also identified on the curve for ACTB, which suggested primer dimer or the presence of a secondary structure - which is more difficult to separate. Standard curves were plotted from the mean Ct values generated across a serially diluted cDNA template and were used to assess the efficiency of each primer pair. It is recommended that all primers within each experiment should have efficiencies between 90 – 110% to generate data that is comparable across targets and suitable for gene quantification. ACTB, PPIA and LINE-1 showed good efficiencies; 93%, 101% and 98%, respectively (Figure 5.3). TDP-43 had an efficiency of 80%, this was just below the recommended threshold however was taken forward due to time restraints.

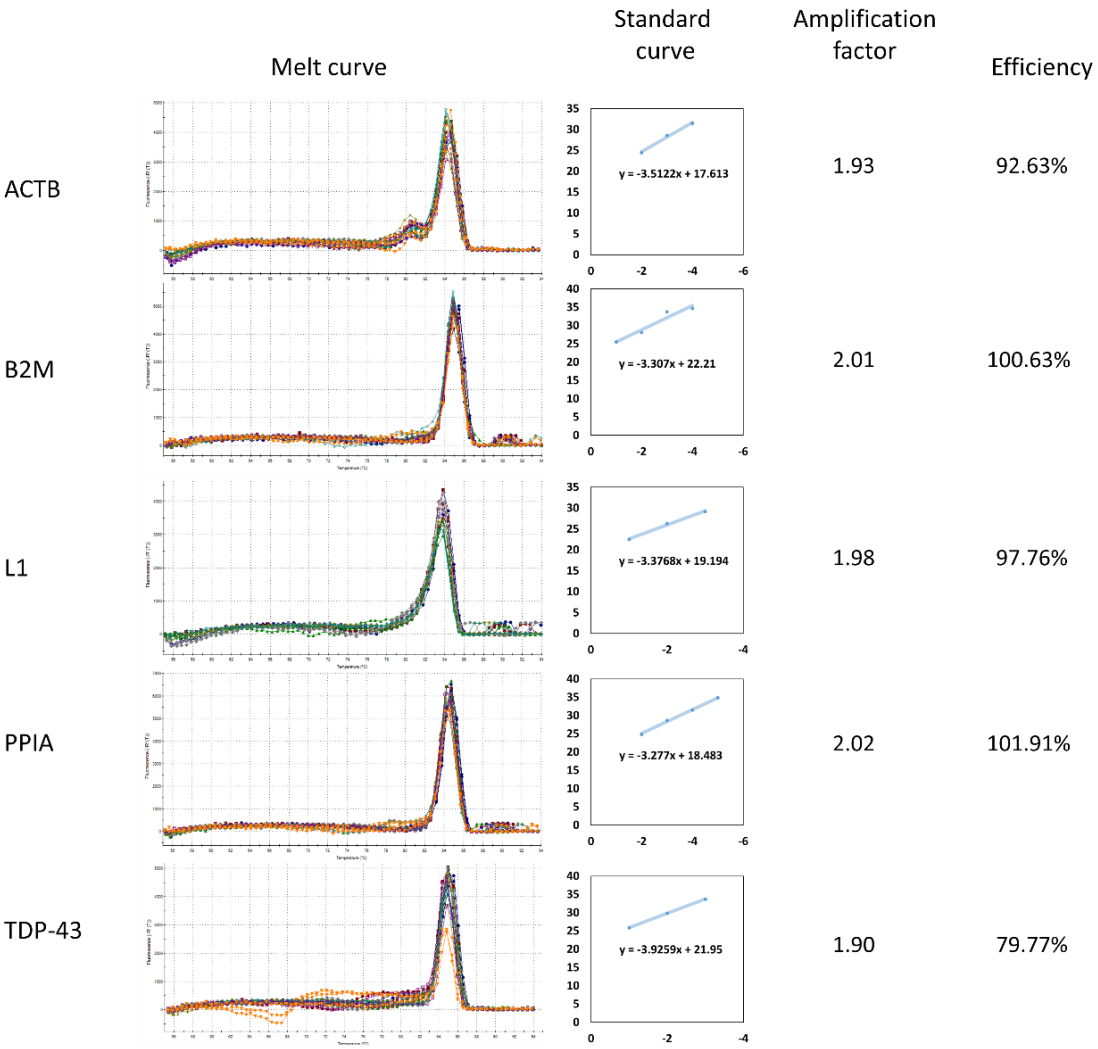
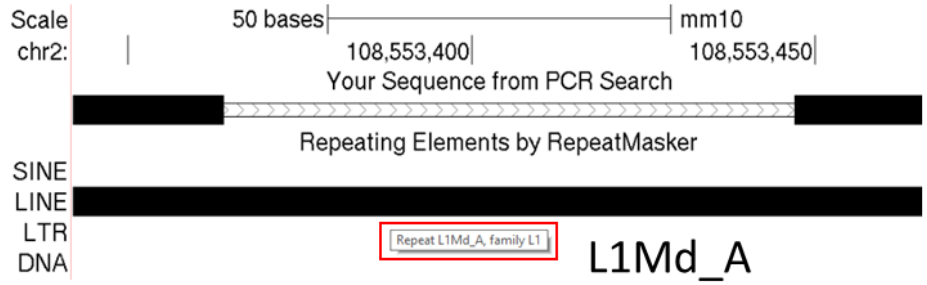


Figure 5.3. qPCR primer validation results. Melt curves are shown for all primer pairs showing single product amplification. Standard curves are shown with dilution factor on X axis and mean Ct value on Y axis. The gradient value from the equations presented on each standard curve were used to calculate amplification factor and efficiency (%).

5.3.1.3 Specificity of L1 primers to subfamilies L1Md_A and L1Md_T

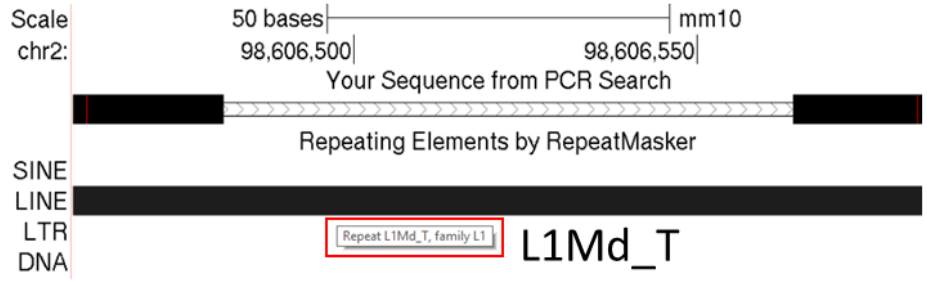
There are many different subclasses of L1 transcripts in the mouse genome that have emerged throughout evolution. The structure of the L1 transcript is well conserved across ORF1 and ORF2 but the 5' UTR can be highly variable. It is estimated that novel promoter sequences have been recruited to the 5' UTR 11 times throughout evolution of the mouse genome, giving rise to multiple L1 subfamilies[300,301]. Most subfamilies of L1 do not mobilise in the genome but the two youngest subfamilies are still actively transcribed and able to mobilise; L1Md_A and L1Md_T[300,301]. Primers were designed to target the 5' UTRs of active retrotransposon subfamilies L1Md_A and L1Md_T (Figure 5.4). These two subfamilies are considered to have non-homologous 5' UTRs overall however homologous sequences were found between the two subfamilies that were suitable for primer design, which generates an amplicon of 124 bp in both subfamilies. Primers and thermal cycles are given in appendix A, entry 12.

A



CGTGATCCTAAGACCTCTAGTGagtggaacacaacttctg
ccaggagtctggttggaacaccagatatctgggtacctgccttgca
agaagagagcttgcctgcaGAGAATACTCTGCCCACTG

B



CGCGATCCTAAGACCTCTAGTGagtggaacacaactcct
accaggagtccggttcaaacaccagatatctgggtaccctccctgc
aagaagagagcttgcctacaGAGAATACTCTGCCCACTC

Figure 5.4. Primers designed to target active L1 subfamilies. In-silico PCR screenshots taken from UCSC genome browser (mm10). Amplicon sequences provided underneath screenshot (primer is capitalised and bold, mismatch is red). (A) In-silico amplification of L1Md_A (e.g. chr2:108,553,343-108,553,466). (B) In-silico amplification of L1Md_T (e.g. chr2:98,606,460-98,606,583).

5.3.1.4 Validation of LINE-1 amplification from cDNA not gDNA

Typical primer design for qPCR involves the primers complimentary to sequences in separate exons to exclude the amplification of larger products containing introns from potential contaminating gDNA (section 2.2.1.2). In the case of L1 mRNA amplification, it was crucial to determine that no gDNA was present in the cDNA sample. This is because L1 contains no introns, therefore any potential amplification from gDNA cannot be distinguished from that amplified from cDNA. The large number of L1 copies in gDNA would also provide a substantial amount of template even in a small amount of contamination, presenting a great potential source of error. To address this, all RNA samples were treated with DNaseI prior to cDNA synthesis. No reverse transcriptase controls and no template controls were included in each qPCR (Figure 5.5). The lack of amplification in the no reverse transcriptase control confirmed the absence of endogenous gDNA in the RNA samples used for cDNA synthesis. The no reverse transcription control was prepared using the RNA sample that yielded the greatest concentration and logically more likely to contain gDNA contamination. The lack of amplification in the no template control also confirmed the absence of exogenous gDNA contamination present in the prepared samples and reagents used. Therefore, we were confident that all amplification measured in this assay is a true reflection of mRNA levels and not due to confounding gDNA contamination.

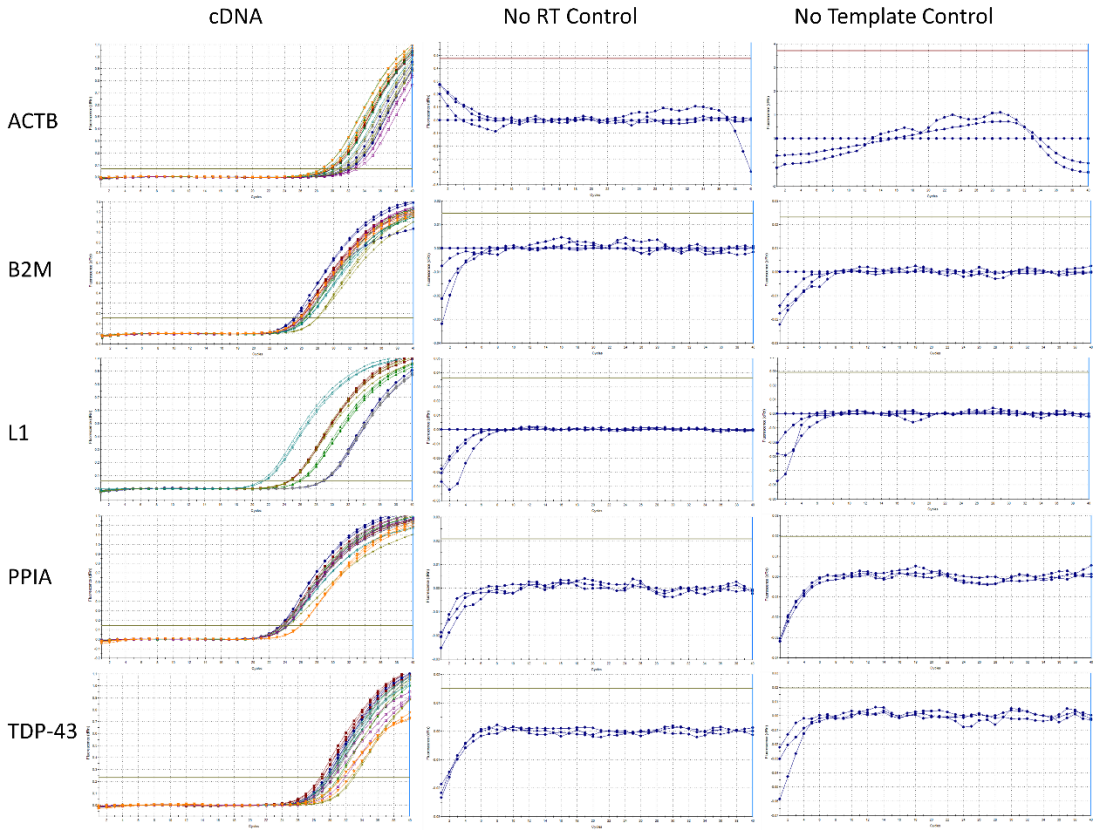


Figure 5.5. Amplification plots from cDNA templates and controls. Amplification from cDNA showing sigmoidal curves. Absence of amplification in no RT (reverse transcriptase) controls and no template controls show cDNA template was prepared properly.

5.3.2 Profiling L1 and TDP-43 expression in DRG from mouse model of ageing

We assessed L1 and TDP-43 mRNA expression in DRG taken from Thy1-YFP mice of increasing age; males ranging from 3 to 49 weeks and females ranging from 3 to 57 weeks. At least three mice of each sex were sampled at each age group. RNA was extracted from DRG samples from individual animals and cDNA was synthesised, keeping biological samples separate (RNA was not pooled). qPCR was performed to assay the expression of target genes L1 and TDP-43. Relative quantification of target genes was normalised against *ACTB* and *PPIA*. Male and female data was analysed separately due to differences in age groups sampled. For instance, older female age groups were 31, 42 and 57 weeks and older male age groups were 20, 35 and 49 weeks. This difference in age groups was due to the universities ethical policy regarding keeping animal usage to a minimum. At the time of study, the animals presented here offered a good range of age groups for both sexes (despite not being identical) and could provide useful biological samples, therefore they were utilised instead of buying and culling additional aged mice. The youngest mice aged 3 weeks were used as the biological control group in both sexes. To assist with data interpretation, the approximate age ranges of mice and their human life equivalencies is given in Table 5.2.

Table 5.2. Life phase equivalencies of human and mice. Equivalencies are representative of C57BL/6 strain. Adapted from https://www.jax.org/news-and-insights/jax-blog/2017/november/when-are-mice-considered-old.

Mouse age range (weeks)	Mouse age range (months)	Human age range (years)	Life phase
13-26	3-6	20-30	Mature adult
43-60	10-14	38-47	Middle aged
78-104	18-24	56-69	Old

5.3.2.1 L1 expression increases in ageing DRG

We measured L1 mRNA expression in ageing DRG (Figure 5.6). We observed a decrease in L1 mRNA in females at 10 weeks that was approaching, but did not reach significance (0.54±0.44, n=3, P=0.08) (Figure 5.6A) . A decrease was also observed in the males at 10 weeks (0.87±0.05, n=3, P=0.59) (Figure 5.6B). Whilst not significant, this decrease was observed in both analyses, perhaps reflecting a true biological effect. We observed no difference in L1 expression in females at 12 weeks (1.12±0.20, n=3, P=0.66) (Figure 5.6A) or in males (1.16±0.23, n=3, P=0.86) (Figure 5.6B) when compared to controls.

However, in both sexes, as age increased, we begun to see a trend of increased expression that was first observed in 'mature adulthood' and was maintained throughout ageing. Older age groups differed between males and females however the same trend was observed. In females age 31 weeks we observed a 16-fold increase in L1 (16.34±4.80, n=3, P=0.08) (Figure 5.6A). A 12-fold increase was also observed in males at 20 weeks (12.52±1.80, n=3, P=0.05) (Figure 5.6B). This increase was maintained in both sexes. Females showed an 8-fold increase at 42 weeks (8.44±1.90, n=3, P=0.08), and a 12-fold increase at 57 weeks (12.36±0.92, n=3, P=0.08) (Figure 5.6B). Males showed a 19-fold increase at 35 weeks (18.81±2.50, n=3, P=0.05) and a 15-fold increase at 49 weeks (15.92±0.05, n=3, P=0.05) (Figure 5.6B).

The overall trend observed here was that L1 expression increased in DRG with age in males and females by several fold. Initially, we were surprised to observe fold changes of such magnitude and suspected it may be due to sample processing errors. Genomic DNA contamination would provide a large source of L1 template and likely result in a huge amount of amplification, and detection at earlier cycles in the qPCR reaction. We did not observe this and ran reverse transcriptase controls during each amplification to confirm absence of gDNA template (Figure 5.5), therefore we were confident that the large fold increases observed in this model are not due to error.

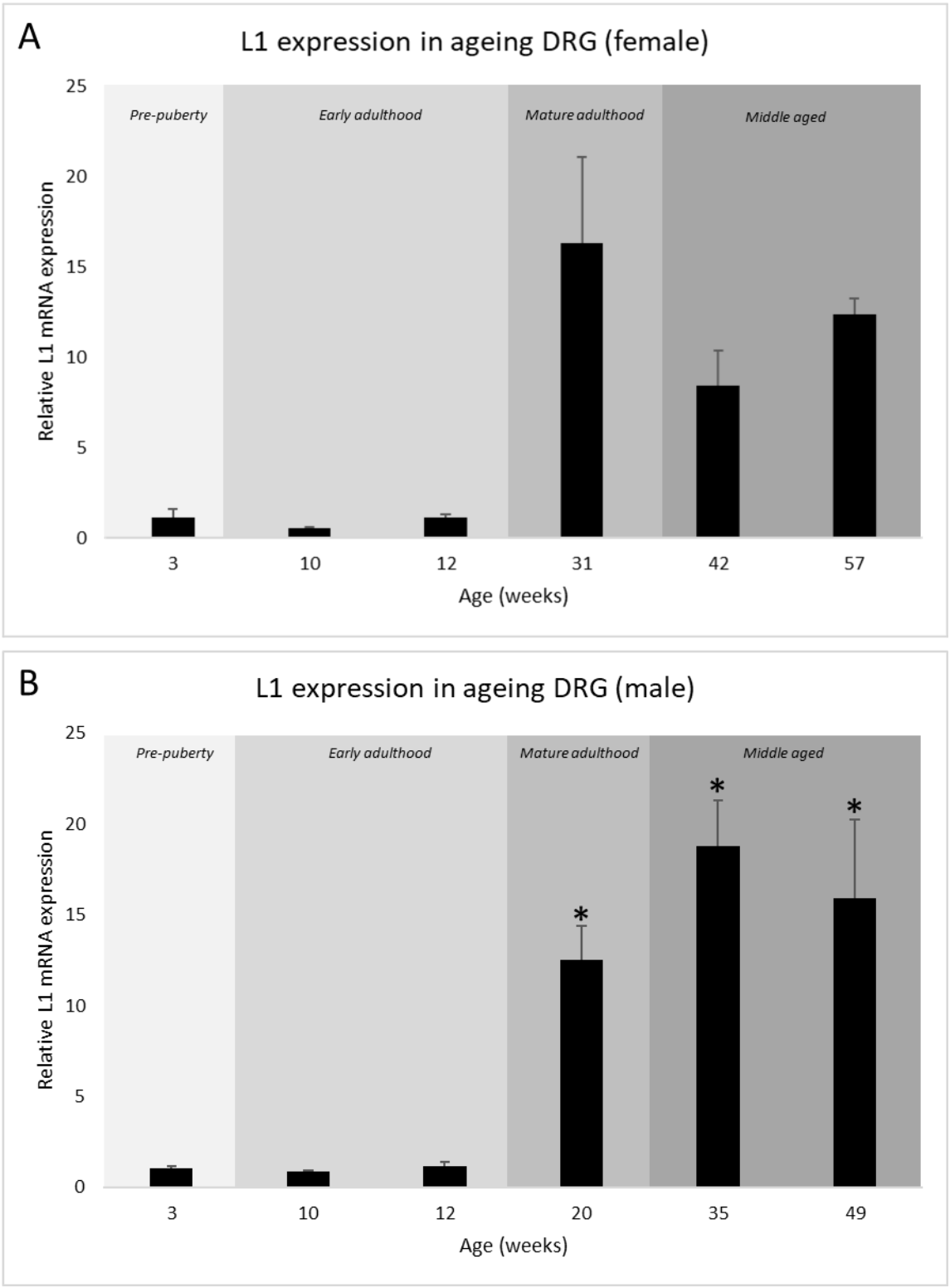


Figure 5.6. L1 mRNA expression increases in ageing DRG. Relative expression of L1 normalised against ACTB and PPIA. Mice at 3 weeks used as biological control. Error bars indicate standard error of the mean. Significance indicated by *P<0.05 (Mann-Whitney U tests).

5.3.2.2 TDP-43 expression increases in ageing DRG

TDP-43 mRNA expression was then measured to identify potential correlations with L1 mRNA expression (Figure 5.7). In females, no difference in relative expression was observed at 10 weeks (1.09±0.21, n=3, P=0.66) or at 12 weeks (0.97±0.14, n=3, P=1.00) (Figure 5.7A). This was also the same in males, with no difference at 10 weeks (0.99±0.05, n=3, P=1.00) or at 12 weeks (1.04±0.07, n=3, P=0.86) (Figure 5.7B).

Similarly to the data trends observed for L1, we began to see an increase in the older age groups. Females showed a 2-fold increase at 31 weeks (2.10±0.23, n=3, P=0.08) (Figure 5.7A). This increase continued at 42 weeks, showing a 3.5-fold increase (3.47±0.72, n=3, P=0.08) and at 57 weeks, showing a 4-fold increase (3.96±1.30, n=3, P=0.08) (Figure 5.7A). The same overall trend was observed in males.

Males at 20 weeks showed a 3-fold increase (2.98±0.77, n=3, P=0.05) (Figure 5.7B). Interestingly, the levels then began to decrease with age yet were still far greater than those at the younger age groups. Males showed a 2.6-fold increase at 35 weeks (2.62±0.11, n=3, P=0.05) and a 2-fold increase at 49 weeks (2.16±0.15, n=3, P=0.05) (Figure 5.7B).

Overall, we consistently observed increased levels in TDP-43 expression in DRG from aged mice compared to young controls. Similarly to L1, this increase was observed during the ‘mature adulthood’ life phase and was consistent throughout ageing.

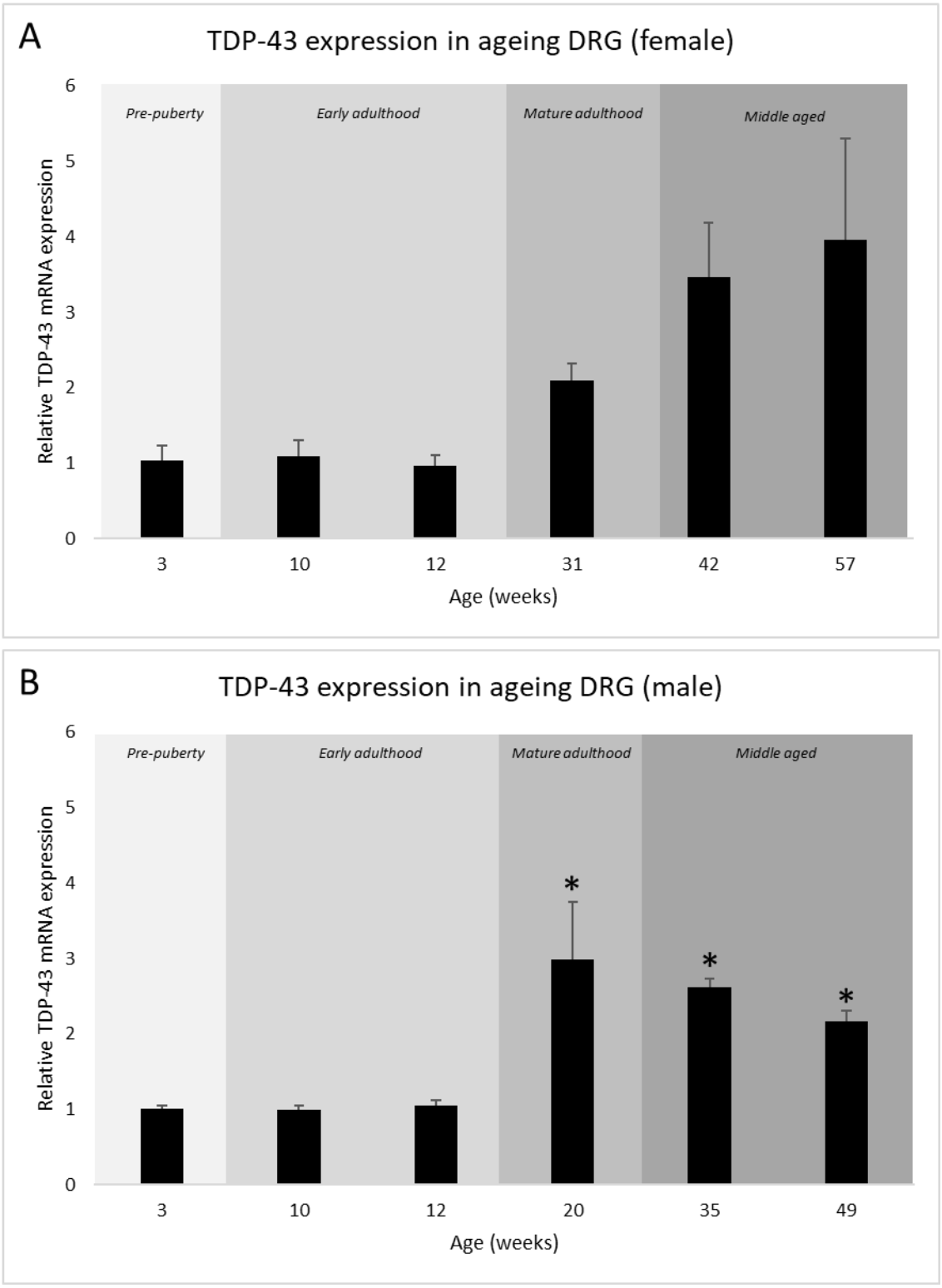


Figure 5.7. TDP-43 expression increases in the ageing DRG. Relative expression of L1 normalised against ACTB and PPIA. Mice at 3 weeks used as biological control. Error bars indicate standard error of the mean. Significance indicated by *P<0.05 (Mann-Whitney U tests).

5.3.3 L1 and TDP-43 expression in brain from mouse model of inflammation

The previous section described L1 and TDP-43 mRNA expression in the DRGs sampled from Thy1-YFP mice of increasing age. This section describes gene expression changes in the C57BL/6 mouse model of the ageing brain, including conditions of dietary induced inflammation and in response to an anti-inflammatory treatment. No DRG were available during this study, which took place after the ageing DRG study in the previous section (therefore comparisons between these models are limited). Please refer to the overview of the biological groups and relevant comparisons made (Figure 5.8). RNA was extracted from the brains of individuals from each group, DNase treated, and cDNA was prepared. qPCR was conducted to measure *L1* and *TDP-43* mRNA, using the same methods developed for the ageing DRG Thy1-YFP model however relative quantification of target genes was normalised against *B2M* and *PPIA*. Male and female data was assessed independently as ES-62 increases median lifespan in males C5BL7/6 but not females[302].

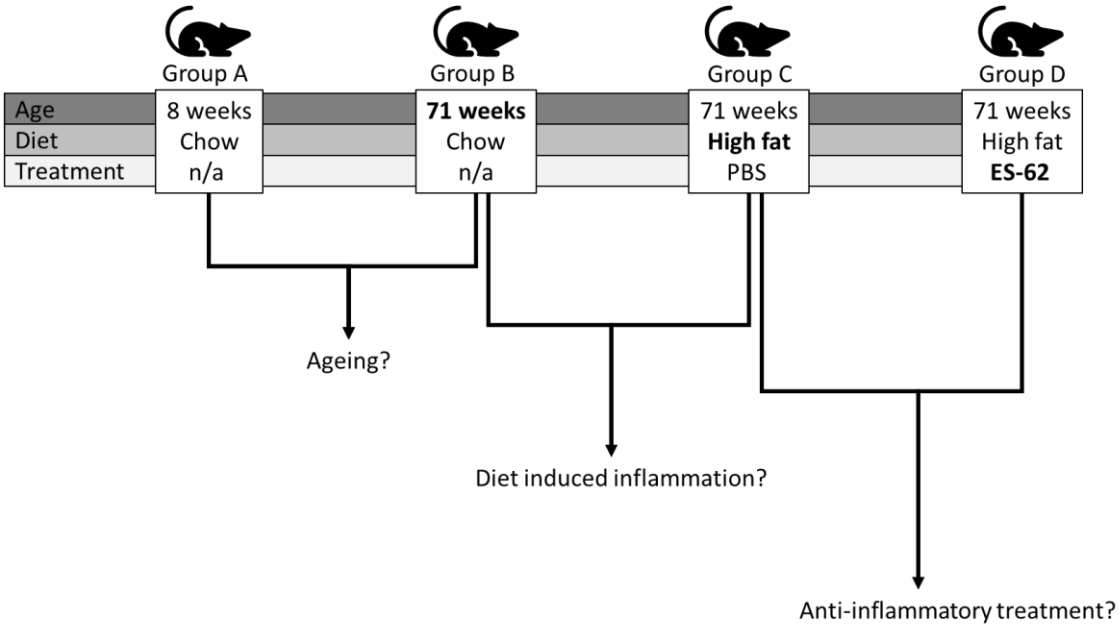


Figure 5.8. Overview of biological groups available in ageing brain C57BL/6 mouse model. 4 groups were available in this model that differed in age, diet and treatment. Comparisons were made between relevant groups to assess if different variables (highlighted in bold) had any effect on *L1* and/or *TDP-43* mRNA expression. Group A and B were suitable compare for effects of ageing. Group B and C were suitable to compare for effect of dietary induced inflammation. Group C and D were suitable to compare for effect of anti-inflammatory ES-62 treatment.

5.3.3.1 L1 expression does not increase in ageing brain

To investigate L1 expression in the ageing brain in this model, brains were harvested from male and female mice at 8 weeks and 71 weeks. All mice were fed a normal chow diet in these groups therefore were assessed for the effect of age only. Unfortunately, female cDNA samples at 71 weeks did not amplify any product during qPCR (for either reference gene or target gene) therefore relative gene expression analysis could not be conducted in this instance. The reason for this failure could not be determined. The male samples were successful, and we observed a 1.1-fold increase in L1 expression in mice at 71 weeks (1.43±0.32, n=5, P=0.52) (Figure 5.9A). Previous studies show L1 expression increases with ageing[84-86], which was also supported by our data from the ageing DRG model showing several fold increase after 12 weeks (Figure 5.6B), therefore we were surprised to not observe a significant increase in L1 mRNA in the brain at 71 weeks. Males at 71 weeks showed a significant 1.4-fold decrease in TDP-43, corresponding to a 30% decrease (0.72±0.09, n=5, P=0.04) (Figure 5.9B).

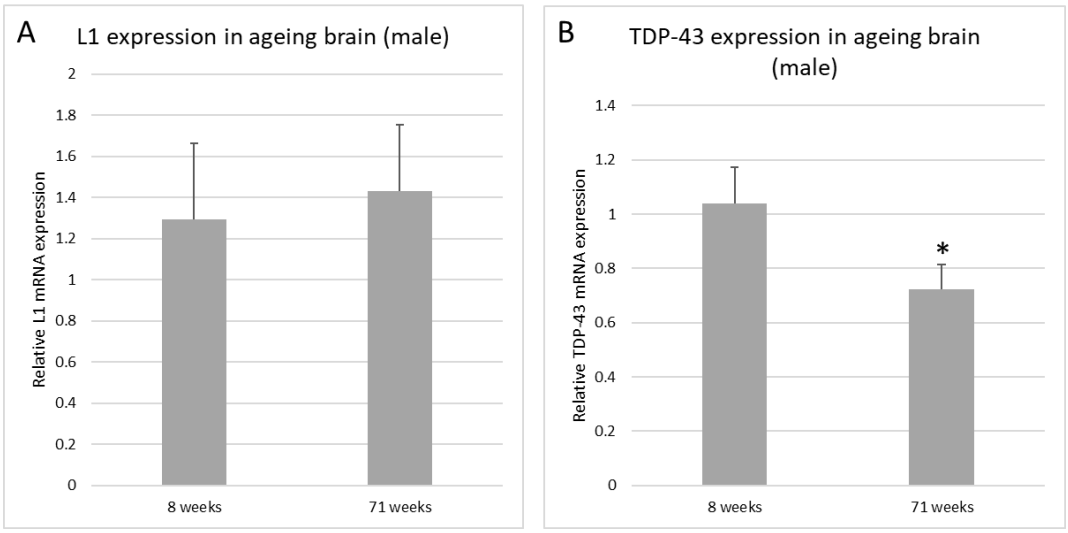


Figure 5.9. L1 and TDP-43 expression in the ageing male brain. (A) L1 expression in brains sampled at 8 weeks and 71 weeks. (B) TDP-43 expression in brains sampled at 8 weeks and 71 weeks. Relative expression of L1 normalised against B2M and PPIA. Error bars indicate standard error of the mean. Significance indicated by *P<0.05 (Mann-Whitney U tests).

5.3.3.2 L1 and TDP-43 expression is unaffected in the ageing brain in response to high fat diet

Previous studies have demonstrated that mice fed a high fat diet have a 5-fold increase in *TDP-43* mRNA in the liver[303]. Calorie restriction reduces L1 mRNA expression approximately 1.6-fold in aged mice skeletal muscle and liver[84]. In addition, L1 hypomethylation in peripheral blood is associated with obesity in humans[304]. Therefore, we were interested to see if the effect of a high fat diet had any impact on *TDP-43* and *L1* mRNA expression in the brain. No data from female mice in the chow fed group was available therefore data analysis could not be conducted. In males, we observed a 1.3-fold increase in L1 expression in response to a high fat diet (1.26±0.24, n=6, P=0.64) (Figure 5.10A). There was no difference in TDP-43 expression in response to a high fat diet (0.98±0.14, n=6, P=0.92) (Figure 5.10B). Our results show no evidence that high fat diet affects *TDP-43* or *L1* expression in this model.

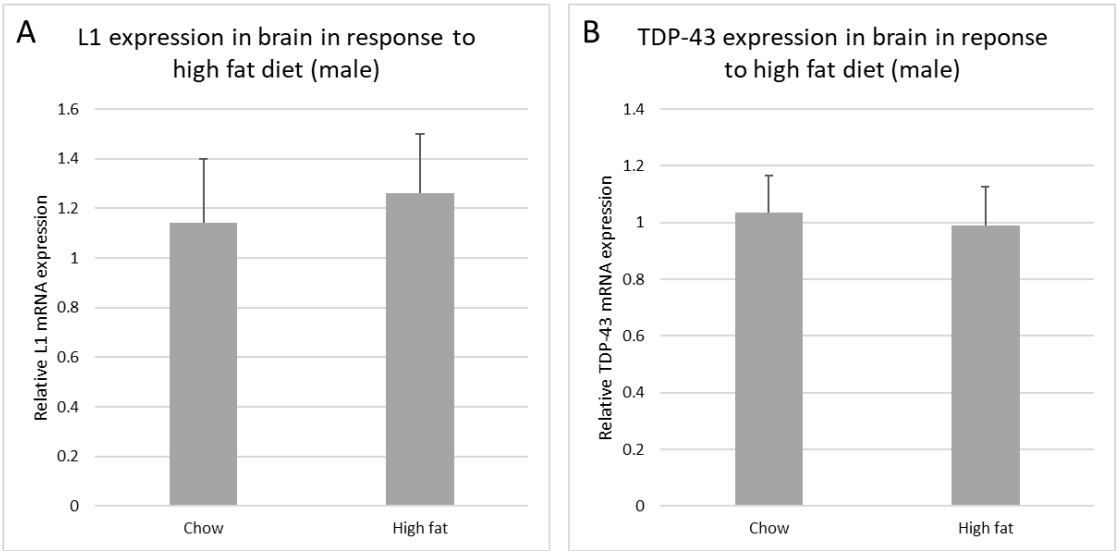


Figure 5.10. L1 and TDP-43 expression in the aged brain and in response to high fat diet. (A) L1 expression in brains of mice fed a chow or high fat diet. (B) TDP-43 expression in brains of mice fed a chow or high fat diet. Relative expression of L1 normalised against B2M and PPIA. Error bars indicate standard error of the mean. No statistical significance identified (Mann-Whitney U tests).

5.3.3.3 L1 and TDP-43 expression increases in response to anti-inflammatory drug ES-62

We assessed if ES-62 treatment had any impact on *L1* and/or *TDP-43* mRNA expression. As the results show in previous section 5.3.3.2, the effect of a high fat diet did not affect *TDP-43* or *L1* expression. However, we proceeded and measured *L1* and *TDP-43* in the aged brains of mice that were fed a high fat diet and then treated with anti-inflammatory drug ES-62 or PBS (refer back to Figure 5.8). Mice injected with PBS were used as the control group.

We observed a 3-fold increase in L1 expression in male mice treated with ES-62 (2.97±1.6, n=5, P=0.31) (Figure 5.11A) and a 3.5-fold increase in female mice treated with ES-62 (3.55±1.1, n=6, P=0.23) (Figure 5.11B) compared to controls. We observed a similar trend when measuring TDP-43 mRNA. In males we observed a 3-fold increase in TDP-43 (3.05±1.8, n=5, P=0.52) (Figure 5.11C) and a 2-fold increase in females (1.84±0.35, n=5, P=0.13) (Figure 5.11D) compared to controls.

None of the results were statistically significant, likely due to the levels of variability in relative expression values. It should be noted that these higher levels of variation were only observed for the ES-62 treated group and not the PBS treated group, nor any of the other biological groups assayed in this model, suggesting variable but biologically significant responses to the ES-62 treatment itself.

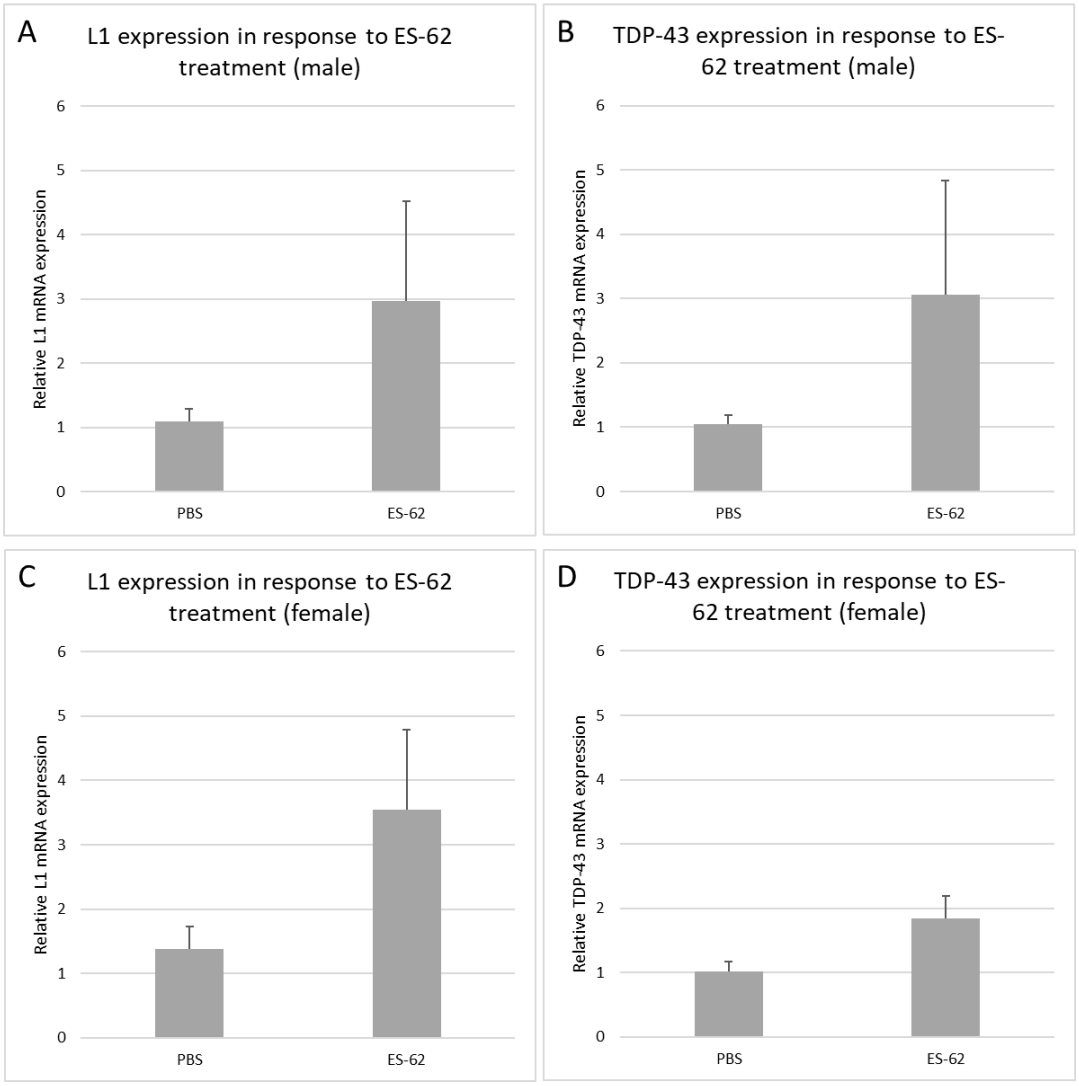


Figure 5.11. L1 and TDP-43 expression in the ageing brain and in response to ES-62 treatment. (A) L1 expression in male brains in response to PBS (control) or ES-62 treatment. (B) TDP-43 expression in brains in male brains in response to PBS (control) or ES-62 treatment. (C) L1 expression in female brains in response to PBS (control) or ES-62 treatment. (D) TDP-43 expression in brains in female brains in response to PBS (control) or ES-62 treatment. Relative expression of L1 normalised against B2M and PPIA. Error bars indicate standard error of the mean. No statistical significance found (Mann-Whitney U tests).

5.4 Discussion

Neuroinflammation in the central and peripheral nervous system is a key driver of neurological disorders and chronic pain[185]. Recent work has identified age-associated accumulation of cytoplasmic L1 cDNA and induction of inflammation[85,86]. However, the majority of the literature regarding L1 in neuroscience has focused on the mechanism of mobilisation and its contribution to neurodegeneration in the central nervous system. Here we aimed to expand on this by developing a qPCR assay to measure L1 mRNA expression in several contexts.

First, we hypothesised that age associated derepression of L1 would occur in the peripheral nervous system. Therefore, we addressed L1 expression in DRG in an ageing mouse model. Secondly, we hypothesised that L1 expression would be affected in response to dietary induced inflammation in the brain, potentially carrying lifestyle implications for the ageing population, therefore we addressed L1 expression in HFD induced inflammation in the brains of mice. Thirdly, we hypothesised that TDP-43 mRNA expression would change with age, and potentially be modulated in response to diet - implicating TDP-43 in a broader context of neuronal dysfunction. Therefore, we also measured TDP-43 mRNA alongside L1 in both DRG and brains of mice.

Here we have successfully developed a robust methodology to measure L1 mRNA expression in the mouse using qPCR. An RNA-seq based approach such as TEtranscripts would offer a complete overview of TE expression however this is not always feasible[305]. We recommended the use of specific reference genes for use in ageing mouse models, based on RT-PCR (Figure 5.1 and Figure 5.2). The primers for this qPCR assay specifically targeted the active murine L1 subfamilies L1Md_T and L1Md_A. The current consensus from human cell line models and human tumours have identified that L1 expression and resulting mobilisation events occur from a few select loci encoding active L1 sequences – referred to as “hot L1s”[306,307]. The profiling of “hot L1s” has not been properly addressed in the mouse genome

(to date there is one defined example at chr4:21650298–21,656,544 (GRCm38/mm10))[308], yet this assay captures expression of such "hot" L1 elements that encode the machinery required to cause DNA damage and retrotranspose. This assay could therefore be useful in future experiments to correlate expression with other data e.g. methylation status of hot L1s and/or interferon levels, to further understand L1 mRNA expression in neuronal models of ageing.

The onset of human chronic pain is associated with increasing age[1,309]. Degeneration of DRG has also been documented to occur in ageing mice, specifically first observed during middle age (11 to 13 months)[283]. L1 activation has previously been documented to occur in ageing mouse liver and skeletal muscle and is a proposed mechanism of neuronal dysfunction. However, to date, no investigation of L1 in the peripheral nervous system has been reported in either mouse or human. We identified a decrease in L1 mRNA expression in DRG in both males and females aged 10 weeks (Figure 5.6). Female mice typically reach sexual maturity by 8 weeks and males by 5 weeks (however can vary between mouse strain) thus the effects seen here may be some reflection of puberty. Hormone levels can affect the levels of L1 expression[310,311]. L1 and Alu methylation status has also been linked to hormone levels and age of puberty onset/sexual maturation[312]. There is currently no literature specifically addressing retrotransposon expression and sexual maturation/puberty however defective piRNA pathways and aberrant L1 expression have been been implicated in Cryptorchidism, a disorder caused by a hormone imbalance during mini-puberty[313]. This highlights an interesting avenue to further explore the expression of L1 not just in development and ageing, but also in other life phases that may be subject to large hormonal influences.

One of the main findings in this chapter was that L1 mRNA expression was detected in DRG and was found to increase with age (Figure 5.6). Our data showed in both sexes

increases showed at least a 12-fold increase after 20 weeks, with the largest increase of 19-fold occurring in males at 35 weeks (Figure 5.6B). De Cecco *et al.* previously described a 6-fold increase in muscle of mice aged 36 months (156 weeks) compared to young controls aged 5 months (21 weeks)[84]. In a follow up study, aged mice at 24 months (104 weeks) and young controls at 4 months (17 weeks) were used and showed the same trend[86]. L1 mRNA increases varied greatly across tissue with age, for example; 3-fold (intestine), 10-fold (liver) and 60-fold (skin)[86]. This also placed into context that the large fold changes observed in our ageing model were not unreasonable. To further this, Bedrosian *et al.* recently reported that L1 mRNA is detected in mouse hippocampus at the early post-natal stage (P0-P7)[44]. However, from these studies it is not clear at which life phase L1 mRNA expression may increase. We used mice that represented several stages of life and profiled a significant increase in L1 mRNA at a similar age to the controls previously used, highlighting changes in L1 likely occur much earlier than previously documented.

To summarise, previous studies placed emphasis on the increase in L1 in aged tissues and imply this is a biological effect of old age, yet our results here highlight that in the dorsal root ganglia there is an increase during 'mature adulthood' that is maintained throughout age. The age ranges addressed in our DRG model demonstrated an increase in L1 mRNA expression in the mouse life phase that equates to 'mature adulthood', which holds relevance for the study of age-associated disorders like cancer and neurodegeneration. Mature adulthood is also the age at which we observed an increase in chronic pain prevalence in humans, therefore it is tempting to hypothesise that this may be a mechanism associated with the onset of chronic pain[1,309].

It has been reported that the increase of L1 expression is associated with an increase in INF-α and INF-β expression[86]. Some evidence supports INF-α as an endogenous pain inhibitor in the spinal cord however other evidence shows it contributes to

neuroinflammation, which also occurs in DRG following nerve injury[314-316]. This highlights that more work is needed regarding the role of type I interferon response in ageing DRG, and to assess its potential contribution to age associated pathology. IFN expression should be measured in a follow up study.

The increase in L1 expression was also correlated with an increase in TDP-43 expression occurring at the same time points (Figure 5.7), supported by the current literature reporting overexpression of TDP-43 and derepression of TEs[92]. TDP-43 mRNA levels have been shown to decrease with ageing however the previous studies only measured mRNA in brains of mice up to 50 days old (7 weeks)[103,104]. Our data extends much further beyond this time point and is more appropriate and reflective of age-associated neuronal changes. TDP-43 has been minimally described in the DRG but no study has yet been done to address mRNA levels in ageing DRG[285-287,317], our data is the first report of age-associated changes in TDP-43 mRNA in the DRG.

Krug *et al.* demonstrated that overexpression of hTDP-43 in *Drosophila* impaired siRNA silencing and resulted in an increase in TE expression, which is consistent with our data showing an increase in TDP-43 and an increase in L1 expression[92]. Neuronal expression of TDP-43 has been shown to affect age-dependent efficacy of siRNA silencing[92]. The mechanism for this effect is yet to be elucidated however it is predicted that TDP-43 directly interacts with siRNA machinery[318]. There is overwhelming evidence that overexpression of TDP-43 results in neurotoxic effects which contributes to neurodegeneration[94-102]. Several case studies have also implicated TDP-43 in sensory neuropathy yet so far it has been generally ignored as a potential contributor to sensory neuron degeneration[285-287]. This study highlights changes in age-dependent TDP-43 mRNA levels in DRG, and thus L1 derepression as a potential mechanism contributing to age associated degeneration in the peripheral nervous system.

It must be highlighted that the Thy1-YFP mice used for this part of the study are transgenic, originating from (C57BL/6J x CBA)F1. The transgene contains the gene for yellow fluorescent protein (YFP), cloned from jellyfish, under the thymus cell antigen 1 promoter[200]. This promoter is a 6.5 kb fragment cloned from the 5' upstream region of the gene Thy1, that drives neuron specific expression, therefore YFP is expressed in DRG[200]. There are several strains of this transgenic mouse. The insertion site in the strain used in this study (B6.Cg-Tg (Thy1-YFP)16Jrs/0) has not yet been mapped (for further information, refer to the MGI website http://www.informatics.jax.org/allele/MGI:3505585). Potential confounding effects of this transgene on global gene expression, and in particular, the effects on L1 expression are currently unknown. Several abnormal neuronal phenotypes have been reported in the YFP strain, including abnormal; cerebral cortex pyramidal cell morphology[319], neuromuscular synapse morphology and synaptic transmission[320], therefore the results from this study should not be extrapolated to non-transgenic models until the appropriate studies have been completed. The Thy1-YFP strain was utilised in this study due to difficulties obtaining wildtype mice of appropriate ages, therefore this study should be repeated in a wildtype mouse strain such as C5BL7/6 to rule out potential confounding effects on gene expression resulting from the YFP transgene.

In contrast to the result observed in DRG, in the brain we observed a decrease in TDP-43 with age (Figure 5.9B), consistent with *in vivo Drosophila* and mouse models that show TDP-43 mRNA expression decreases with age[103,104]. *In vitro* studies have also previously identified a 32% decrease in aged mouse primary cortical and motor cells[321]. The data from this model supports the literature describing a decrease of TDP-43 mRNA in the ageing male brain. The literature also describes an increase in TE expression in response to loss of nuclear TDP-43[93], attributed to loss of Dicer-2 mediated siRNA silencing[322] and heterochromatin maintenance[323], however we only observed a small increase in L1 mRNA expression which was not statistically significant or biologically significant due to similar levels of variability

seen in controls. Differential L1 copy numbers have been identified in certain regions of the brain (e.g. hippocampus has elevated levels)[324] therefore it could be possible that as we processed the entire left hemisphere of the brain, any increases in mRNA levels in specific brain regions were diluted during processing. In contrast, TDP-43 is ubiquitously expressed therefore age associated changes would be easier to detect.

Diseases associated with obesity such as type II diabetes, cardiovascular disease and cancer are the fastest rising in terms of prevalence in the western developed world[325]. Excessive consumption of high fat diets contributes to these health problems and mechanistically induces chronic low-grade inflammation in the central nervous system[326]. Dietary changes can affect the methylation status and expression of L1 and TDP-43, however this has previously only been addressed in muscle and adipose tissue[84,303,304]. Therefore, we were interested to assess if a high fat diet had any effect on expression of L1 and TDP-43 mRNA specifically in the brain. We found no evidence that high fat diet induced changes in L1 or TDP-43 expression using the HFD induced inflammation model.

The HFD induced inflammation model was fairly limited in assessing changes in expression as it was based on comparison of mice at a single age point in response to diet. No young controls that were fed different diets were available. The effect of dietary induced inflammation affects certain regions of the brain more than others (e.g. cerebellum but not cortex), therefore processing the whole brain may have dampened any observable changes, as discussed previously[326]. Upon reflection, a better study would be to profile L1 expression in response to a high fat diet with age using several time points specifically in the cerebellum (adopting a similar approach to the DRG model). As we previously documented age-associated increase in L1 in DRG, it would be relatively easy to extend the relevance of this study to pain, by sampling DRG taken from the HFD induced inflammation mouse model. Unfortunately, it was not possible to assess expression changes in DRG in the HFD model at

the time of study (due to difficulties obtaining tissues from collaborators), however it is highly recommended that these expanded studies should be completed in the future to offer more insight into the interplay between ageing, diet and retrotransposon expression and how this may affect the peripheral nervous system.

The anti-inflammatory secretory product of *A. viteae,* ES-62, has successfully been used to treat some inflammatory conditions (e.g. arthritis[291], asthma[292] and systemic lupus erythematosus[293]) however in some murine autoimmune models it has failed, including type I diabetes, multiple sclerosis and inflammatory bowel disease[327]. It has been postulated that the MyD88-dependent signalling that is targeted by ES-62, may be compensated for by some other inflammatory pathway, resulting in the intended anti-inflammatory effects to fail. In this model, we observed an increase in both L1 and TDP-43 expression in response to ES-62 treatment. This was not observed in response to high fat diet in aged mice, therefore the effects here either are because of ES-62 alone or a combinatorial effect between ageing, high fat diet and treatment. The relevant controls were not available in this model to decipher which variables contributed to this effect.

ES-62 disrupts MyD88 dimerisation and impairs the induction of inflammatory cascades involving NF-κB and MAPK[328]. Inhibition of MyD88 leads to defective *ERCC1*-DNA repair (via NER) and accumulation of DNA damage[61,329]. *Ercc1* deficient mice contain increased numbers of γ-H2AX foci in cells (indicative of DSBs) and *in vitro* studies have shown that *ERRC1* deficiency results in increased levels of L1 mobilisation[330,331]. *ERRC1* has also been suggested to play a role in the orientation bias observed in L1 insertions in transcribed genes[332]. This highlights a direct role between *ERCC1* and L1 activity. MyD88 is involved in immune activation in response to pathogenic nucleic-acids (detection via TLR7 and TLR9) and has been shown to play a role in systemic lupus erythematosus (SLE)[333]. However since beginning this investigation, studies have shown that the induction of inflammation via L1 is

mediated through cGAS/STING signalling[85,86]. Deficiency of MyD88 has no effect on the induction of cGAS, suggesting that these two inflammatory pathways are independent[334]. Therefore, we propose a hypothetical mechanism by which ES-62 could potentially disrupt *ERCC1*-mediated DNA repair and result in an accumulation of DNA damage, loss of suppression of L1 and ultimately result in an increase of L1 mRNA (Figure 5.12). It must be emphasised that this is a theoretical explanation for the increases observed in L1 expression and more studies would need to be conducted to test this hypothesis. There is no relevant information available in the literature regarding the relationship between TDP-43 and MyD88, however in the same *Ercc1*-deficient mouse model as previously described, a follow up study found that TDP-43 pathology (namely cytoplasmic inclusions/aggregates) were not affected[335]. Inhibition of MyD88 (via ES-62) supresses activation of NF-κB, which has been shown in a separate study to decrease TDP-43 expression[336]. This was not the effect seen in our study, suggesting that another unknown mechanism may be responsible. It has not yet been confirmed whether the ES-62 SMAs can cross the blood-brain barrier and target MyD88 as described. The blood-brain barrier is compromised in the case of inflammation, resulting in increased permeability[337]. Even if ES-62 cannot cross the barrier and directly affect MyD88 in the brain, some response in the brain is evident, therefore further research should still try to dissect what mechanism is responsible.

The sample sizes used in this study were relatively small, as this study was intended to act as an initial proof of principle. Despite trends being observed, relative expression values were variable which when combined with a small sample size, led to statistical significance not being reached. This study would benefit from increasing the sample sizes, which could also be ethically justified now that the initial results have given cause for further investigation. It must also be emphasised that differences in L1 and TDP-43 trends observed in DRG and brain should be interpreted with caution. The differences observed may simply be due to tissue specific effects or they could be due to difference in strain (genetic

background) and housing conditions between Liverpool and Strathclyde. All C5BL7/6 mice used in the high fat diet model in Strathclyde were treated with fenbendazole to control pinworm, whereas the Th1-YFP mice used in the ageing model in Liverpool were not. Fenbendazole has been reported to affect mRNA expression of transcription factor E2A and also impact the activation of B lymphocytes in aged mice (22 months compared against 2 months)[338]. As the work conducted in this chapter focused on both neuroinflammation and ageing, this treatment may have contributed to differences observed between animal models used in these studies. The potential impact on the targets studied in this project are unknown. This is a potential effect that may have and is therefore an important consideration for future experimental design. Many of these mice sampled (e.g. males and females used in ageing DRG study) were litter mates. Bedrosian *et al.* demonstrated that murine L1 mRNA expression can be influenced in the first week of life by the level of maternal care received by the animal, e.g. less maternal care resulted higher L1 mRNA levels compared to those that received high maternal care[44]. Less maternal care was also associated with greater L1 mRNA variability between individuals. Maternal care was not controlled for in this study therefore is an important factor to consider, especially with small sample numbers. Future experiments should be conducted using a single mouse model to enable direct comparison of data across tissues.

Our collaborators recently published a preprint documenting a >10% increase in lifespan in males treated with ES-62[302]. The ideas that both L1 mRNA and lifespan increased in the same model are not in agreement if we assume that increased L1 expression is detrimental to cellular function, which seems to be the dominating trend in the literature at present. Earlier work has demonstrated that increases in L1 retrotransposition is most common in the dentate gyrus, which is associated with memory and learning and adult neurogenesis[339]. Therefore increased L1 expression in the brain in non-pathological ageing may simply be a mechanism to generate neuronal diversity in response to stress and not

necessarily be detrimental to the survival of an organism[340]. Further work is needed to quantify L1 mRNA expression levels with specific markers of cellular degeneration, which would rely on more advanced techniques than relative quantification in order to establish 'healthy' basal levels and those considered 'toxic'. To summarise, we have demonstrated L1 and TDP-43 changes in various neuronal models of ageing and inflammation, highlighting that these targets may have relevance in non-pathological age-related neuroinflammation (i.e. not just in neurodegenerative disorders of the CNS), in both the peripheral nervous system, and in response to anti-inflammatory treatment ES-62 in the brain.

In ageing DRG, we observed an increase in L1 in association with TDP-43 increase, however this may be a normal observation in 'healthy' animals and not necessarily reflective of any age-associated neurodegeneration in these animals. This study should be expanded to assess for markers of degeneration in DRG such as TrkC expression as described by *Vaughan et al.*[283]. To further this, IFN-α and IFN-β should be assayed to correlate with changes in L1 expression and assess the presence of L1 mediated neuroinflammation. TDP-43 pathology research highlights the loss of nuclear TDP-43 and accumulation of cytoplasmic TDP-43 as a pathological mechanism, therefore this should also be assayed to assess if the age-associated changes are causative of any cytoplasmic inclusions, which may be of relevance to peripheral neuropathy and chronic pain pathology.

Expression of L1 has two downstream consequences that we currently understand, the first is mobilisation which can cause genomic instability/novel insertions which ultimately rewires the genetic circuitry that dictates neurobiological processes. The second is induction of an interferon response via accumulation of L1 cDNA in the cytoplasm. We have provided the first evidence that L1 is expressed in the peripheral system, specifically DRG. If further studies can attribute L1 expression with neuroinflammation in the peripheral nervous system, this may present an exciting and novel opportunity for intervention through reverse

transcriptase inhibitors (originally developed for HIV), which are currently being trialled in Aicardi–Goutières Syndrome, and showing effective reduction of interferon response[341]. We also documented an age-dependent increase in L1 mRNA associated with an increase in TDP-43 expression, highlighting a target for further analysis regarding age associated neuronal dysfunction in the peripheral nervous system. Importantly we identified that changes in L1 and TDP-43 expression occur in mature adulthood, not old age. Some recent studies have failed to identify any association between TDP-43 and L1 expression[342,343]. However, these methods were based on RNA-seq data which is subject to multiple errors (e.g. false positives resulting from exonisation). The literature places a large focus on L1 accumulation with ageing, but the relevant longitudinal studies in mammals have not yet been done. In addition, we have also documented an increase in L1 in association with ES-62 treatment, which highlights a potential alternative inflammatory pathway that may explain its failure in several autoimmune models.

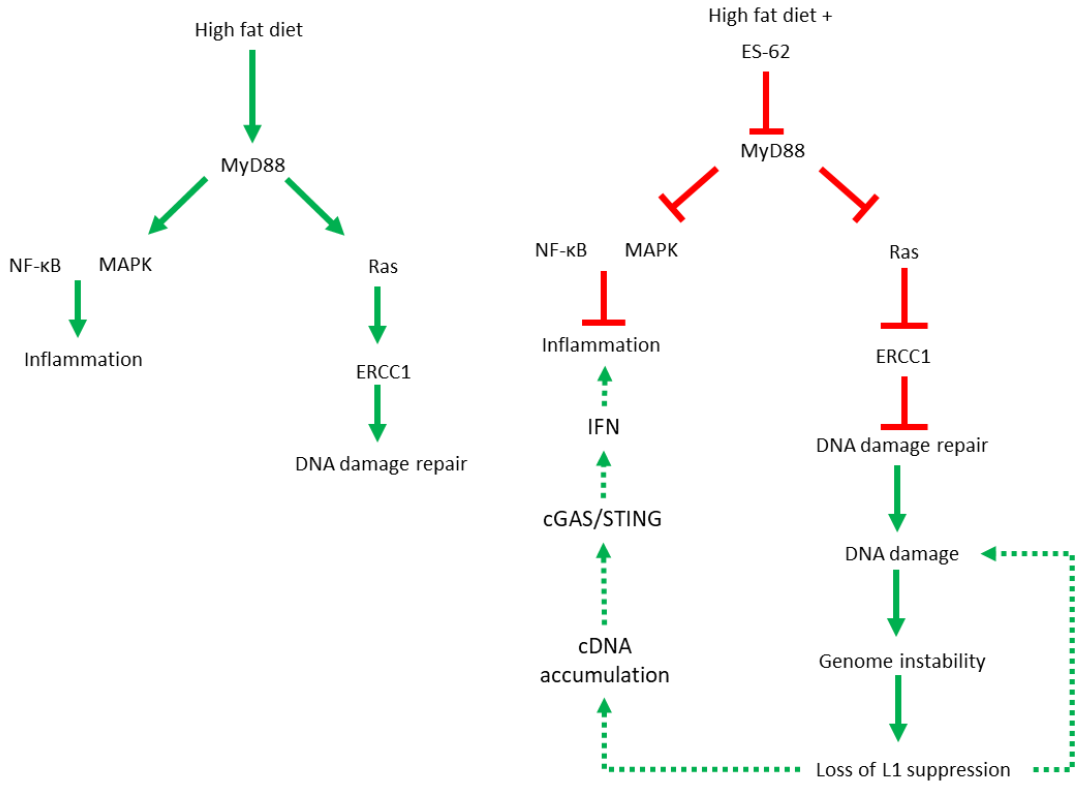


Figure 5.12. Hypothetical model of ES-62 induction of L1 expression. (Left) Activation of inflammatory response via MyD88 in response to high fat diet. Activation of NF-κB and MAPK mediates inflammatory response. MyD88 also enables ERCC1-DNA repair via Ras pathway. (Right) Inhibition of MyD88 via ES-62, and inhibition of inflammation through NF-κB and MAPK. Loss of ERCC1-mediated DNA repair could hypothetically lead to accumulation of DNA damage, genome instability and loss of L1 suppression. This in theory could induce inflammation via cGAS/STING or contribute further to DNA damage.

Chapter 6. Discussion

6.1 Thesis summary

The ultimate goal of this thesis was to conduct the first investigation into the potential role of retrotransposons in the context of chronic pain. To date, the field of pain genetics has made great use of genetic studies and animal models to identify the key molecular components of pain mechanisms, however the development of effective analgesics has not reflected this effort, partly due to a lack of translational research from rodent models to human. The work conducted in the first part of this thesis addressed the role of a human specific SVA insertion as a candidate regulatory domain for species specific differences observed in the expression patterns of pain genes *TRPV1* and *TRPV3*. We provided evidence to support the SVA is functional as a human specific regulatory domain at this locus. Furthermore, we identified this SVA as polymorphic and demonstrated potential for its application as a biomarker for pain related phenotypes.

The work conducted in the second part of this thesis explored neuronal L1 expression in different mouse models. L1 expression is associated with neuroinflammation in CNS neurodegenerative disorders and has been shown to increase in normal ageing in non-neuronal tissues however no study has investigated the potential contribution of L1 activity to the neuroinflammation that is a driving force in the development of chronic pain. We identified age-associated changes in L1 mRNA expression in DRG, presenting the first insight into L1 activity in the peripheral nervous system. Our data also highlighted a role for L1 activity in the brains of mice in response to the anti-inflammatory treatment ES-62, which is typically associated with increased longevity. This raised further questions regarding the pathogenicity of L1 in the ageing brain in the context of anti-inflammatory intervention. To summarise, the work undertaken in this thesis has highlighted the relevance of retrotransposons in several aspects of chronic pain research including species specific gene regulation, response to therapeutics, and age associated activity in the peripheral nervous system.

6.2 Conclusions

6.2.1.1 The SVA at the TRPV1 and TRPV3 locus is a human specific transcriptional regulator

Translating findings from preclinical rodent models into effective therapeutics often fails – due to a lack of efficacy or unsafe side effects when therapeutics are tested in humans during clinical trials[126,127]. This in part due to genomic and transcriptomic differences between mice and humans. The mouse and human genome are highly similar in terms synteny, however many differences are apparent including the content of transposable element specifically in terms of lineage specific repeats (Figure 1.6), which has contributed to the evolution of each species (Figure 1.8).

In the context of pain, RNA-seq studies comparing human and mouse DRG have identified a moderate level of conservation between the transcriptomes of humans and mice (63.8% genes shared similar tissue specific gene expression patterns) however have also highlighted a large degree of divergence between transcriptomes[6], which is highly relevant in the context of understanding human pain pathways and therapeutic discovery. An example of this is demonstrated in the case of *TRPV1* and *TRPV3*, two pain genes that are adjacent to one another on the same syntenic block between humans and mice (Figure 1.10), *TRPV1* is expressed in DRG in both human and mice, yet *TRPV3* is only expressed in human DRG (and not in mouse DRG) - highlighting species specific gene regulation[6,9,10,152,174]. TRPV1 and TRPV3 are symptomatic of the lack of translational research resulting from rodent models as several pharmaceuticals targeting these channels have failed to reach the clinic. Transposable elements influence species specific gene regulation, however have not been previously investigated in regards to species specific differences that may influence pain genes[7,34].

Chapter 3 focused on the transcriptional regulation of *TRPV1* and *TRPV3.* Bioinformatic analysis identified a human specific SVA retrotransposon at the intergenic

region between *TRPV1* and *TRPV3*, that was adjacent to a conserved ECR containing a MIR element which was predicted to function as a CRE (Figure 3.1). We hypothesised the SVA was functional as a novel CRE in the human genome, affecting human specific gene regulation and responsible for the species-specific differences described in the literature. Using reporter gene assays, the SVA was shown to function as a transcriptional regulator in the cell line HEK293 (Figure 3.6) and in murine DRG (Figure 3.8). This demonstrated the SVA sequence was functional in a neuronal cell type, which holds relevance for future studies investigating this element further in the context of pain. We did not obtain any evidence to support the role of the ECR as a transcriptional regulator in HEK293 cells, but reporter gene assays can vary greatly based on the cell line used[215], therefore it should be tested in others.

The SVA was then deleted in the cell line HEK293 using CRISPR (Figure 3.20). Cell lines containing a heterozygous deletion showed increased yet variable *TRPV3* expression compared to unedited cell lines (Figure 3.26). We also observed variable *TRPV1* expression in cells containing a heterozygous deletion (Figure 3.27). Homozygous deleted cell lines did not show the same variability as heterozygous deleted cell lines for either *TRPV1* or *TRPV3.* HEK293 cells are reported to carry four copies of chromosome 17 therefore we estimated there were four copies of the SVA[227]. A different number of deleted SVA copies may explain this variability due to differential transcriptional influences. For example, in the context of TRPV3, a deletion of a single SVA copy may result in a slight increase in expression but the deletion of 2 copies may result in a greater increase.

In chapter 4, it was demonstrated that the SVA was polymorphic in sequence length at the CT-rich domain (Figure 4.5). We concluded the variable results obtained in the heterozygous deletions may have been affected by different combinations of remaining alleles. We did not genotype the cell lines for polymorphism at the time of study however this should be completed to offer a better interpretation of this result. TEs have also been

shown to function in response to stimuli, therefore *TRPV1* and *TRPV3* mRNA changes this should be addressed in future experiments should assess potential changes in response to inflammatory stimuli, which may give insight into the potential role of the SVA in the process of sensitisation. We were unable to generate any ECR deletions using CRISPR (Figure 3.23). The efficiency of the protocol may have been dampened by genomic structure or poor guide design.

Taken together, the data presented in chapter 3 gives evidence to support the role of the SVA at the *TRPV1* and *TRPV3* locus as a human specific CRE which may in part explain the species-specific differences observed in the literature. To the best of our knowledge, at the time of writing this is the first report of a CRISPR mediated SVA deletion that is present in the reference genome (i.e. not an insertional mutation as previously demonstrated by the SVA in *TAF1).* The insertion of the SVA into this locus means it could potentially affect gene expression by the introduction of novel TFBS, affect DNA structure (e.g. G4 quadruplex formation as shown in chapter 4), or confer differential epigenetic modifications to the region.

Ultimately, the SVA may be an important regulatory factor relevant to consider in human conditions characterised by increased levels of *TRPV1* and *TRPV3* expression. This work was focused on *TRPV1* and *TRPV3* as they are well known pain genes, however this study was not conducted in a pain model, therefore the findings presented here can also be expanded to other conditions in which aberrant regulation of these genes is relevant, e.g. increased *TRPV3* expression is also implicated in psoriasis[344]. Going forwards, studies like this offer insight into the regulatory differences between rodents and humans, and highlight the functionality of TEs in species differences, which is crucial to understand in order to develop effective therapeutics. The future work to be completed regarding the functional activity of the SVA at TRPV1 and TRPV3 is presented in section “future work” in section 6.3.

6.2.1.2 The SVA at the *TRPV1* and *TRPV* locus shows potential as a biomarker for pain phenotypes

Increased *TRPV1* and *TRPV3* mRNA expression is evident in tissues associated with pain[163-168,170,171]. Multiple SNPs have been identified across the *TRPV1* and *TRPV3* locus in association with pain phenotypes[245-249]. Some SNPs have also been identified in association with *TRPV1* and *TRPV3* mRNA changes in asthma (Supplementary figure 2), however the causal mechanism has not yet been identified. TEs represent a large source of polymorphism in the genome however are largely unexplored, due to their repetitive nature, which means they are often excluded from genomic data analyses and are therefore overlooked.

In chapter 4, we investigated the SVA at the intergenic region between *TRPV1* and *TRPV3* for sequence length polymorphism. Using PCR, we identified at least 4 distinct SVA alleles which were conferred at least in part due to 4 different sized CT-rich domains (Figure 4.3), which were all identified as common genetic variants in the general population. We identified using DNA sequencing the CT-rich domain was composed of different TCTCCC hexamer repeat copy numbers but also different copy numbers of a novel 17mer repeat (Figure 4.5). We showed this novel 17mer was present across multiple SVA subclasses throughout the genome, with a specific enrichment in subclass D (Figure 4.6). Subsequently, from this analysis we identified SVA subclass E and F are burdened with annotation errors in the current genome build hg38. This carries implications for evolutionary and genome wide studies of SVAs and further improvement to SVA annotation are required.

Based on the CT variants identified, we genotyped two cohorts that were relevant to pain to look for associations between genotype/allele frequency and pain phenotypes and assess its potential use as a biomarker. The prevalence of pain is associated with ageing[1] therefore we first genotyped a healthy aged cohort (Dyne-Steele) with self-reported pain measures and identified significant associations between a heterozygous SVA genotype and self-reported pain (Figure 4.8A). To determine its use as a biomarker to assess risk of pain

with age, this would have to be genotyped in a far larger cohort with more robust pain measures. This however is unlikely as chronic pain is such a complex and heterogenous disease. As demonstrated by GWAS, identifying genetic variants in association with broad chronic pain phenotypes is difficult and likely not very informative in a biological context. If biomarkers can be developed to identify those at risk of pain with ageing, measures could be taken earlier in life to reduce risk (e.g. changes to diet to reduce neuroinflammation). Perhaps as our understanding of chronic pain develops in the era of personalised medicine, we will be able to more appropriately stratify chronic pain subgroups and manage this condition better.

Tramadol, a strong opioid used to treat pain has been shown to function as an agonist of TRPV1[258]. Many patients treated with tramadol experience ADRs and poor analgesic efficacy[257], hypothesised to be mediated in part by the activation of TRPV1. No genetic variants at *TRPV1* have been studied with regards to tramadol response therefore we investigated SVA genotypes with these response outcomes in an osteoporosis cohort being treated with tramadol for bone pain. We also observed the same heterozygous SVA genotype (associated with self-reported pain in healthy aged individuals) trending in association with ADRs (Figure 4.12) and poor analgesic efficacy (Figure 4.10) in osteoporosis patients treated with tramadol for bone pain. ADRs and poor analgesic efficacy resulting from tramadol are a large problem not only for osteoporosis patients but for a patients suffering from pain in general[257]. Tramadol has also been shown to be a prime contributor to opioid misuse in the US[4]. As a result, if patients that are predicted to respond poorly to tramadol can be identified, this would prevent unnecessary administration of the drug. The SVA should therefore be genotyped in much larger cohorts to assess its utility as a biomarker for ADRs and efficacy in response to tramadol. In addition, we also identified an increased frequency of a specific SVA allele in males with osteoporosis (Figure 4.14B), but not in healthy aged individuals (Figure 4.9). It is tempting to infer that this allele could perhaps be a biomarker of osteoporosis in

males however due to the small sample size this would have to be investigated in a much larger cohort. In summary, regarding the polymorphism of the SVA we found interesting associations that warrant further investigation. These studies are discussed in the "future work" section.

In chapter 4, we also conducted a preliminary investigation into the methylation status of the SVA and identified there is an increased frequency of unmethylated copies in both temporal cortex and whole blood (Figure 4.19A and B), which could indicate the SVA is active. This would support the results observed in chapter 3, which indicated the SVA was functional as a transcriptional regulator at the *TRPV1* and *TRPV3* locus (Figure 3.6, Figure 3.27). Interestingly, we also found variable methylation patterns across tissues between individuals (Figure 4.19D). No association could be made regarding links between the methylation status of the SVA and its genotype however it could be hypothesised that different SVA variants could be subject to different epigenetic modifications and thus confer different regulatory properties, i.e. function as an epiallele.

We can apply the findings from chapter 4 to the variable gene expression results observed in the CRISPR modified cell lines generated in chapter 3 (Figure 3.20). Cell lines that carried a heterozygous SVA deletion displayed the most variable TRPV1 and TRPV3 expression when compared to both wild-type controls and modified cell lines carrying a homozygous deletion (Figure 3.27). We discussed in the previous section (6.2.1.1) this could have been the result of a different number of SVA copies being deleted across the four copies of chromosome 17, however this could also be influenced by the different alleles remaining (if shown to be heterozygous). Furthermore, as we demonstrated the SVA is functional, the original hypothesis can be expanded beyond a simple genetic association of SVA variants and phenotypes. Due to the mechanisms in which TEs can affect gene regulation (Figure 1.7D), the polymorphic SVA alleles could potentially confer different regulatory properties to the

locus, and thus act as a functional genetic variant affecting *TRPV1* and *TRPV3* expression, therefore may be a causal genetic variant at this locus.

6.2.1.3 L1 expression increases in mouse DRG with ageing

Neuroinflammation is classed a hallmark of ageing and is an underlying driver of chronic pain[1,280]. L1 expression has also been shown to increase in ageing mouse non-neuronal tissues, which leads to an accumulation of L1 cDNA in the cytoplasm – which ultimately induces an inflammatory response[85,86]. In the CNS, a large body of research has emerged describing the loss of TE silencing and the induction of inflammatory markers as a contributor to neurodegenerative disorders like ALS[89]. Despite the increasing evidence that L1 is a contributor to neuroinflammation in the CNS, it has not yet been explored in the peripheral nervous system.

In chapter 5, we developed a robust protocol to profile L1 mRNA expression in mouse tissues (as it was not possible for us to obtain human tissues) and produced an ageing DRG RNA resource representing mice from early life phases (e.g. pre-puberty) to middle aged. This functioned as a model to study age related changes in the peripheral nervous system, which is a key driver of pain. We found L1 was expressed in DRG across the lifespan of both male and female mice and observed a significant increase in L1 mRNA expression during the transition between early and mature adulthood (Figure 5.6). We were surprised to observe relatively large fold L1 increases as early as mature adulthood as previous studies had indicated this was symptomatic of old age. This observation is very interesting in regards to chronic pain, as this is the same age/life phase at which we observe an increase in chronic pain prevalence in humans[1,309]. However, before any conclusions regarding the biological impact in ageing DRG and links to pain can be made, several inflammatory markers must be profiled in this model. This is discussed in the "future work" section (6.3.3).

TDP-43 is an RNA binding protein which has been heavily associated with L1 regulation. We observed a positive correlation between increasing L1 and TDP-43 mRNA with ageing in this mouse model (Figure 5.7). This could point to a response of TDP-43 in an attempt to regulate L1, or age associated TDP-43 deregulation in the DRG may be causative of the

increase in L1 – an observation documented throughout the literature. Whilst the relationship between L1 and TDP-43 is not clear until further studies have been completed, an overwhelming theme in the literature is that overexpression of TDP-43 is neurotoxic[94-102]. The bottom line is that we observed age-associated increase in both TDP-43 and L1 mRNA expression in DRG and increased expression of both has been heavily implicated in age associated neuronal dysfunction.

A key study missing from the literature is DNA methylation profiling in the DRG with increasing age. This would offer insight into a potential cause of L1 expression in the ageing DRG, which has been discussed in multiple studies[79-83]. If the role of L1 in the ageing DRG can be demonstrated in the future to be a mechanism underlying sterile inflammation that contributes to the development of chronic pain, it would offer a novel therapeutic strategy. For example, reverse transcriptase inhibitors have been shown to limit interferon response in Aicardi-Goutieres syndrome[341] (which is characterised by L1 activity) – thus highlighting a way to dampen L1 induced inflammation in the peripheral nervous system.

Whilst limited, there is evidence to support the role of L1 activation and retrotransposition in learning and memory, particularly in the dentate gyrus, and that L1 can be induced in response to stress to generate neuronal diversity[339,340]. Therefore, it must be considered that not all L1 activity is detrimental to an organism's fitness. Based on the method of relative quantification, we cannot be certain if the increase level of L1 mRNA is problematic for the cell without the follow up studies to profile inflammatory or neurodegenerative markers. We can however conclude that L1 plays a role in the peripheral nervous system and is a relevant target for future study.

6.2.1.4 L1 expression is modulated in the ageing brain in response to treatment with anti-inflammatory drug ES-62

Age-associated neuroinflammation can further be exacerbated in obesity, which is often a comorbidity that occurs with pain[280,288,289]. Therefore, the role of diet in L1 modulation was of interest to this project. In chapter 5, we expanded on the theme of L1 and age-associated neuroinflammation by measuring L1 and TDP-43 mRNA in the brains taken from a mouse model of HFD induced inflammation. We observed a decrease in TDP-43 mRNA in the brain of aged male fed normal chow (Figure 5.9B). No female samples were available for comparison however this did reflect reports in the literature that TDP-43 protein expression decreases with age[103,104]. We did not identify an increase in L1 mRNA expression in the brain (Figure 5.9A) however we postulated this may have been due to methodological issues – i.e. purifying RNA from the entire left hemisphere when L1 activity may depend upon brain region (e.g. increased activity in the hippocampus)[324]. We did not find any increase in L1 mRNA or TDP-43 in response to high fat diet (Figure 5.10)– which did not support our hypothesis, however this study and the interpretation of the data could have been improved with the addition of young control mice fed normal chow or HFD.

The most interesting finding from the HFD induced inflammation mouse model was an increase in L1 mRNA expression in HFD induced mice that were also subject to treatment with anti-inflammatory drug ES-62 (Figure 5.11). An increase in TDP-43 was also observed in this condition (Figure 5.11), again demonstrating a link between L1 and TDP-43 in a non-neurodegenerative context. Previous studies using this drug had success when tested in various mouse models of inflammatory disease[291-293] (e.g. asthma) but failure in autoimmune models (e.g. multiple sclerosis)[345]. It had been proposed that the MyD88 inflammatory pathway that is inhibit by ES-62 was compensated for by some other inflammatory pathway, as a possible explanation for the drugs failure[327]. From the data and a literature search of the crossover between the targeted inflammatory pathways, DNA repair pathways and L1, we

proposed a model involving the DNA repair protein ERRC1 – which already has well established links with L1 activity[330-332]. In mice fed a high fat diet, a central inflammatory signalling protein named MyD88 triggers inflammation via NF-κβ and MAPK pathways[328]. It also induces the Ras pathway, which promotes DNA repair via ERCC1[61,329]. In the presence of ES-62, these inflammatory and DNA repair pathways are inhibited[61,329]. The absence of ERCC1 has been shown to result in an increase in both DNA damage and L1 mobilisation[330,331]. Previous work has further shown that L1 activity (and cDNA accumulation in the cytoplasm) induces inflammation through activation interferon response via cGAS/STING signalling[85,86]. Therefore in this model we hypothesised that ERCC1 is potentially inhibited by ES-62 via indirect inhibition of MyD88, resulting in DNA damage, loss of L1 suppression and induction of the interferon response (Figure 5.12). This alternate route to inflammation could explain the lack of efficacy in certain inflammatory models.

Interestingly, ES-62 treatment in this HFD induced inflammation model results in a >10% increase in the lifespan of males[302]. One current theory is that somatic L1 activity associated with ageing negatively impacts longevity however this model suggests otherwise[45]. Our results did not identify any major sex specific differences in L1 at the time at which these animals were culled (500 days) (Figure 5.11), however it would be interesting to profile the L1 mRNA and markers of DNA damage in mice which are long-lived (approximately 700 days) after natural death to assess if increased levels of L1 can be tolerated and if differences occur between sexes. This may be an important indicator of healthspan, which is important to maintain with increased lifespan in any case. These findings demonstrate the interplay between DNA repair mechanisms and inflammatory responses and the importance of both mechanisms of cell maintenance and defence in terms of longevity. As a result of these findings, and of those in the ageing DRG model, further studies are now underway to explore age associated L1 induced inflammation in broader neuronal context compared to previous models of CNS neurodegeneration.

6.3 Future work

6.3.1 Future experiments using CRISPR and cell lines containing SVA deletion

To expand of the work conducted in chapter 3, we can explore the effects of the SVA deletion beyond the *TRPV1* and *TRPV3* locus and look more broadly across chr17p13.2 (Figure 3.1). This region was identified by Gianfrancesco *et al.* as one of 30 loci in the genome that were enriched for SVA insertions[216]. All SVAs at chr17p13.2 were enriched for SVA D, and the majority were human specific – indicating that this region of the genome had undergone significant change in recent evolutionary history. In addition, other genes at this 1 Mb region were described in pathways that are relevant to pain (Supplementary table 1).

The CRISPR and qPCR protocol developed in this thesis provides a good framework to assess the impact of the SVAs on the regulation of other genes at this region. For example, does the deletion of the SVA at *TRPV1* and *TRPV3* affect the expression of other genes across the region? Or, how might deletion of all the SVAs at this region influence gene expression? This CRISPR protocol should also be applied to induced pluripotent stem cells (iPSCs), which can be differentiated into sensory neurons[241], to further test the effects of the SVA in a model more appropriate to pain biology. This assay could also be applied to a HaCaT (immortalized human keratinocyte cell line) M5-stimulated model of inflammation to assess the SVA function in relevance to *TRPV3* in psoriasis.

6.3.2 Linking SVA genotype to function

To expand on the work conducted in chapter 4, the development of a tagging SNP is currently in the pipeline with our collaborators Neil Pendleton and Antony Payton at the University of Manchester. This involves a bioinformatic approach imputing SNPs using a software called PLINK (http://zzz.bwh.harvard.edu/plink/download.shtml). This will identify SNPs that are in linkage disequilibrium with specific SVA alleles and enable us to use the SNP as a tag. If successful, this will enable us to computationally genotype larger cohorts must faster than the PCR based approach and will enable us to explore associations with more phenotypes and/or correlate with other datasets to give insight into the function of the SVA.

Our group also has access to data from the NABEC: North American Brain Expression Consortium. This dataset has SNP array, gene expression and methylation data from the brains of 382 neurologically healthy individuals. This would enable investigations into SVA genotype, methylation status and effects on *TRPV1* and *TRPV3* expression (which are both expressed in brain in human). The sample size of the preliminary study conducted in chapter 4, in which we looked at the methylation status of the SVA, was too small to identify any associations between genotype and methylation status. The development of tagging SNPs would enable us to overcome this limitation. For example, we could explore questions such as: does SVA genotype associate with methylation status, and does this affect *TRPV1* and/or *TRPV3* expression?

This methodology can also be applied to data generated by the 1000 genomes project which contains whole genome sequencing and RNA-seq for over 4000 individuals. This would enable us to discern links between SVA genotype and molecular phenotype. In chapter 4, we showed data to suggest that SVA genotype was potentially associated with complex phenotypes such as pain and response to tramadol. The development of a tagging SNP would enable other researchers to explore this variant in much larger cohorts and stratify potential associations with other pain phenotypes relevant to TRPV1 and TRPV3.

6.3.3 Further studies addressing L1 mRNA expression

We are extending our collaboration with the University of Strathclyde to further assess the role of L1 and TDP-43 expression in the brain. Initial studies will profile inflammatory markers such as INFα and INFβ using qPCR, indicating an inducible interferon response from the presence of L1 cDNA[86]. This will also be conducted in the ageing DRG resource to address neuroinflammation associated with L1 in the peripheral nervous system. In addition, the roles of previously characterised genes that are associated with L1 activity in neurodegenerative disorders, such as *SIRT6* and *TREX1*, will be profiled using qPCR to identify potential roles in the L1 mRNA changes observed in these models. To further assay the impact of L1 mRNA increases in neuronal models, genomic L1 copy number will also be measured using a PCR assay.

We have also proposed a collaboration with the Pain Neurobiology Research Group at the School of Behavioural and Brain Science at the University of Texas Dallas (UTD), USA. This group perform RNA-seq in human DRG from healthy controls and individuals with well characterised pain. Utilising resources generated by collaborators and protocols developed throughout this project, we can address if L1 mRNA is expressed in human DRG, as was documented in mice, and profile the methylation status of "hot L1s" in relation to L1 mRNA expression. There is also scope within this project to measure genomic L1 copy number or even somatic insertions in the DRG (compared against whole blood) to give insight into the level of somatic variation that may exist in the peripheral nervous system. We can further correlate these findings with gene expression to address the functional relevance.

In addition to work on L1, we can also address questions regarding the genotype of the SVA at TRPV1 and TRPV3 in relation to specific DRG molecular phenotypes using RNA-seq data. This will enable us to further address the relevance of this SVA insertion in pain pathways. Using a valuable resource like this will undoubtedly raise the profile of retrotransposon biology in the context of chronic pain in humans, creating a novel

opportunity for therapeutic intervention and hopefully offer better understanding of age-associated changes that contribute to its development in humans.

Bibliography

1 Fayaz, A., Croft, P., Langford, R. M., Donaldson, L. J. & Jones, G. T. Prevalence of chronic pain in the UK: a systematic review and meta-analysis of population studies. *BMJ Open* **6**, e010364, doi:10.1136/bmjopen-2015-010364 (2016).

2 Kuehn, B. Chronic Pain Prevalence. *Jama* **320**, 1632, doi:10.1001/jama.2018.16009 (2018).

3 Gaskin, D. J. & Richard, P. The Economic Costs of Pain in the United States. doi:https://www.ncbi.nlm.nih.gov/books/NBK92521/ (2011).

4 Lyden, J. & Binswanger, I. A. The United States opioid epidemic. *Semin Perinatol* **43**, 123-131, doi:10.1053/j.semperi.2019.01.001 (2019).

5 Zorina-Lichtenwalter, K., Meloto, C. B., Khoury, S. & Diatchenko, L. Genetic predictors of human chronic pain conditions. *Neuroscience* **338**, 36-62, doi:10.1016/j.neuroscience.2016.04.041 (2016).

6 Ray, P. *et al.* Comparative transcriptome profiling of the human and mouse dorsal root ganglia: an RNA-seq-based resource for pain and sensory neuroscience research. *Pain* **159**, 1325-1345, doi:10.1097/j.pain.0000000000001217 (2018).

7 Sundaram, V. *et al.* Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* **24**, 1963-1976, doi:10.1101/gr.168872.113 (2014).

8 Bennett, E. A., Coleman, L. E., Tsui, C., Pittard, W. S. & Devine, S. E. Natural Genetic Variation Caused by Transposable Elements in Humans. *Genetics* **168**, 933-951, doi:10.1534/genetics.104.031757 (2004).

9 Smith, G. D. *et al.* TRPV3 is a temperature-sensitive vanilloid receptor-like protein. *Nature* **418**, 186-190, doi:10.1038/nature00894 (2002).

10 Xu, H. *et al.* TRPV3 is a calcium-permeable temperature-sensitive cation channel. *Nature* **418**, 181-186, doi:10.1038/nature00882 (2002).

11 McClintock, B. Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol* **21**, 197-216, doi:10.1101/sqb.1956.021.01.017 (1956).

12 de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genet* **7**, doi:10.1371/journal.pgen.1002384 (2011).

13 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).

14 Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703, doi:10.1038/nrg2640 (2009).

15 Tang, W., Mun, S., Joshi, A., Han, K. & Liang, P. Mobile elements contribute to the uniqueness of human genome with 15,000 human-specific insertions and 14 Mbp sequence increase. *DNA Res* **25**, 521-533, doi:10.1093/dnares/dsy022 (2018).

16 Misiak, B., Ricceri, L. & Sasiadek, M. M. Transposable Elements and Their Epigenetic Regulation in Mental Disorders: Current Evidence in the Field. *Front Genet* **10**, 580, doi:10.3389/fgene.2019.00580 (2019).

17 Kury, P. *et al.* Human Endogenous Retroviruses in Neurological Diseases. *Trends Mol Med* **24**, 379-394, doi:10.1016/j.molmed.2018.02.007 (2018).

18 Li, W. *et al.* Human endogenous retrovirus-K contributes to motor neuron disease. *Sci Transl Med* **7**, 307ra153, doi:10.1126/scitranslmed.aac8201 (2015).

19 Wildschutte, J. H. *et al.* Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc Natl Acad Sci U S A* **113**, E2326-2334, doi:10.1073/pnas.1602336113 (2016).

20 Kazazian, H. H., Jr. & Moran, J. V. Mobile DNA in Health and Disease. *N Engl J Med* **377**, 361-370, doi:10.1056/NEJMra1510092 (2017).

21 Majumdar, S. & Rio, D. C. P transposable elements in Drosophila and other eukaryotic organisms. *Microbiol Spectr* **3**, MDNA3-0004-2014, doi:10.1128/microbiolspec.MDNA3-0004-2014 (2015).

22 Savage, A. L. *et al.* Retrotransposons in the development and progression of amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry* **90**, 284-293, doi:10.1136/jnnp-2018-319210 (2019).

23 Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* **100**, 5280-5285, doi:10.1073/pnas.0831042100 (2003).

24 Deininger, P. Alu elements: know the SINEs. *Genome Biol* **12**, 236, doi:10.1186/gb-2011-12-12-236 (2011).

25 Wang, H. *et al.* SVA elements: a hominid-specific retroposon family. *J Mol Biol* **354**, 994-1007, doi:10.1016/j.jmb.2005.09.085 (2005).

26 Hancks, D. C. & Kazazian, H. SVA retrotransposons: Evolution and genetic instability. *Semin Cancer Biol* **20**, 234-245, doi:10.1016/j.semcancer.2010.04.001 (2010).

27 Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562, doi:10.1038/nature01262 (2002).

28 Goodier, J. L., Ostertag, E. M., Du, K. & Kazazian, H. H., Jr. A novel active L1 retrotransposon subfamily in the mouse. *Genome Res* **11**, 1677-1685, doi:10.1101/gr.198301 (2001).

29 Buckley, R. M., Kortschak, R. D., Raison, J. M. & Adelson, D. L. Similar Evolutionary Trajectories for Retrotransposon Accumulation in Mammals. *Genome Biol Evol* **9**, 2336-2353, doi:10.1093/gbe/evx179 (2017).
30 Romanish, M. T., Nakamura, H., Lai, C. B., Wang, Y. & Mager, D. L. A novel protein isoform of the multicopy human NAIP gene derives from intragenic Alu SINE promoters. *PLoS One* **4**, e5761, doi:10.1371/journal.pone.0005761 (2009).
31 Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165**, 349-357, doi:10.1126/science.165.3891.349 (1969).
32 Nikitin, D. *et al.* Profiling of Human Molecular Pathways Affected by Retrotransposons at the Level of Regulation by Transcription Factor Proteins. *Front Immunol* **9**, 30, doi:10.3389/fimmu.2018.00030 (2018).
33 Kagawa, T. *et al.* Recessive inheritance of population-specific intronic LINE-1 insertion causes a rotor syndrome phenotype. *Hum Mutat* **36**, 327-332, doi:10.1002/humu.22745 (2015).
34 Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083-1087, doi:10.1126/science.aad5497 (2016).
35 Mirabelli, P., Coppola, L. & Salvatore, M. Cancer Cell Lines Are Useful Model Systems for Medical Research. *Cancers (Basel)* **11**, doi:10.3390/cancers11081098 (2019).
36 Yu, K. *et al.* Comprehensive transcriptomic analysis of cell lines as models of primary tumors across 22 tumor types. *Nat Commun* **10**, 3574, doi:10.1038/s41467-019-11415-2 (2019).
37 Cosby, R. L., Chang, N. C. & Feschotte, C. Host-transposon interactions: conflict, cooperation, and cooption. *Genes Dev* **33**, 1098-1116, doi:10.1101/gad.327312.119 (2019).
38 Jacobs, F. M. *et al.* An evolutionary arms race between KRAB zinc finger genes 91/93 and SVA/L1 retrotransposons. *Nature* **516**, 242-245, doi:10.1038/nature13760 (2014).
39 Deniz, O., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat Rev Genet* **20**, 417-431, doi:10.1038/s41576-019-0106-6 (2019).
40 Rajan, K. S. & Ramasamy, S. Retrotransposons and piRNA: the missing link in central nervous system. *Neurochem Int* **77**, 94-102, doi:10.1016/j.neuint.2014.05.017 (2014).
41 Fasching, L. *et al.* TRIM28 Represses Transcription of Endogenous Retroviruses in Neural Progenitor Cells. *Cell Rep* **10**, 20-28, doi:10.1016/j.celrep.2014.12.004 (2015).
42 Xie, M. *et al.* DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* **45**, 836-841, doi:10.1038/ng.2649 (2013).
43 Teneng, I., Montoya-Durango, D. E., Quertermous, J. L., Lacy, M. E. & Ramos, K. S. Reactivation of L1 retrotransposon by benzo(a)pyrene involves complex genetic and epigenetic regulation. *Epigenetics* **6**, 355-367, doi:10.4161/epi.6.3.14282 (2011).
44 Bedrosian, T. A., Quayle, C., Novaresi, N. & Gage, F. H. Early life experience drives structural variation of neural genomes in mice. *Science* **359**, 1395-1399, doi:10.1126/science.aah3378 (2018).
45 Laurent, G. S., Hammell, N. & McCaffrey, T. A. A LINE-1 Component to Human Aging: Do LINE elements exact a longevity cost for evolutionary advantage? *Mech Ageing Dev* **131**, 299-305, doi:10.1016/j.mad.2010.03.008 (2010).
46 Faulkner, G. J. Retrotransposons: mobile and mutagenic from conception to death. *FEBS Lett* **585**, 1589-1594, doi:10.1016/j.febslet.2011.03.061 (2011).
47 Wang, L., Norris, E. T. & Jordan, I. K. Human Retrotransposon Insertion Polymorphisms Are Associated with Health and Disease via Gene Regulatory Phenotypes. *Front Microbiol* **8**, doi:10.3389/fmicb.2017.01418 (2017).
48 Hancks, D. C. & Kazazian, H. H. Roles for retrotransposon insertions in human disease. *Mobile DNA* **7**, 9, doi:10.1186/s13100-016-0065-9 (2016).
49 Baillie, J. K. *et al.* Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**, 534-537, doi:10.1038/nature10531 (2011).
50 Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127-1131, doi:10.1038/nature08248 (2009).
51 Evrony, G. D. *et al.* Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **151**, 483-496, doi:10.1016/j.cell.2012.09.035 (2012).
52 Grandi, F. C. & An, W. Non-LTR retrotransposons and microsatellites: Partners in genomic variation. *Mob Genet Elements* **3**, doi:10.4161/mge.25674 (2013).
53 Savage, A. L. *et al.* An evaluation of a SVA retrotransposon in the FUS promoter as a transcriptional regulator and its association to ALS. *PLoS One* **9**, e90833, doi:10.1371/journal.pone.0090833 (2014).
54 Bragg, D. C. *et al.* Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. *Proc Natl Acad Sci U S A* **114**, E11020-e11028, doi:10.1073/pnas.1712526114 (2017).
55 Michaud, E. J. *et al.* Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. *Genes Dev* **8**, 1463-1472, doi:10.1101/gad.8.12.1463 (1994).

56 Gardner, E. J. *et al.* The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res* **27**, 1916-1929, doi:10.1101/gr.218032.116 (2017).

57 Niedernhofer, L. J. *et al.* Nuclear Genomic Instability and Aging. *Annu Rev Biochem* **87**, 295-322, doi:10.1146/annurev-biochem-062917-012239 (2018).

58 Driver, C. J. & McKechnie, S. W. Transposable elements as a factor in the aging of Drosophila melanogaster. *Ann N Y Acad Sci* **673**, 83-91, doi:10.1111/j.1749-6632.1992.tb27439.x (1992).

59 De Cecco, M. *et al.* Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* **12**, 247-256, doi:10.1111/acel.12047 (2013).

60 Gasior, S. L., Wakeman, T. P., Xu, B. & Deininger, P. L. The Human LINE-1 Retrotransposon Creates DNA Double-strand Breaks. *J Mol Biol* **357**, 1383-1393, doi:10.1016/j.jmb.2006.01.089 (2006).

61 Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **273**, 5858-5868, doi:10.1074/jbc.273.10.5858 (1998).

62 Rauch, T. A. *et al.* High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc Natl Acad Sci U S A* **105**, 252-257, doi:10.1073/pnas.0710735105 (2008).

63 Saito, K. *et al.* Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer. *Clin Cancer Res* **16**, 2418-2426, doi:10.1158/1078-0432.ccr-09-2819 (2010).

64 Mima, K. *et al.* Tumor LINE-1 methylation level and colorectal cancer location in relation to patient survival. *Oncotarget* **7**, 55098-55109, doi:10.18632/oncotarget.10398 (2016).

65 Park, S. Y. *et al.* Alu and LINE-1 hypomethylation is associated with HER2 enriched subtype of breast cancer. *PLoS One* **9**, e100429, doi:10.1371/journal.pone.0100429 (2014).

66 Iwagami, S. *et al.* LINE-1 hypomethylation is associated with a poor prognosis among patients with curatively resected esophageal squamous cell carcinoma. *Ann Surg* **257**, 449-455, doi:10.1097/SLA.0b013e31826d8602 (2013).

67 De Luca, C. *et al.* Enhanced expression of LINE-1-encoded ORF2 protein in early stages of colon and prostate transformation. *Oncotarget* **7**, 4048-4061, doi:10.18632/oncotarget.6767 (2016).

68 Rodic, N. *et al.* Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat Med* **21**, 1060-1064, doi:10.1038/nm.3919 (2015).

69 Scott, E. C. *et al.* A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res* **26**, 745-755, doi:10.1101/gr.201814.115 (2016).

70 Pisanic, T. R., 2nd *et al.* Long Interspersed Nuclear Element 1 Retrotransposons Become Deregulated during the Development of Ovarian Cancer Precursor Lesions. *Am J Pathol* **189**, 513-520, doi:10.1016/j.ajpath.2018.11.005 (2019).

71 Briggs, E. M. *et al.* Long interspersed nuclear element-1 expression and retrotransposition in prostate cancer cells. *Mob DNA* **9**, doi:10.1186/s13100-017-0106-z (2018).

72 Coufal, N. G. *et al.* Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc Natl Acad Sci U S A* **108**, 20382-20387, doi:10.1073/pnas.1100273108 (2011).

73 Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468**, 443-446, doi:10.1038/nature09544 (2010).

74 Jacob-Hirsch, J. *et al.* Whole-genome sequencing reveals principles of brain retrotransposition in neurodevelopmental disorders. *Cell Res* **28**, 187-203, doi:10.1038/cr.2018.8 (2018).

75 Bundo, M. *et al.* Increased l1 retrotransposition in the neuronal genome in schizophrenia. *Neuron* **81**, 306-313, doi:10.1016/j.neuron.2013.10.053 (2014).

76 Guo, C. *et al.* Tau Activates Transposable Elements in Alzheimer's Disease. *Cell Rep* **23**, 2874-2880, doi:10.1016/j.celrep.2018.05.004 (2018).

77 Liu, S. *et al.* Inverse changes in L1 retrotransposons between blood and brain in major depressive disorder. *Sci Rep* **6**, 37530, doi:10.1038/srep37530 (2016).

78 Maynard, S., Fang, E. F., Scheibye-Knudsen, M., Croteau, D. L. & Bohr, V. A. DNA Damage, DNA Repair, Aging, and Neurodegeneration. *Cold Spring Harb Perspect Med* **5**, doi:10.1101/cshperspect.a025130 (2015).

79 Van Meter, M. *et al.* SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat Commun* **5**, 5011, doi:10.1038/ncomms6011 (2014).

80 Dennis, S., Sheth, U., Feldman, J. L., English, K. A. & Priess, J. R. C. elegans germ cells show temperature and age-dependent expression of Cer1, a Gypsy/Ty3-related retrotransposon. *PLoS Pathog* **8**, e1002591, doi:10.1371/journal.ppat.1002591 (2012).

81 Maxwell, P. H., Burhans, W. C. & Curcio, M. J. Retrotransposition is associated with genome instability during chronological aging. *Proc Natl Acad Sci U S A* **108**, 20376-20381, doi:10.1073/pnas.1100271108 (2011).

82 Li, W. *et al.* Activation of transposable elements during aging and neuronal decline in Drosophila. *Nat Neurosci* **16**, 529-531, doi:10.1038/nn.3368 (2013).

83 Chen, H., Zheng, X., Xiao, D. & Zheng, Y. Age-associated de-repression of retrotransposons in the Drosophila fat body, its potential cause and consequence. *Aging Cell* **15**, 542-552, doi:10.1111/acel.12465 (2016).

84 De Cecco, M. *et al.* Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging (Albany NY)* **5**, 867-883 (2013).

85 De Cecco, M. *et al.* L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73-78, doi:10.1038/s41586-018-0784-9 (2019).

86 Simon, M. *et al.* LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation. *Cell Metab* **29**, 871-885.e875, doi:10.1016/j.cmet.2019.02.014 (2019).

87 Mavragani, C. P. *et al.* Expression of Long Interspersed Nuclear Element 1 Retroelements and Induction of Type I Interferon in Patients With Systemic Autoimmune Disease. *Arthritis Rheumatol* **68**, 2686-2696, doi:10.1002/art.39795 (2016).

88 Mavragani, C. P. *et al.* Defective regulation of L1 endogenous retroelements in primary Sjogren's syndrome and systemic lupus erythematosus: Role of methylating enzymes. *J Autoimmun* **88**, 75-82, doi:10.1016/j.jaut.2017.10.004 (2018).

89 Saleh, A., Macia, A. & Muotri, A. R. Transposable Elements, Inflammation, and Neurological Disease. *Front Neurol* **10**, doi:10.3389/fneur.2019.00894 (2019).

90 Prasad, A., Bharathi, V., Sivalingam, V., Girdhar, A. & Patel, B. K. Molecular Mechanisms of TDP-43 Misfolding and Pathology in Amyotrophic Lateral Sclerosis. *Front Mol Neurosci* **12**, doi:10.3389/fnmol.2019.00025 (2019).

91 Li, W., Jin, Y., Prazak, L., Hammell, M. & Dubnau, J. Transposable elements in TDP-43-mediated neurodegenerative disorders. *PLoS One* **7**, e44099, doi:10.1371/journal.pone.0044099 (2012).

92 Krug, L. *et al.* Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS. *PLoS Genet* **13**, e1006635, doi:10.1371/journal.pgen.1006635 (2017).

93 Liu, E. Y. *et al.* Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. *Cell Rep* **27**, 1409-1421 e1406, doi:10.1016/j.celrep.2019.04.003 (2019).

94 Estes, P. S. *et al.* Wild-type and A315T mutant TDP-43 exert differential neurotoxicity in a Drosophila model of ALS. *Hum Mol Genet* **20**, 2308-2321, doi:10.1093/hmg/ddr124 (2011).

95 Xu, Y. F. *et al.* Wild-type human TDP-43 expression causes TDP-43 phosphorylation, mitochondrial aggregation, motor deficits, and early mortality in transgenic mice. *J Neurosci* **30**, 10851-10859, doi:10.1523/jneurosci.1630-10.2010 (2010).

96 Zhang, T., Mullane, P. C., Periz, G. & Wang, J. TDP-43 neurotoxicity and protein aggregation modulated by heat shock factor and insulin/IGF-1 signaling. *Hum Mol Genet* **20**, 1952-1965, doi:10.1093/hmg/ddr076 (2011).

97 Vaccaro, A. *et al.* in *PLoS Genet* Vol. 8 (2012).

98 Ash, P. E. *et al.* Neurotoxic effects of TDP-43 overexpression in C. elegans. *Hum Mol Genet* **19**, 3206-3218, doi:10.1093/hmg/ddq230 (2010).

99 Tsuiji, H. *et al.* TDP-43 accelerates age-dependent degeneration of interneurons. *Sci Rep* **7**, 14972, doi:10.1038/s41598-017-14966-w (2017).

100 Igaz, L. M. *et al.* Dysregulation of the ALS-associated gene TDP-43 leads to neuronal death and degeneration in mice. *J Clin Invest* **121**, 726-738, doi:10.1172/jci44867 (2011).

101 Wang, W. *et al.* The ALS disease-associated mutant TDP-43 impairs mitochondrial dynamics and function in motor neurons. *Hum Mol Genet* **22**, 4706-4719, doi:10.1093/hmg/ddt319 (2013).

102 Stallings, N. R., Puttaparthi, K., Luther, C. M., Burns, D. K. & Elliott, J. L. Progressive motor weakness in transgenic mice expressing human TDP-43. *Neurobiol Dis* **40**, 404-414, doi:10.1016/j.nbd.2010.06.017 (2010).

103 Cragnaz, L. *et al.* An age-related reduction of brain TBPH/TDP-43 levels precedes the onset of locomotion defects in a Drosophila ALS model. *Neuroscience* **311**, 415-421, doi:10.1016/j.neuroscience.2015.10.037 (2015).

104 Liu, Y. *et al.* Changes in TDP-43 expression in development, aging, and in the neurofilament light protein knockout mouse. *Neurobiol Aging* **36**, 1151-1159, doi:10.1016/j.neurobiolaging.2014.10.001 (2015).

105 Cox, J. J. *et al.* An SCN9A channelopathy causes congenital inability to experience pain. *Nature* **444**, 894-898, doi:10.1038/nature05413 (2006).

106 Price, T. J. & Dussor, G. Evolution: The Advantage of 'Maladaptive' Pain Plasticity. *Curr Biol* **24**, R384-386, doi:10.1016/j.cub.2014.04.011 (2014).

107 Haberberger, R. V., Barry, C., Dominguez, N. & Matusica, D. Human Dorsal Root Ganglia. *Front Cell Neurosci* **13**, doi:10.3389/fncel.2019.00271 (2019).

108 Williams, L. *et al.* Prevalence and impact of depression and pain in neurology outpatients. *J Neurol Neurosurg Psychiatry* **74**, 1587-1589, doi:10.1136/jnnp.74.11.1587 (2003).

109 Hocking, L. J., Morris, A. D., Dominiczak, A. F., Porteous, D. J. & Smith, B. H. Heritability of chronic pain in 2195 extended families. *Eur J Pain* **16**, 1053-1063, doi:10.1002/j.1532-2149.2011.00095.x (2012).

110 Nielsen, C. S., Knudsen, G. P. & Steingrimsdottir, O. A. Twin studies of pain. *Clin Genet* **82**, 331-340, doi:10.1111/j.1399-0004.2012.01938.x (2012).
111 Diatchenko, L. *et al.* Genetic basis for individual variations in pain perception and the development of a chronic pain condition. *Hum Mol Genet* **14**, 135-143, doi:10.1093/hmg/ddi013 (2005).
112 Lacroix-Fralish, M. L., Ledoux, J. B. & Mogil, J. S. The Pain Genes Database: An interactive web browser of pain-related transgenic knockout studies. *Pain* **131**, 3.e1-4, doi:10.1016/j.pain.2007.04.041 (2007).
113 Meloto, C. B. *et al.* Human pain genetics database: a resource dedicated to human pain genetics research. *Pain* **159**, 749-763, doi:10.1097/j.pain.0000000000001135 (2018).
114 Freilinger, T. *et al.* Genome-wide association analysis identifies susceptibility loci for migraine without aura. *Nat Genet* **44**, 777-782, doi:10.1038/ng.2307 (2012).
115 Cox, H. C. *et al.* A genome-wide analysis of 'Bounty' descendants implicates several novel variants in migraine susceptibility. *Neurogenetics* **13**, 261-266, doi:10.1007/s10048-012-0325-x (2012).
116 Chasman, D. I. *et al.* Genome-wide association study reveals three susceptibility loci for common migraine in the general population. *Nat Genet* **43**, 695-698, doi:10.1038/ng.856 (2011).
117 Anttila, V. *et al.* Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1. *Nat Genet* **42**, 869-873, doi:10.1038/ng.652 (2010).
118 Gormley, P. *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat Genet* **48**, 856-866, doi:10.1038/ng.3598 (2016).
119 Anttila, V. *et al.* Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat Genet* **45**, 912-917, doi:10.1038/ng.2676 (2013).
120 Ligthart, L. *et al.* Meta-analysis of genome-wide association for migraine in six population-based European cohorts. *Eur J Hum Genet* **19**, 901-907, doi:10.1038/ejhg.2011.48 (2011).
121 Peters, M. J. *et al.* in *Ann Rheum Dis* Vol. 72 427-436 (2013).
122 Meng, W. *et al.* A Genome-wide Association Study Provides Evidence of Sex-specific Involvement of Chr1p35.1 (ZSCAN20-TLR12P) and Chr8p23.1 (HMGB1P46) With Diabetic Neuropathic Pain. *EBioMedicine* **2**, 1386-1393, doi:10.1016/j.ebiom.2015.08.001 (2015).
123 Johnston, K. J. A. *et al.* Genome-wide association study of multisite chronic pain in UK Biobank. *PLoS Genet* **15**, e1008164, doi:10.1371/journal.pgen.1008164 (2019).
124 Docampo, E. *et al.* Genome-wide analysis of single nucleotide polymorphisms and copy number variants in fibromyalgia suggest a role for the central nervous system. *Pain* **155**, 1102-1109, doi:10.1016/j.pain.2014.02.016 (2014).
125 Sexton, J. E., Cox, J. J., Zhao, J. & Wood, J. N. The Genetics of Pain: Implications for Therapeutics. *Annu Rev Pharmacol Toxicol* **58**, 123-142, doi:10.1146/annurev-pharmtox-010617-052554 (2018).
126 Harrison, R. K. Phase II and phase III failures: 2013-2015. *Nat Rev Drug Discov* **15**, 817-818, doi:10.1038/nrd.2016.184 (2016).
127 Mogil, J. S. Animal models of pain: progress and challenges. *Nat Rev Neurosci* **10**, 283-294, doi:10.1038/nrn2606 (2009).
128 Gottschalk, A. *et al.* Amantadine, a N-methyl-D-aspartate receptor antagonist, does not enhance postoperative analgesia in women undergoing abdominal hysterectomy. *Anesth Analg* **93**, 192-196, doi:10.1097/00000539-200107000-00038 (2001).
129 Wallace, M. S. *et al.* A randomized, double-blind, placebo-controlled trial of a glycine antagonist in neuropathic pain. *Neurology* **59**, 1694-1700, doi:10.1212/01.wnl.0000036273.98213.34 (2002).
130 Wallace, M. S. *et al.* A multicenter, double-blind, randomized, placebo-controlled crossover evaluation of a short course of 4030W92 in patients with chronic neuropathic pain. *J Pain* **3**, 227-233 (2002).
131 Galer, B. S., Lee, D., Ma, T., Nagle, B. & Schlagheck, T. G. MorphiDex (morphine sulfate/dextromethorphan hydrobromide combination) in the treatment of chronic pain: three multicenter, randomized, double-blind, controlled clinical trials fail to demonstrate enhanced opioid analgesia or reduction in tolerance. *Pain* **115**, 284-295, doi:10.1016/j.pain.2005.03.004 (2005).
132 Gavva, N. R. *et al.* Pharmacological blockade of the vanilloid receptor TRPV1 elicits marked hyperthermia in humans. *Pain* **136**, 202-210, doi:10.1016/j.pain.2008.01.024 (2008).
133 Shackelford, S. *et al.* A randomized, double-blind, placebo-controlled trial of a selective COX-2 inhibitor, GW406381, in patients with postherpetic neuralgia. *J Pain* **10**, 654-660, doi:10.1016/j.jpain.2009.01.328 (2009).
134 Ostenfeld, T. *et al.* A randomized, controlled study to investigate the analgesic efficacy of single doses of the cannabinoid receptor-2 agonist GW842166, ibuprofen or placebo in patients with acute pain following third molar tooth extraction. *Clin J Pain* **27**, 668-676, doi:10.1097/AJP.0b013e318219799a (2011).
135 Tuveson, B., Leffler, A. S. & Hansson, P. Ondansetron, a 5HT3-antagonist, does not alter dynamic mechanical allodynia or spontaneous ongoing pain in peripheral neuropathy. *Clin J Pain* **27**, 323-329, doi:10.1097/AJP.0b013e31820215c5 (2011).
136 Kalliomaki, J. *et al.* A randomized, double-blind, placebo-controlled trial of a chemokine receptor 2 (CCR2) antagonist in posttraumatic neuralgia. *Pain* **154**, 761-767, doi:10.1016/j.pain.2013.02.003 (2013).

137 Martinez, V. *et al.* The efficacy of a glial inhibitor, minocycline, for preventing persistent pain after lumbar discectomy: a randomized, double-blind, controlled study. *Pain* **154**, 1197-1203, doi:10.1016/j.pain.2013.03.028 (2013).
138 Bratus-Neuenschwander, A. *et al.* Pain-Associated Transcriptome Changes in Synovium of Knee Osteoarthritis Patients. *Genes (Basel)* **9**, doi:10.3390/genes9070338 (2018).
139 Alvarado, S. *et al.* Peripheral nerve injury is accompanied by chronic transcriptome-wide changes in the mouse prefrontal cortex. *Mol Pain* **9**, 21, doi:10.1186/1744-8069-9-21 (2013).
140 Wang, H. *et al.* Chronic neuropathic pain is accompanied by global changes in gene expression and shares pathobiology with neurodegenerative diseases. *Neuroscience* **114**, 529-546 (2002).
141 Athie, M. C. P. *et al.* Transcriptome analysis of dorsal root ganglia's diabetic neuropathy reveals mechanisms involved in pain and regeneration. *Life Sci* **205**, 54-62, doi:10.1016/j.lfs.2018.05.016 (2018).
142 Sapio, M. R., Goswami, S. C., Gross, J. R., Mannes, A. J. & Iadarola, M. J. Transcriptomic analyses of genes and tissues in inherited sensory neuropathies. *Exp Neurol* **283**, 375-395, doi:10.1016/j.expneurol.2016.06.023 (2016).
143 Manteniotis, S. *et al.* Comprehensive RNA-Seq expression analysis of sensory ganglia with a focus on ion channels and GPCRs in Trigeminal ganglia. *PLoS One* **8**, e79523, doi:10.1371/journal.pone.0079523 (2013).
144 Flegel, C. *et al.* RNA-Seq Analysis of Human Trigeminal and Dorsal Root Ganglia with a Focus on Chemoreceptors. *PLoS One* **10**, e0128951, doi:10.1371/journal.pone.0128951 (2015).
145 Zhang, F. *et al.* Heat activation is intrinsic to the pore domain of TRPV1. *Proc Natl Acad Sci U S A* **115**, E317-e324, doi:10.1073/pnas.1717192115 (2018).
146 Yao, J., Liu, B. & Qin, F. Kinetic and energetic analysis of thermally activated TRPV1 channels. *Biophys J* **99**, 1743-1753, doi:10.1016/j.bpj.2010.07.022 (2010).
147 Lakshmi, S. & Joshi, P. G. Co-activation of P2Y2 receptor and TRPV channel by ATP: implications for ATP induced pain. *Cell Mol Neurobiol* **25**, 819-832, doi:10.1007/s10571-005-4936-8 (2005).
148 Caterina, M. J. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816-824, doi:10.1038/39807 (1997).
149 Smart, D. *et al.* The endogenous lipid anandamide is a full agonist at the human vanilloid receptor (hVR1). *Br J Pharmacol* **129**, 227-230, doi:10.1038/sj.bjp.0703050 (2000).
150 Caterina, M. J. *et al.* Impaired nociception and pain sensation in mice lacking the capsaicin receptor. *Science* **288**, 306-313, doi:10.1126/science.288.5464.306 (2000).
151 Davis, J. B. *et al.* Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia. *Nature* **405**, 183-187, doi:10.1038/35012076 (2000).
152 Marics, I., Malapert, P., Reynders, A., Gaillard, S. & Moqrich, A. Acute Heat-Evoked Temperature Sensation Is Impaired but Not Abolished in Mice Lacking TRPV1 and TRPV3 Channels. *PLoS One* **9**, doi:10.1371/journal.pone.0099828 (2014).
153 Nassini, R., Materazzi, S., Benemei, S. & Geppetti, P. The TRPA1 channel in inflammatory and neuropathic pain and migraine. *Rev Physiol Biochem Pharmacol* **167**, 1-43, doi:10.1007/112_2014_18 (2014).
154 Ranade, S. S., Syeda, R. & Patapoutian, A. Mechanically Activated Ion Channels. *Neuron* **87**, 1162-1179, doi:10.1016/j.neuron.2015.08.032 (2015).
155 Liu, B. & Qin, F. Single-residue molecular switch for high-temperature dependence of vanilloid receptor TRPV3. *Proc Natl Acad Sci U S A* **114**, 1589-1594, doi:10.1073/pnas.1615304114 (2017).
156 Sherkheli, M. A., Vogt-Eisele, A. K., Weber, K. & Hatt, H. Camphor modulates TRPV3 cation channels activity by interacting with critical pore-region cysteine residues. *Pak J Pharm Sci* **26**, 431-438 (2013).
157 Gao, L. *et al.* Selective potentiation of 2-APB-induced activation of TRPV1–3 channels by acid. *Sci Rep* **6**, doi:10.1038/srep20791 (2016).
158 Xiao, R. *et al.* Calcium plays a central role in the sensitization of TRPV3 channel to repetitive stimulations. *J Biol Chem* **283**, 6162-6174, doi:10.1074/jbc.M706535200 (2008).
159 Marwaha, L. *et al.* TRP channels: potential drug target for neuropathic pain. *Inflammopharmacology*, doi:10.1007/s10787-016-0288-x (2016).
160 Ji, R.-R., Xu, Z.-Z. & Gao, Y.-J. Emerging targets in neuroinflammation-driven chronic pain. *Nature Reviews Drug Discovery* **13**, 533-548, doi:doi:10.1038/nrd4334 (2014).
161 Basbaum, A. I., Bautista, D. M., Scherrer, G. & Julius, D. Cellular and molecular mechanisms of pain. *Cell* **139**, 267-284, doi:10.1016/j.cell.2009.09.028 (2009).
162 Zhang, X., Huang, J. & McNaughton, P. A. NGF rapidly increases membrane expression of TRPV1 heat-gated ion channels. *EMBO J* **24**, 4211-4223, doi:10.1038/sj.emboj.7600893 (2005).
163 Wang, C. *et al.* Pirt Together with TRPV1 Is Involved in the Regulation of Neuropathic Pain. *Neural Plast* **2018**, doi:10.1155/2018/4861491 (2018).
164 Akbar, A. *et al.* Increased capsaicin receptor TRPV1-expressing sensory fibres in irritable bowel syndrome and their correlation with abdominal pain. *Gut* **57**, 923-929, doi:10.1136/gut.2007.138982 (2008).

165 Zhu, Y. *et al.* Nerve growth factor modulates TRPV1 expression and function and mediates pain in chronic pancreatitis. *Gastroenterology* **141**, 370-377, doi:10.1053/j.gastro.2011.03.046 (2011).
166 Tympanidis, P. *et al.* Increased vanilloid receptor VR1 innervation in vulvodynia. *Eur J Pain* **8**, 129-133, doi:10.1016/s1090-3801(03)00085-5 (2004).
167 Facer, P. *et al.* Differential expression of the capsaicin receptor TRPV1 and related novel receptors TRPV3, TRPV4 and TRPM8 in normal human tissues and changes in traumatic and diabetic neuropathy. *BMC Neurol* **7**, 11, doi:10.1186/1471-2377-7-11 (2007).
168 Gopinath, P. *et al.* Increased capsaicin receptor TRPV1 in skin nerve fibres and related vanilloid receptors TRPV3 and TRPV4 in keratinocytes in human breast pain. *BMC Womens Health* **5**, 2, doi:10.1186/1472-6874-5-2 (2005).
169 Pang, Z. *et al.* Selective keratinocyte stimulation is sufficient to evoke nociception in mice. *Pain* **156**, 656-665, doi:10.1097/j.pain.0000000000000092 (2015).
170 Sukenaga, N. & Hirose, M. Transient Receptor Potential Channels as Blood Biomarkers for Pain Characteristics in Patients with Chronic Pain. *Anesth Essays Res* **12**, 279-281, doi:10.4103/aer.AER_179_17 (2018).
171 Yang, Y. S. *et al.* Increased expression of three types of transient receptor potential channels (TRPA1, TRPV4 and TRPV3) in burn scars with post-burn pruritus. *Acta Derm Venereol* **95**, 20-24, doi:10.2340/00015555-1858 (2015).
172 Moqrich, A. *et al.* Impaired thermosensation in mice lacking TRPV3, a heat and camphor sensor in the skin. *Science* **307**, 1468-1472, doi:10.1126/science.1108609 (2005).
173 Mandadi, S. *et al.* TRPV3 in keratinocytes transmits temperature information to sensory neurons via ATP. *Pflugers Arch* **458**, 1093-1102, doi:10.1007/s00424-009-0703-x (2009).
174 Peier, A. M. *et al.* A heat-sensitive TRP channel expressed in keratinocytes. *Science* **296**, 2046-2049, doi:10.1126/science.1073140 (2002).
175 Guedes, V., Castro, J. P. & Brito, I. Topical capsaicin for pain in osteoarthritis: A literature review. *Reumatol Clin* **14**, 40-45, doi:10.1016/j.reuma.2016.07.008 (2018).
176 Anand, P. & Bley, K. Topical capsaicin for pain management: therapeutic potential and mechanisms of action of the new high-concentration capsaicin 8% patch. *Br J Anaesth* **107**, 490-502, doi:10.1093/bja/aer260 (2011).
177 Garami, A. *et al.* TRPV1 antagonists that cause hypothermia, instead of hyperthermia, in rodents: Compounds' pharmacological profiles, in vivo targets, thermoeffectors recruited and implications for drug development. *Acta Physiol (Oxf)* **223**, e13038, doi:10.1111/apha.13038 (2018).
178 Rea, I. M. *et al.* Age and Age-Related Diseases: Role of Inflammation Triggers and Cytokines. *Front Immunol* **9**, 586, doi:10.3389/fimmu.2018.00586 (2018).
179 Khan, N. & Smith, M. T. Multiple sclerosis-induced neuropathic pain: pharmacological management and pathophysiological insights from rodent EAE models. *Inflammopharmacology* **22**, 1-22, doi:10.1007/s10787-013-0195-3 (2014).
180 O'Neill, T. W. & Felson, D. T. Mechanisms of Osteoarthritis (OA) Pain. *Curr Osteoporos Rep* **16**, 611-616, doi:10.1007/s11914-018-0477-1 (2018).
181 Grippo, A. J. & Scotti, M. A. Stress and neuroinflammation. *Mod Trends Pharmacopsychiatry* **28**, 20-32, doi:10.1159/000343965 (2013).
182 Pistell, P. J. *et al.* Cognitive impairment following high fat diet consumption is associated with brain inflammation. *J Neuroimmunol* **219**, 25-32, doi:10.1016/j.jneuroim.2009.11.010 (2010).
183 Valcarcel-Ares, M. N. *et al.* Obesity in Aging Exacerbates Neuroinflammation, Dysregulating Synaptic Function-Related Genes and Altering Eicosanoid Synthesis in the Mouse Hippocampus: Potential Role in Impaired Synaptic Plasticity and Cognitive Decline. *J Gerontol A Biol Sci Med Sci* **74**, 290-298, doi:10.1093/gerona/gly127 (2019).
184 Zajączkowska, R. *et al.* Mechanisms of Chemotherapy-Induced Peripheral Neuropathy. *Int J Mol Sci* **20**, doi:10.3390/ijms20061451 (2019).
185 Ji, R. R., Nackley, A., Huh, Y., Terrando, N. & Maixner, W. Neuroinflammation and central sensitization in chronic and widespread pain. *Anesthesiology* **129**, 343-366, doi:10.1097/aln.0000000000002130 (2018).
186 Jin, X. & Gereau, R. W. t. Acute p38-mediated modulation of tetrodotoxin-resistant sodium channels in mouse sensory neurons by tumor necrosis factor-alpha. *J Neurosci* **26**, 246-255, doi:10.1523/jneurosci.3858-05.2006 (2006).
187 Ji, R. R., Samad, T. A., Jin, S. X., Schmoll, R. & Woolf, C. J. p38 MAPK activation by NGF in primary sensory neurons after inflammation increases TRPV1 levels and maintains heat hyperalgesia. *Neuron* **36**, 57-68, doi:10.1016/s0896-6273(02)00908-x (2002).
188 Liang, L., Lutz, B. M., Bekker, A. & Tao, Y. X. Epigenetic regulation of chronic pain. *Epigenomics* **7**, 235-245, doi:10.2217/epi.14.75 (2015).
189 Descalzi, G. *et al.* Epigenetic Mechanisms of Chronic Pain. *Trends Neurosci* **38**, 237-246, doi:10.1016/j.tins.2015.02.001 (2015).

190 Matsushita, Y., Araki, K., Omotuyi, O., Mukae, T. & Ueda, H. HDAC inhibitors restore C-fibre sensitivity in experimental neuropathic pain model. *Br J Pharmacol* **170**, 991-998, doi:10.1111/bph.12366 (2013).
191 Laumet, G. *et al.* G9a is essential for epigenetic silencing of K(+) channel genes in acute-to-chronic pain transition. *Nat Neurosci* **18**, 1746-1755, doi:10.1038/nn.4165 (2015).
192 Garriga, J. *et al.* Nerve Injury-Induced Chronic Pain Is Associated with Persistent DNA Methylation Reprogramming in Dorsal Root Ganglion. *J Neurosci* **38**, 6090-6101, doi:10.1523/jneurosci.2616-17.2018 (2018).
193 Pollema-Mays, S. L., Centeno, M. V., Apkarian, A. V. & Martina, M. Expression of DNA methyltransferases in adult dorsal root ganglia is cell-type specific and up regulated in a rodent model of neuropathic pain. *Front Cell Neurosci* **8**, 217, doi:10.3389/fncel.2014.00217 (2014).
194 Manners, M. T., Ertel, A., Tian, Y. & Ajit, S. K. Genome-wide redistribution of MeCP2 in dorsal root ganglia after peripheral nerve injury. *Epigenetics Chromatin* **9**, 23, doi:10.1186/s13072-016-0073-5 (2016).
195 Hong, S., Zheng, G. & Wiley, J. W. Epigenetic regulation of genes that modulate chronic stress-induced visceral pain in the peripheral nervous system. *Gastroenterology* **148**, 148-157.e147, doi:10.1053/j.gastro.2014.09.032 (2015).
196 Crow, M. *et al.* HDAC4 is required for inflammation-associated thermal hypersensitivity. *Faseb j* **29**, 3370-3378, doi:10.1096/fj.14-264440 (2015).
197 Wang, F., Stefano, G. B. & Kream, R. M. Epigenetic modification of DRG neuronal gene expression subsequent to nerve injury: Etiological contribution to complex regional pain syndromes (Part II). *Med Sci Monit* **20**, 1188-1200, doi:10.12659/msm.890707 (2014).
198 Galbavy, W., Kaczocha, M., Puopolo, M., Liu, L. & Rebecchi, M. J. Neuroimmune and Neuropathic Responses of Spinal Cord and Dorsal Root Ganglia in Middle Age. *PLoS One* **10**, doi:10.1371/journal.pone.0134394 (2015).
199 Cruz-Almeida, Y. *et al.* Epigenetic aging is associated with clinical and experimental pain in community-dwelling older adults. *Mol Pain* **15**, 1744806919871819, doi:10.1177/1744806919871819 (2019).
200 Feng, G. *et al.* Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28**, 41-51, doi:10.1016/s0896-6273(00)00084-2 (2000).
201 Stephens, K. E. *et al.* Sex differences in gene regulation in the dorsal root ganglion after nerve injury. *BMC Genomics* **20**, 147, doi:10.1186/s12864-019-5512-9 (2019).
202 McInnes, I. B. *et al.* A novel therapeutic approach targeting articular inflammation using the filarial nematode-derived phosphorylcholine-containing glycoprotein ES-62. *J Immunol* **171**, 2127-2133, doi:10.4049/jimmunol.171.4.2127 (2003).
203 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006, doi:10.1101/gr.229102 (2002).
204 Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res* **40**, e115, doi:10.1093/nar/gks596 (2012).
205 Wang, X., Spandidos, A., Wang, H. & Seed, B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res* **40**, D1144-1149, doi:10.1093/nar/gkr1013 (2012).
206 Owczarzy, R. *et al.* IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res* **36**, W163-169, doi:10.1093/nar/gkn198 (2008).
207 Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-1108 (2008).
208 Vrtacnik, P., Kos, S., Bustin, S. A., Marc, J. & Ostanek, B. Influence of trypsinization and alternative procedures for cell preparation before RNA extraction on RNA integrity. *Anal Biochem* **463**, 38-44, doi:10.1016/j.ab.2014.06.017 (2014).
209 Sampaio-Silva, F., Magalhães, T., Carvalho, F., Dinis-Oliveira, R. J. & Silvestre, R. Profiling of RNA Degradation for Estimation of Post Morterm Interval. *PLoS One* **8**, doi:10.1371/journal.pone.0056507 (2013).
210 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).
211 Platt, R. N., Vandewege, M. W. & Ray, D. A. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res* **26**, 25-43, doi:10.1007/s10577-017-9570-z (2018).
212 Quinn, J. P., Savage, A. L. & Bubb, V. J. Non-coding genetic variation shaping mental health. *Curr Opin Psychol* **27**, 18-24, doi:10.1016/j.copsyc.2018.07.006 (2019).
213 Warren, I. A. *et al.* Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. *Chromosome Res* **23**, 505-531, doi:10.1007/s10577-015-9493-5 (2015).
214 Horvath, V., Merenciano, M. & Gonzalez, J. Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response. *Trends Genet* **33**, 832-841, doi:10.1016/j.tig.2017.08.007 (2017).

215 Savage, A. L., Bubb, V. J., Breen, G. & Quinn, J. P. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC Evol Biol* **13**, 101, doi:10.1186/1471-2148-13-101 (2013).
216 Gianfrancesco, O. *et al.* The Role of SINE-VNTR-Alu (SVA) Retrotransposons in Shaping the Human Genome. *Int J Mol Sci* **20**, doi:10.3390/ijms20235977 (2019).
217 Trizzino, M. *et al.* Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res* **27**, 1623-1633, doi:10.1101/gr.218149.116 (2017).
218 Tang , W. & Liang, P. Comparative genomics analysis reveals high levels of differential DNA transposition among primates. *bioRxiv* **520387**, doi:doi: https://doi.org/10.1101/520387 (2019).
219 Jjingo, D. *et al.* Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. *Mob DNA* **5**, 14, doi:10.1186/1759-8753-5-14 (2014).
220 Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**, D794-801, doi:10.1093/nar/gkx1081 (2018).
221 Essletzbichler, P. *et al.* Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res* **24**, 2059-2065, doi:10.1101/gr.177220.114 (2014).
222 Kovalevich, J. & Langford, D. Considerations for the Use of SH-SY5Y Neuroblastoma Cells in Neurobiology. *Methods Mol Biol* **1078**, 9-21, doi:10.1007/978-1-62703-640-5_2 (2013).
223 Krishna, A. *et al.* Systems genomics evaluation of the SH-SY5Y neuroblastoma cell line as a model for Parkinson's disease. *BMC Genomics* **15**, 1154, doi:10.1186/1471-2164-15-1154 (2014).
224 Bylund, L., Kytola, S., Lui, W. O., Larsson, C. & Weber, G. Analysis of the cytogenetic stability of the human embryonal kidney cell line 293 by cytogenetic and STR profiling approaches. *Cytogenet Genome Res* **106**, 28-32, doi:10.1159/000078556 (2004).
225 Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun* **5**, 4767, doi:10.1038/ncomms5767 (2014).
226 Stepanenko, A. A. & Dmitrenko, V. V. HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene* **569**, 182-190, doi:10.1016/j.gene.2015.05.065 (2015).
227 Binz, R. L. *et al.* Identification of novel breakpoints for locus- and region-specific translocations in 293 cells by molecular cytogenetics before and after irradiation. *Sci Rep* **9**, doi:10.1038/s41598-019-47002-0 (2019).
228 Rohn, G. *et al.* ACTB and SDHA Are Suitable Endogenous Reference Genes for Gene Expression Studies in Human Astrocytomas Using Quantitative RT-PCR. *Technol Cancer Res Treat* **17**, 1533033818802318, doi:10.1177/1533033818802318 (2018).
229 Long, Y., Wang, X., Youmans, D. T. & Cech, T. R. in *Sci Adv* Vol. 3 (2017).
230 Guo, T. *et al.* Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing. *Genome Biol* **19**, 170, doi:10.1186/s13059-018-1518-x (2018).
231 Zhu, S. *et al.* Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat Biotechnol* **34**, 1279-1286, doi:10.1038/nbt.3715 (2016).
232 Rakovic, A. *et al.* Genome editing in induced pluripotent stem cells rescues TAF1 levels in X-linked dystonia-parkinsonism. *Mov Disord* **33**, 1108-1118, doi:10.1002/mds.27441 (2018).
233 Westenberger, A. *et al.* A hexanucleotide repeat modifies expressivity of X-linked dystonia parkinsonism. *Ann Neurol* **85**, 812-822, doi:10.1002/ana.25488 (2019).
234 Allen, B., Pezone, A., Porcellini, A., Muller, M. T. & Masternak, M. M. Non-homologous end joining induced alterations in DNA methylation: A source of permanent epigenetic change. *Oncotarget* **8**, 40359-40372, doi:10.18632/oncotarget.16122 (2017).
235 Strichman-Almashanu, L. Z. *et al.* A Genome-Wide Screen for Normally Methylated Human CpG Islands That Can Identify Novel Imprinted Genes. *Genome Res* **12**, 543-554, doi:10.1101/gr.224102 (2002).
236 Okae, H. *et al.* Genome-Wide Analysis of DNA Methylation Dynamics during Early Human Development. *PLoS Genet* **10**, doi:10.1371/journal.pgen.1004868 (2014).
237 Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* **36**, 765-771, doi:10.1038/nbt.4192 (2018).
238 Hay, E. A. *et al.* An analysis of possible off target effects following CAS9/CRISPR targeted deletions of neuropeptide gene enhancers from the mouse genome. *Neuropeptides* **64**, 101-107, doi:10.1016/j.npep.2016.11.003 (2017).
239 Pontis, J. *et al.* Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell* **24**, 724-735.e725, doi:10.1016/j.stem.2019.03.012 (2019).
240 Trizzino, M., Kapusta, A. & Brown, C. D. Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genomics* **19**, doi:10.1186/s12864-018-4850-3 (2018).
241 Guimarães, M. Z. P. *et al.* Generation of iPSC-Derived Human Peripheral Sensory Neurons Releasing Substance P Elicited by TRPV1 Agonists. *Front Mol Neurosci* **11**, doi:10.3389/fnmol.2018.00277 (2018).

242 Witherspoon, D. J. *et al.* Mobile element scanning (ME-Scan) identifies thousands of novel Alu insertions in diverse human populations. *Genome Res* **23**, 1170-1181, doi:10.1101/gr.148973.112 (2013).
243 Lee, J. M. *et al.* CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology* **78**, 690-695, doi:10.1212/WNL.0b013e318249f683 (2012).
244 Gianfrancesco, O., Bubb, V. J. & Quinn, J. P. SVA retrotransposons as potential modulators of neuropeptide gene expression. *Neuropeptides* **64**, 3-7, doi:10.1016/j.npep.2016.09.006 (2017).
245 Binder, A. *et al.* Transient receptor potential channel polymorphisms are associated with the somatosensory function in neuropathic pain patients. *PLoS One* **6**, e17387, doi:10.1371/journal.pone.0017387 (2011).
246 Carreno, O. *et al.* SNP variants within the vanilloid TRPV1 and TRPV3 receptor genes are associated with migraine in the Spanish population. *Am J Med Genet B Neuropsychiatr Genet* **159b**, 94-103, doi:10.1002/ajmg.b.32007 (2012).
247 Okamoto, N. *et al.* Effect of single-nucleotide polymorphisms in TRPV1 on burning pain and capsaicin sensitivity in Japanese adults. *Mol Pain* **14**, doi:10.1177/1744806918804439 (2018).
248 Deering-Rice, C. E. *et al.* Characterization of Transient Receptor Potential Vanilloid-1 (TRPV1) Variant Activation by Coal Fly Ash Particles and Associations with Altered Transient Receptor Potential Ankyrin-1 (TRPA1) Expression and Asthma*. *J Biol Chem* **291**, 24866-24879, doi:10.1074/jbc.M116.746156 (2016).
249 Zhang, J. *et al.* Establishment of preliminary regulatory network of TRPV1 and related cytokines. *Saudi J Biol Sci* **24**, 582-588, doi:10.1016/j.sjbs.2017.01.029 (2017).
250 Wang, S., He, S., Yuan, F. & Zhu, X. Tagging SNP-set selection with maximum information based on linkage disequilibrium structure in genome-wide association studies. *Bioinformatics* **33**, 2078-2081, doi:10.1093/bioinformatics/btx151 (2017).
251 Quinn, J. P. & Bubb, V. J. SVA retrotransposons as modulators of gene expression. *Mob Genet Elements* **4**, e32102, doi:10.4161/mge.32102 (2014).
252 Makino, S. *et al.* Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *Am J Hum Genet* **80**, 393-406, doi:10.1086/512129 (2007).
253 Bell, J. T. *et al.* Differential methylation of the TRPA1 promoter in pain sensitivity. *Nat Commun* **5**, 2978, doi:10.1038/ncomms3978 (2014).
254 Paolucci, T., Saraceni, V. M. & Piccinini, G. Management of chronic pain in osteoporosis: challenges and solutions. *J Pain Res* **9**, 177-186, doi:10.2147/jpr.s83574 (2016).
255 Cline, M. E., Herman, J., Shaw, E. R. & Morton, R. D. Standardization of the visual analogue scale. *Nurs Res* **41**, 378-380 (1992).
256 Melzack, R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain* **1**, 277-299, doi:10.1016/0304-3959(75)90044-5 (1975).
257 Thiels, C. A., Habermann, E. B., Hooten, W. M. & Jeffery, M. M. Chronic use of tramadol after acute pain episode: cohort study. *BMJ* **365**, doi:10.1136/bmj.l1849 (2019).
258 Marincsak, R. *et al.* The analgesic drug, tramadol, acts as an agonist of the transient receptor potential vanilloid-1. *Anesth Analg* **106**, 1890-1896, doi:10.1213/ane.0b013e318172fefc (2008).
259 Gan, S. H., Ismail, R., Wan Adnan, W. A. & Zulmi, W. Impact of CYP2D6 genetic polymorphism on tramadol pharmacokinetics and pharmacodynamics. *Mol Diagn Ther* **11**, 171-181, doi:10.1007/bf03256239 (2007).
260 Kanaya, K., Iba, K., Dohke, T., Okazaki, S. & Yamashita, T. TRPV1, ASICs and P2X2/3 expressed in bone cells simultaneously regulate bone metabolic markers in ovariectomized mice. *J Musculoskelet Neuronal Interact* **16**, 145-151 (2016).
261 Rossi, F. *et al.* The genetic ablation or pharmacological inhibition of TRPV1 signalling is beneficial for the restoration of quiescent osteoclast activity in ovariectomized mice. *Br J Pharmacol* **171**, 2621-2630, doi:10.1111/bph.12542 (2014).
262 Yoshino, K. *et al.* Increase of TRPV1-Immunoreactivity in Dorsal Root Ganglia Neurons Innervating the Femur in a Rat Model of Osteoporosis. *Yonsei Med J* **55**, 1600-1605, doi:10.3349/ymj.2014.55.6.1600 (2014).
263 He, L. H. *et al.* TRPV1 deletion impaired fracture healing and inhibited osteoclast and osteoblast differentiation. *Sci Rep* **7**, 42385, doi:10.1038/srep42385 (2017).
264 Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* **43**, 8627-8637, doi:10.1093/nar/gkv862 (2015).
265 Ravichandran, S., Ahn, J. H. & Kim, K. K. Unraveling the Regulatory G-Quadruplex Puzzle: Lessons From Genome and Transcriptome-Wide Studies. *Front Genet* **10**, doi:10.3389/fgene.2019.01002 (2019).
266 Ostertag, E. M., Goodier, J. L., Zhang, Y. & Kazazian Jr, H. H. SVA Elements Are Nonautonomous Retrotransposons that Cause Disease in Humans. *Am J Hum Genet* **73**, 1444-1451 (2003).
267 Wildschutte, J. H., Baron, A., Diroff, N. M. & Kidd, J. M. Discovery and characterization of Alu repeat sequences via precise local read assembly. *Nucleic Acids Res* **43**, 10292-10307, doi:10.1093/nar/gkv1089 (2015).

268 Kryatova, M. S., Steranka, J. P., Burns, K. H. & Payer, L. M. Insertion and deletion polymorphisms of the ancient AluS family in the human genome. *Mob DNA* **8**, doi:10.1186/s13100-017-0089-9 (2017).
269 Keum, J. W. *et al.* The HTT CAG-Expansion Mutation Determines Age at Death but Not Disease Duration in Huntington Disease. *Am J Hum Genet* **98**, 287-298, doi:10.1016/j.ajhg.2015.12.018 (2016).
270 Lee, Y. M., Kim, Y. K. & Chung, J. H. Increased expression of TRPV1 channel in intrinsically aged and photoaged human skin in vivo. *Exp Dermatol* **18**, 431-436, doi:10.1111/j.1600-0625.2008.00806.x (2009).
271 Hu, Q. *et al.* TRPV1 Channel Contributes to the Behavioral Hypersensitivity in a Rat Model of Complex Regional Pain Syndrome Type 1. *Front Pharmacol* **10**, doi:10.3389/fphar.2019.00453 (2019).
272 Alswat, K. A. Gender Disparities in Osteoporosis. *J Clin Med Res* **9**, 382-387, doi:10.14740/jocmr2970w (2017).
273 Hong, E. P. & Park, J. W. Sample Size and Statistical Power Calculation in Genetic Association Studies. *Genomics Inf* **10**, 117-122, doi:10.5808/gi.2012.10.2.117 (2012).
274 Lin, J. *et al.* Stabilization of G-quadruplex DNA by C-5-methyl-cytosine in bcl-2 promoter: implications for epigenetic regulation. *Biochem Biophys Res Commun* **433**, 368-373, doi:10.1016/j.bbrc.2012.12.040 (2013).
275 Zamiri, B., Mirceta, M., Bomsztyk, K., Macgregor, R. B. & Pearson, C. E. Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC)8•(GGCCCC)8 repeat: effect of CpG methylation. *Nucleic Acids Res* **43**, 10055-10064, doi:10.1093/nar/gkv1008 (2015).
276 Sun, F. J. *et al.* Increased expression of TRPV1 in the cortex and hippocampus from patients with mesial temporal lobe epilepsy. *J Mol Neurosci* **49**, 182-193, doi:10.1007/s12031-012-9878-2 (2013).
277 Xiao, F. H., Wang, H. T. & Kong, Q. P. Dynamic DNA Methylation During Aging: A "Prophet" of Age-Related Outcomes. *Front Genet* **10**, 107, doi:10.3389/fgene.2019.00107 (2019).
278 Rhein, M. *et al.* DNA methylation results depend on DNA integrity—role of post mortem interval. *Front Genet* **6**, doi:10.3389/fgene.2015.00182 (2015).
279 Sjoholm, L. K., Ransome, Y., Ekstrom, T. J. & Karlsson, O. Evaluation of Post-Mortem Effects on Global Brain DNA Methylation and Hydroxymethylation. *Basic Clin Pharmacol Toxicol* **122**, 208-213, doi:10.1111/bcpt.12875 (2018).
280 Higgins, D. M. *et al.* Persistent pain and comorbidity among Operation Enduring Freedom/Operation Iraqi Freedom/operation New Dawn veterans. *Pain Med* **15**, 782-790, doi:10.1111/pme.12388 (2014).
281 Ramer, M. S. & Bisby, M. A. Normal and injury-induced sympathetic innervation of rat dorsal root ganglia increases with age. *J Comp Neurol* **394**, 38-47 (1998).
282 Hou, Y. *et al.* Ageing as a risk factor for neurodegenerative disease. *Nat Rev Neurol* **15**, 565-581, doi:10.1038/s41582-019-0244-7 (2019).
283 Vaughan, S. K., Stanley, O. L. & Valdez, G. Impact of Aging on Proprioceptive Sensory Neurons and Intrafusal Muscle Fibers in Mice. *J Gerontol A Biol Sci Med Sci* **72**, 771-779, doi:10.1093/gerona/glw175 (2017).
284 Lin, Y. T., Ro, L. S., Wang, H. L. & Chen, J. C. Up-regulation of dorsal root ganglia BDNF and trkB receptor in inflammatory pain: an in vivo and in vitro study. *J Neuroinflammation* **8**, 126, doi:10.1186/1742-2094-8-126 (2011).
285 Camdessanche, J. P. *et al.* Sensory and motor neuronopathy in a patient with the A382P TDP-43 mutation. *Orphanet J Rare Dis* **6**, 4, doi:10.1186/1750-1172-6-4 (2011).
286 Rossor, A. M., Jaunmuktane, Z., Rossor, M. N., Hoti, G. & Reilly, M. M. TDP43 pathology in the brain, spinal cord, and dorsal root ganglia of a patient with FOSMN. *Neurology* **92**, e951-956, doi:10.1212/wnl.0000000000007008 (2019).
287 Vaughan, S. K. *et al.* The ALS-inducing factors, TDP43(A315T) and SOD1(G93A), directly affect and sensitize sensory neurons to stress. *Sci Rep* **8**, 16582, doi:10.1038/s41598-018-34510-8 (2018).
288 Deere, K. C. *et al.* Obesity is a risk factor for musculoskeletal pain in adolescents: findings from a population-based cohort. *Pain* **153**, 1932-1938, doi:10.1016/j.pain.2012.06.006 (2012).
289 Hitt, H. C., McMillen, R. C., Thornton-Neaves, T., Koch, K. & Cosby, A. G. Comorbidity of obesity and pain in a general population: results from the Southern Pain Prevalence Study. *J Pain* **8**, 430-436, doi:10.1016/j.jpain.2006.12.003 (2007).
290 Moreno-Navarrete, J. M. *et al.* Neuroinflammation in obesity: circulating lipopolysaccharide-binding protein associates with brain structure and cognitive performance. *Int J Obes (Lond)* **41**, 1627-1635, doi:10.1038/ijo.2017.162 (2017).
291 Pineda, M. A., Lumb, F., Harnett, M. M. & Harnett, W. ES-62, a therapeutic anti-inflammatory agent evolved by the filarial nematode Acanthocheilonema viteae. *Mol Biochem Parasitol* **194**, 1-8, doi:10.1016/j.molbiopara.2014.03.003 (2014).
292 Janicova, L. *et al.* Testing small molecule analogues of the Acanthocheilonema viteae immunomodulator ES-62 against clinically relevant allergens. *Parasite Immunol* **38**, 340-351, doi:10.1111/pim.12322 (2016).

293 Rodgers, D. T., Pineda, M. A., Suckling, C. J., Harnett, W. & Harnett, M. M. Drug-like analogues of the parasitic worm-derived immunomodulator ES-62 are therapeutic in the MRL/Lpr model of systemic lupus erythematosus. *Lupus* **24**, 1437-1442, doi:10.1177/0961203315591031 (2015).
294 Liu, X. J. *et al.* TLR signaling adaptor protein MyD88 in primary sensory neurons contributes to persistent inflammatory and neuropathic pain and neuroinflammation. *Sci Rep* **6**, 28188, doi:10.1038/srep28188 (2016).
295 Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**, 611-622, doi:10.1373/clinchem.2008.112797 (2009).
296 Boda, E., Pini, A., Hoxha, E., Parolisi, R. & Tempia, F. Selection of reference genes for quantitative real-time RT-PCR studies in mouse brain. *J Mol Neurosci* **37**, 238-253, doi:10.1007/s12031-008-9128-9 (2009).
297 Cheung, T. T., Weston, M. K. & Wilson, M. J. Selection and evaluation of reference genes for analysis of mouse (Mus musculus) sex-dimorphic brain development. *PeerJ* **5**, doi:10.7717/peerj.2909 (2017).
298 Timaru-Kast, R., Herbig, E. L., Luh, C., Engelhard, K. & Thal, S. C. Influence of Age on Cerebral Housekeeping Gene Expression for Normalization of Quantitative Polymerase Chain Reaction after Acute Brain Injury in Mice. *J Neurotrauma* **32**, 1777-1788, doi:10.1089/neu.2014.3784 (2015).
299 Gong, H. *et al.* Evaluation of candidate reference genes for RT-qPCR studies in three metabolism related tissues of mice after caloric restriction. *Sci Rep* **6**, 38513, doi:10.1038/srep38513 (2016).
300 Sookdeo, A., Hepp, C. M., McClure, M. A. & Boissinot, S. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob DNA* **4**, 3, doi:10.1186/1759-8753-4-3 (2013).
301 Prior, S. *et al.* Densely ionizing radiation affects DNA methylation of selective LINE-1 elements1. *Environ Res* **150**, 470-481, doi:10.1016/j.envres.2016.06.043 (2016).
302 Crowe , J. *et al.* Parasitic worm product ES-62 promotes healthspan and lifespan in a mouse model of obesity-induced ageing. *bioRxiv* **622753**, doi:https://doi.org/10.1101/622753 (2019).
303 Lee, S., Lee, T. A., Song, S. J., Park, T. & Park, B. Hyperproduction of IL-6 caused by aberrant TDP-43 overexpression in high-fat diet-induced obese mice. *FEBS Lett* **589**, 1825-1831, doi:10.1016/j.febslet.2015.05.040 (2015).
304 Garcia-Lacarte, M., Milagro, F. I., Zulet, M. A., Martinez, J. A. & Mansego, M. L. LINE-1 methylation levels, a biomarker of weight loss in obese subjects, are influenced by dietary antioxidant capacity. *Redox Rep* **21**, 67-74, doi:10.1179/1351000215y.0000000029 (2016).
305 Jin, Y. & Hammell, M. Analysis of RNA-Seq Data Using TEtranscripts. *Methods Mol Biol* **1751**, 153-167, doi:10.1007/978-1-4939-7710-9_11 (2018).
306 Philippe, C. *et al.* Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *eLife* **5**, doi:10.7554/eLife.13926 (2016).
307 Tubio, J. M. C. *et al.* Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**, 1251343, doi:10.1126/science.1251343 (2014).
308 Gagnier, L., Belancio, V. P. & Mager, D. L. Mouse germ line mutations due to retrotransposon insertions. *Mob DNA* **10**, 15, doi:10.1186/s13100-019-0157-4 (2019).
309 Dahlhamer, J. *et al.* Prevalence of Chronic Pain and High-Impact Chronic Pain Among Adults — United States, 2016. *MMWR Morb Mortal Wkly Rep* **67**, 1001-1006, doi:10.15585/mmwr.mm6736a2 (2018).
310 Papasotiriou, I., Pantopikou, K. & Apostolou, P. L1 retrotransposon expression in circulating tumor cells. *PLoS One* **12**, e0171466, doi:10.1371/journal.pone.0171466 (2017).
311 Chaiwongwatanakul, S., Yanatatsaneejit, P., Tongsima, S., Mutirangura, A. & Boonyaratanakornkit, V. Sex Steroids Regulate Expression of Genes Containing Long Interspersed Elements-1s in Breast Cancer Cells. *Asian Pac J Cancer Prev* **17**, 4003-4007 (2016).
312 Huen, K. *et al.* DNA methylation of LINE-1 and Alu repetitive elements in relation to sex hormones and pubertal timing in Mexican-American children. *Pediatr Res* **79**, 855-862, doi:10.1038/pr.2016.31 (2016).
313 Hadziselimovic, F., Hadziselimovic, N. O., Demougin, P., Krey, G. & Oakeley, E. Piwi-pathway alteration induces LINE-1 transposon derepression and infertility development in cryptorchidism. *Sex Dev* **9**, 98-104, doi:10.1159/000375351 (2015).
314 Liu, C. C. *et al.* Interferon alpha inhibits spinal cord synaptic and nociceptive transmission via neuronal-glial interactions. *Sci Rep* **6**, 34356, doi:10.1038/srep34356 (2016).
315 Abdullah, A. *et al.* STING-mediated type-I interferons contribute to the neuroinflammatory process and detrimental effects following traumatic brain injury. *J Neuroinflammation* **15**, 323, doi:10.1186/s12974-018-1354-7 (2018).
316 McLachlan, E. M. & Hu, P. Inflammation in dorsal root ganglia after peripheral nerve injury: effects of the sympathetic innervation. *Auton Neurosci* **182**, 108-117, doi:10.1016/j.autneu.2013.12.009 (2014).
317 Sephton, C. F. *et al.* TDP-43 is a developmentally regulated protein essential for early embryonic development. *J Biol Chem* **285**, 6826-6834, doi:10.1074/jbc.M109.061846 (2010).
318 Kawahara, Y. & Mieda-Sato, A. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proc Natl Acad Sci U S A* **109**, 3347-3352, doi:10.1073/pnas.1112427109 (2012).

319 Gu, C. *et al.* Neuropilin-1 conveys semaphorin and VEGF signaling during neural and cardiovascular development. *Dev Cell* **5**, 45-57, doi:10.1016/s1534-5807(03)00169-2 (2003).

320 Spaulding, E. L. *et al.* Synaptic Deficits at Neuromuscular Junctions in Two Mouse Models of Charcot-Marie-Tooth Type 2d. *J Neurosci* **36**, 3254-3267, doi:10.1523/jneurosci.1762-15.2016 (2016).

321 Khalfallah, Y. *et al.* TDP-43 regulation of stress granule dynamics in neurodegenerative disease-relevant cell types. *Sci Rep* **8**, 7551, doi:10.1038/s41598-018-25767-0 (2018).

322 Romano , G., Klima , R., Profile, V. O. & Feiguin, F. Endogenous TDP-43 prevents retrotransposons activation through Dicer-2 activity and the RNA silencing machinery in Drosophila neurons. (2019).

323 Saldi, T. K. *et al.* The Caenorhabditis elegans Ortholog of TDP-43 Regulates the Chromatin Localization of the Heterochromatin Protein 1 Homolog HPL-2. *Mol Cell Biol* **38**, doi:10.1128/mcb.00668-17 (2018).

324 Sur, D. *et al.* Detection of the LINE-1 retrotransposon RNA-binding protein ORF1p in different anatomical regions of the human brain. *Mob DNA* **8**, doi:10.1186/s13100-017-0101-4 (2017).

325 Afshin, A. *et al.* Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *N Engl J Med* **377**, 13-27, doi:10.1056/NEJMoa1614362 (2017).

326 Guillemot-Legris, O. *et al.* High-fat diet feeding differentially affects the development of inflammation in the central nervous system. *J Neuroinflammation* **13**, 206, doi:10.1186/s12974-016-0666-8 (2016).

327 Doonan, J. *et al.* Failure of the Anti-Inflammatory Parasitic Worm Product ES-62 to Provide Protection in Mouse Models of Type I Diabetes, Multiple Sclerosis, and Inflammatory Bowel Disease. *Molecules* **23**, doi:10.3390/molecules23102669 (2018).

328 Suckling, C. J. *et al.* Small Molecule Analogues of the parasitic worm product ES-62 interact with the TIR domain of MyD88 to inhibit pro-inflammatory signalling. *Sci Rep* **8**, doi:10.1038/s41598-018-20388-z (2018).

329 Kfoury, A. *et al.* MyD88 in DNA Repair and Cancer Cell Resistance to Genotoxic Drugs. *J Natl Cancer Inst* **105**, 937-946, doi:10.1093/jnci/djt120 (2013).

330 Vermeij, W. *et al.* Diet restriction delays accelerated aging and genomic stress in DNA repair deficient mice. *Nature* **537**, 427-431, doi:10.1038/nature19329 (2016).

331 Gasior, S. L. & Deininger, P. L. ERCC1/XPF limits L1 retrotransposition. *DNA Repair (Amst)* **7**, 983-989, doi:10.1016/j.dnarep.2008.02.006 (2008).

332 Servant, G., Streva, V. A. & Deininger, P. L. Transcription coupled repair and biased insertion of human retrotransposon L1 in transcribed genes. *Mob DNA* **8**, 18, doi:10.1186/s13100-017-0100-5 (2017).

333 Teichmann, L. L., Schenten, D., Medzhitov, R., Kashgarian, M. & Shlomchik, M. J. Signals via the adaptor MyD88 in B cells and DCs make distinct and synergistic contributions to immune activation and tissue damage in lupus. *Immunity* **38**, 528-540, doi:10.1016/j.immuni.2012.11.017 (2013).

334 Ma, F. *et al.* Positive feedback regulation of type I IFN production by the IFN-inducible DNA sensor cGAS. *J Immunol* **194**, 1545-1554, doi:10.4049/jimmunol.1402066 (2015).

335 de Waard, M. C. *et al.* Age-related motor neuron degeneration in DNA repair-deficient Ercc1 mice. *Acta Neuropathol* **120**, 461-475, doi:10.1007/s00401-010-0715-9 (2010).

336 Swarup, V. *et al.* Deregulation of TDP-43 in amyotrophic lateral sclerosis triggers nuclear factor κB–mediated pathogenic pathways. *J Exp Med* **208**, 2429-2447, doi:10.1084/jem.20111313 (2011).

337 Elwood, E., Lim, Z., Naveed, H. & Galea, I. The effect of systemic inflammation on human brain barrier function. *Brain Behav Immun* **62**, 35-40, doi:10.1016/j.bbi.2016.10.020 (2017).

338 Landin, A. M. *et al.* Effects of fenbendazole on the murine humoral immune system. *J Am Assoc Lab Anim Sci* **48**, 251-257 (2009).

339 Kurnosov, A. A. *et al.* The Evidence for Increased L1 Activity in the Site of Human Adult Brain Neurogenesis. *PLoS One* **10**, doi:10.1371/journal.pone.0117854 (2015).

340 Singer, T., McConnell, M. J., Marchetto, M. C., Coufal, N. G. & Gage, F. H. LINE-1 Retrotransposons: Mediators of Somatic Variation in Neuronal Genomes? *Trends Neurosci* **33**, 345-354, doi:10.1016/j.tins.2010.04.001 (2010).

341 Rice, G. I. *et al.* Reverse-Transcriptase Inhibitors in the Aicardi-Goutieres Syndrome. *N Engl J Med* **379**, 2275-2277, doi:10.1056/NEJMc1810983 (2018).

342 Pereira, G. C. *et al.* Properties of LINE-1 proteins and repeat element expression in the context of amyotrophic lateral sclerosis. *Mob DNA* **9**, 35, doi:10.1186/s13100-018-0138-z (2018).

343 Prudencio, M. *et al.* Repetitive element transcripts are elevated in the brain of C9orf72 ALS/FTLD patients. *Hum Mol Genet* **26**, 3421-3431, doi:10.1093/hmg/ddx233 (2017).

344 Szollosi, A. G. *et al.* Activation of TRPV3 Regulates Inflammatory Actions of Human Epidermal Keratinocytes. *J Invest Dermatol* **138**, 365-374, doi:10.1016/j.jid.2017.07.852 (2018).

345 Doonan, J. *et al.* Protection Against Arthritis by the Parasitic Worm Product ES-62, and Its Drug-Like Small Molecule Analogues, Is Associated With Inhibition of Osteoclastogenesis. *Front Immunol* **9**, 1016, doi:10.3389/fimmu.2018.01016 (2018).

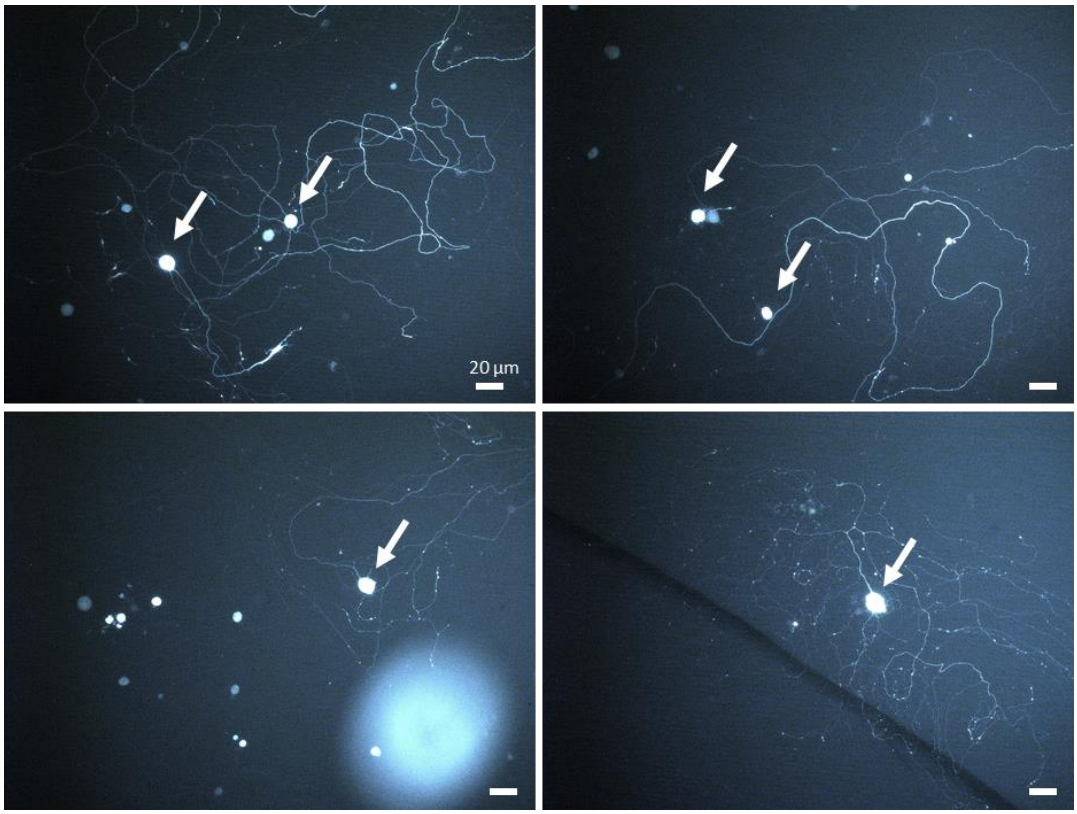
346 Surendran, S. & Bhatnagar, M. Upregulation of N-acetylaspartic acid induces oxidative stress to contribute in disease pathophysiology. *Int J Neurosci* **121**, 305-309, doi:10.3109/00207454.2011.558225 (2011).

347 Stincone, A. *et al.* The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biol Rev Camb Philos Soc* **90**, 927-963, doi:10.1111/brv.12140 (2015).
348 Itokazu, T., Hayano, Y., Takahashi, R. & Yamashita, T. Involvement of Wnt/beta-catenin signaling in the development of neuropathic pain. *Neurosci Res* **79**, 34-40, doi:10.1016/j.neures.2013.12.002 (2014).
349 Li, Y. *et al.* A novel ER-localized transmembrane protein, EMC6, interacts with RAB5A and regulates cell autophagy. *Autophagy* **9**, 150-163, doi:10.4161/auto.22742 (2013).
350 Berliocchi, L. *et al.* Spinal autophagy is differently modulated in distinct mouse models of neuropathic pain. *Mol Pain* **11**, doi:10.1186/1744-8069-11-3 (2015).
351 Alsalem, A. B., Halees, A. S., Anazi, S., Alshamekh, S. & Alkuraya, F. S. Autozygome Sequencing Expands the Horizon of Human Knockout Research and Provides Novel Insights into Human Phenotypic Variation. *PLoS Genet* **9**, doi:10.1371/journal.pgen.1004030 (2013).
352 Ley, K., Pramod, A. B., Croft, M., Ravichandran, K. S. & Ting, J. P. How Mouse Macrophages Sense What Is Going On. *Front Immunol* **7**, doi:10.3389/fimmu.2016.00204 (2016).
353 Joseph, E. K. & Levine, J. D. Caspase signalling in neuropathic and inflammatory pain in the rat. *Eur J Neurosci* **20**, 2896-2902, doi:10.1111/j.1460-9568.2004.03750.x (2004).
354 Sekiguchi, M. *et al.* Comparison of neuropathic pain and neuronal apoptosis following nerve root or spinal nerve compression. *Eur Spine J* **18**, 1978-1985, doi:10.1007/s00586-009-1064-z (2009).
355 Harrison, B. J. *et al.* IB4-binding sensory neurons in the adult rat express a novel 3' UTR-extended isoform of CaMK4 that is associated with its localization to axons. *J Comp Neurol* **522**, doi:10.1002/cne.23398 (2014).
356 Duncan, C. *et al.* Painful nerve injury decreases sarco-endoplasmic reticulum Ca(2)(+)-ATPase activity in axotomized sensory neurons. *Neuroscience* **231**, 247-257, doi:10.1016/j.neuroscience.2012.11.055 (2013).
357 Bezprozvanny, I. Calcium signaling and neurodegenerative diseases. *Trends Mol Med* **15**, 89-100, doi:10.1016/j.molmed.2009.01.001 (2009).
358 Chazin, W. J. Relating Form and Function of EF-hand Calcium Binding Proteins. *Acc Chem Res* **44**, 171-179, doi:10.1021/ar100110d (2011).
359 Zhang, M. D. *et al.* Neuronal calcium-binding proteins 1/2 localize to dorsal root ganglia and excitatory spinal neurons and are regulated by nerve injury. *Proc Natl Acad Sci U S A* **111**, E1149-1158, doi:10.1073/pnas.1402318111 (2014).

Supplementary data

Supplementary table 1. Associations between genes at chr17p13.2 with pain and inflammation.

Gene	Name	Comment	Reference
ASPA	Aspartocyclase	Deregulation causes accumulation of N-acetyl aspartate and nitric oxide synthase resulting in neurodegeneration. Cause of Canavan disease.	346
SHPK	Sedoheptulokinase	Newly identified protein in the PPP pathway, which is implicated in neurodegeneration.	347
TAX1BP3	Tax-1 binding protein	Overexpressed in osteoarthritis and plays a role in Wnt/B-catenin signalling which is deregulated in pain states	348
EMC6	ER membrane protein complex subunit 6	A novel regulator of autophagy, which is altered in experimental models of neuropathic pain	349,350
P2RX5	Purinergic receptor P2X ligand gated ion channel 5	Purinergic receptors are associated with neuropathic and inflammatory pain. This receptor is associated with ischemic pain resulting from exercise.	351
ITGAE	Integrin subunit alpha E	Role of integrins implicated in the maintenance of inflammatory and neuropathic hyperalgesia.	352,353
CAMKK1	Calcium/Calmodulin dependent protein kinase kinase 1	Role in regulating apoptosis, a process which is disrupted in DRG in association with neuropathic pain. Also plays a role in endosome trafficking in response to intracellular calcium signalling	354 355
ATP2A3	ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 3	Role in intracellular calcium homeostasis. Numerous links between calcium dyshomeostasis and neurological disease like Alzheimer's, Huntington's and Parkinson's disease have been established therefore the role of ATP2A3 and calcium homeostasis may also be relevant in chronic pain pathways.	356,357
ZZEF1	Zinc finger ZZ-type and EF-hand domain vontaining 1	Its function is unknown however the family of EF-hand Ca2+-binding proteins are one of the most common structural motifs in the cell and have a direct association with neurological disorders like Alzheimer's, diabetes and chronic inflammatory disorders. Other family members have an established role in pain pathways in DRG.	358,359



Supplementary figure 1. Examples of DRG cultures. DRG sensory neurons were purified from Thy1-YFP mice and maintained in culture for up to one month.

Supplementary table 2. Summary statistics for relative luciferase activity values in reporter gene assays conducted in HEK293 and DRG primary cells (Mann-Whitney U tests). Data assessed was $2^{-\Delta\Delta Ct}$ values.

HEK293

Construct	N	Mean	Median	SD	SEM	Difference	CI for difference	Achieved confidence	W-value	P-value
Empty	4	1.00	1.02	0.06	0.03					
Baseline	4	0.03	0.03	0.00	0.00	0.98	(0.88, 1.02)	96.96%	26.00	0.03
Empty	4	1.00	1.02	0.06	0.03					
SVA (endogenous)	4	0.48	0.46	0.05	0.03	0.55	(0.36, 0.62)	96.96%	26.00	0.03
Empty	4	1.00	1.02	0.06	0.03					
SVA (reverse)	4	0.62	0.62	0.07	0.03	0.35	(0.22, 0.49)	96.96%	26.00	0.03
SVA (endogenous)	4	0.48	0.46	0.05	0.03					
SVA (reverse)	4	0.62	0.62	0.07	0.03	-0.13	(-0.26, -0.00)	96.96%	10.00	0.03

DRG

Construct	N	Mean	Median	SD	SEM	Difference	CI for difference	Achieved confidence	W-value	P-value
Empty	4	1.00	0.99	0.04	0.02					
Baseline	4	0.50	0.49	0.02	0.01	0.50	(0.43, 0.58)	96.96%	26	0.03
Empty	4	1.00	0.99	0.04	0.02					
SVA (endogenous)	4	0.64	0.66	0.12	0.06	0.35	(0.21, 0.56)	96.96%	26	0.03

Supplementary table 3. Summary TRPV1 and TRPV3 expression differences in CRISPR modified HEK293 cell lines.

Sample	Mean Ct			TRPV1 ddCt						TRPV3 ddCt					
	ACTB	TRPV1	TRPV3	dCt	Control dCt	ddCt	$2^{-\Delta\Delta Ct}$	Mean	SEM	dCt	Control dCt	ddCt	$2^{-\Delta\Delta Ct}$	Mean	SEM
Unedited	21.04	31.85	32.81	-10.81	-10.89	0.41	0.75	0.80	0.03	-11.77	-11.85	0.10	0.93	0.97	0.05
Unedited	21.68	32.68	33.65	-11.00		0.22	0.86			-11.97		-0.10	1.08		
Unedited	21.87	32.73	33.60	-10.87		0.35	0.78			-11.74		0.13	0.91		
ntgRNA control	21.49	32.52	33.17	-11.03	-11.22	0.19	0.88	1.01	0.09	-11.67	-11.87	0.19	0.88	1.02	0.14
ntgRNA control	21.51	32.98	33.74	-11.46		-0.25	1.19			-12.23		-0.36	1.29		
ntgRNA control	21.57	32.73	33.26	-11.16		0.06	0.96			-11.69		0.17	0.89		
Edited (heterozygous)	22.83	33.68	35.78	-10.84		0.37	0.77	1.18	0.51	-12.94		-1.08	2.11	4.56	2.87
Edited (heterozygous)	23.31	35.66	38.54	-12.35		-1.13	2.19			-15.23		-3.36	10.29		
Edited (heterozygous)	24.25	34.67	36.47	-10.43		0.79	0.58			-12.22		-0.35	1.28		
Edited (homozygous)	23.25	34.89	27.57	-11.64		-0.43	1.34	1.12	0.12	-4.33		7.54	0.01	0.66	0.39
Edited (homozygous)	23.71	34.86	34.94	-11.15		0.07	0.95			-11.23		0.64	0.64		
Edited (homozygous)	23.00	34.29	35.29	-11.29		-0.08	1.05			-12.29		-0.43	1.35		

Supplementary table 4. Summary statistics (Mann-Whitney U tests) for relative expression analysis of L1 mRNA in ageing DRG Thy1-YFP mouse model. Data assessed was $2^{-\Delta\Delta Ct}$ values. Data assessed separately for male and female mice.

Females

Age	N	Mean	Median	SD	SEM	Difference	CI for difference	Achieved confidence	W-value	P-value
3	3	1.15	0.72	0.77	0.44	0.24	(0.09, 1.59)	91.91%	15.00	0.08
10	3	0.54	0.58	0.08	0.05					
3	3	1.15	0.72	0.77	0.44	-0.17	(-0.83, 1.17)	91.91%	9.00	0.66
12	3	1.12	0.99	0.35	0.20					
3	3	1.15	0.72	0.77	0.44	-14.00	(-24.62, -6.96)	91.91%	6.00	0.08
31	3	16.34	14.72	8.27	4.80					
3	3	1.15	0.72	0.77	0.44	-5.96	(-11.55, -4.39)	91.91%	6.00	0.08
42	3	8.44	6.68	3.29	1.90					
3	3	1.15	0.72	0.77	0.44	-11.16	(-13.45, -9.02)	91.91%	6.00	0.08
57	3	12.36	11.88	1.59	0.92					

Males

Age	N	Mean	Median	SD	SEM	Difference	CI for difference	Achieved confidence	W-value	P-value
3	3	1.02	1.04	0.25	0.12	0.18	(-0.20, 0.49)	94.82%	18.00	0.59
10	3	0.87	0.09	0.08	0.05					
3	4	1.02	1.04	0.25	0.12	-0.11	(-0.86, 0.48)	94.82%	15.00	0.86
12	3	1.16	1.08	0.41	0.23					
3	4	1.02	1.04	0.25	0.12	-12.86	(-14.10, -7.59)	94.82%	10.00	0.05
20	3	12.52	13.91	3.19	1.80					
3	4	1.02	1.04	0.25	0.12	-19.49	(-21.22, -12.66)	94.82%	10.00	0.05
35	3	18.81	20.53	4.27	2.50					
3	4	1.02	1.04	0.25	0.12	-12.38	(-23.57, -8.76)	94.82%	10.00	0.05
49	3	15.92	13.42	7.45	4.30					

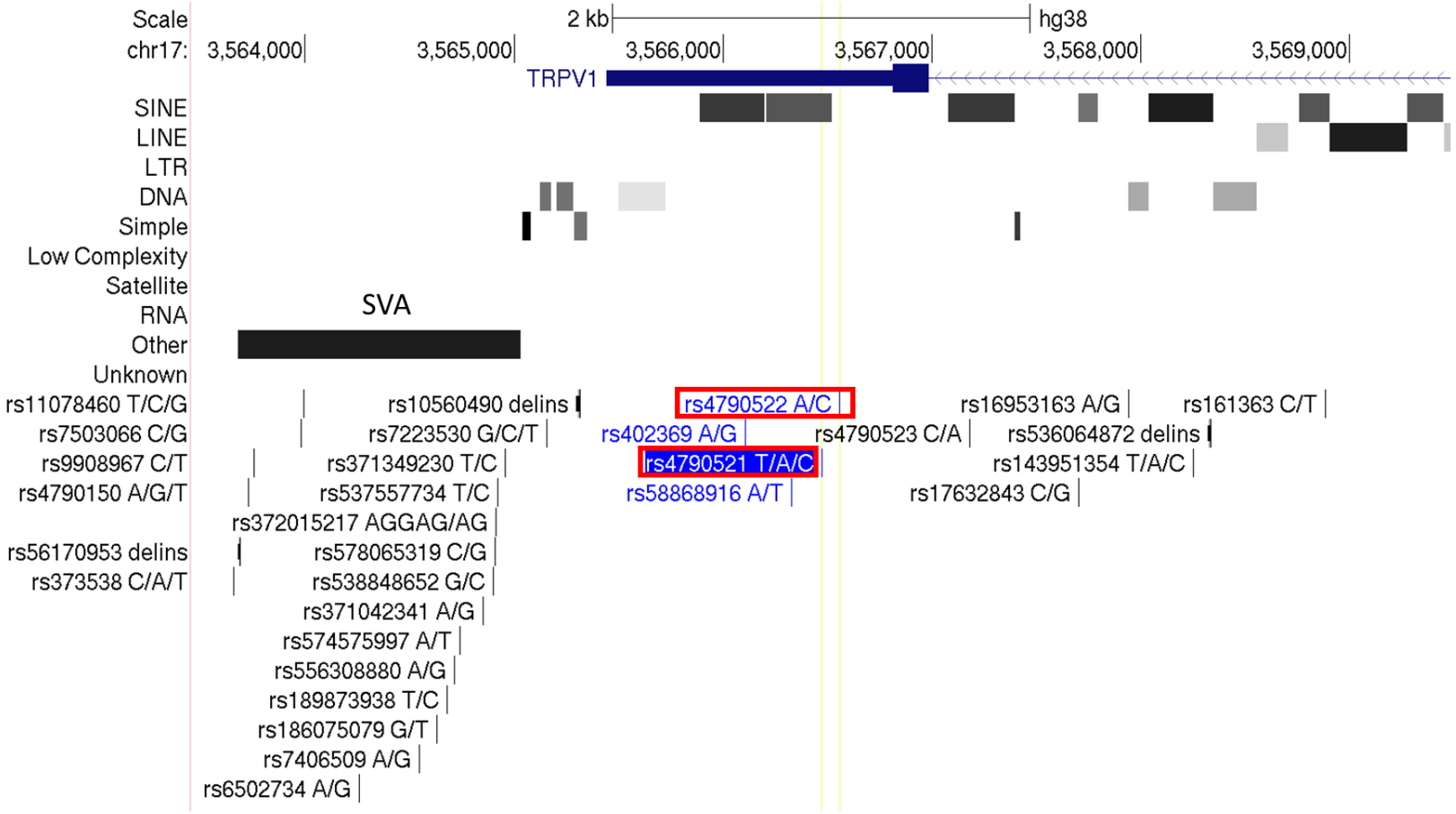
Supplementary table 5. Summary statistics for relative L1 and TDP-43 expression values in ageing brain C57BL/6 mouse model. Data assessed was $2^{-\Delta\Delta Ct}$ values.

LINE-1

Investigation	Sex	Group	Condition	N	Mean	Median	SD	SEM	Difference	CI for difference	Achieved confidence	W-value	P-value
Ageing	Male	Group A	8 weeks	6	1.29	1.16	0.91	0.37	-0.15	(-1.50, 1.38)	96.42%	32.00	0.52
		Group B	71 weeks	5	1.43	1.83	0.72	0.32					
Diet	Male	Group B	Chow	5	1.14	1.46	0.57	0.26	-0.15	(-1.16, 0.78)	96.42%	27.00	0.64
		Group C	High fat	6	1.26	1.07	0.58	0.24					
ES-62 treatment	Male	Group C	PBS	6	1.08	0.92	0.50	0.21	-0.77	(-7.92, 0.69)	96.42%	30.00	0.31
		Group D	ES-62	5	2.97	2.16	3.47	1.6					
	Female	Group C	PBS	5	1.37	1.70	0.84	0.38	-1.69	(-6.23, 0.97)	96.42%	23.00	0.23
		Group D	ES-62	6	3.55	2.94	2.78	1.1					

TDP-43

Investigation	Sex	Group	Condition	N	Mean	Median	SD	SEM	Difference	CI for difference	Achieved confidence	W-value	P-value
Ageing	Male	Group A	8 weeks	6	1.03	0.96	0.33	0.13	0.25	(-0.00, 0.73)	96.42%	47.50	0.04
		Group B	71 weeks	5	0.72	0.71	0.20	0.09					
Diet	Male	Group B	Chow	5	1.03	1.02	0.29	0.13	0.07	(-0.52, 0.50)	96.42%	31.00	0.92
		Group C	High fat	6	0.98	0.87	0.33	0.14					
ES-62 treatment	Male	Group C	PBS	6	1.04	0.92	0.35	0.15	-0.71	(-9.11, 0.61)	96.42%	32.00	0.52
		Group D	ES-62	5	3.05	1.86	3.98	1.8					
	Female	Group C	PBS	3	1.02	0.95	0.26	0.15	-0.73	(-2.21, 0.23)	96.31%	8.00	0.13
		Group D	ES-62	5	1.84	1.86	0.77	0.35					



Supplementary figure 2. Screenshot of SNPs in 3'UTR of *TRPV1* associated with changes in *TRPV1* mRNA expression. Screenshot of UCSC genome browser hg38 showing SNPs at the 3' UTR of *TRPV1* which are associated with mRNA expression differences[249]. SNPs are highlighted in red boxes. SVA is denoted.

Supplementary figure 3. DNA sequences used to identify polymorphism in CT-rich domain of SVA.

```
>allele_a_seq
TGTACATTCCTATGGTTGCATTCCTGCTATAATGCAGAAGGGAATAGTTGC
AACAGAGACCAAATAGTTTGCAAAGCCTAAAATATAAGCTTACCTAAAATA
TAAAAAGCTTACCTAAAATATAAAAAACATACTATCTGGCCCTTTACAGAA
GAGTTTGCTGACTCTTCCTTTGTAGAATAGAGCTTAAGAAACAGATACTCT
GCTCTCCCTCTCCCTCTCCCTCTCCCTCTCCCCATGGTCTCCCTCTCCCTC
TCTTT CCACGGTCTCCCTCTCATGCCGAGCCGAAGCTGGACTGTACTGA
>allele_b_ref_seq
TGTACATTCCTATGGTTGCATTCCTGCTATAATGCAGAAGGGAATAGTTG
CAACAGAGACCAAATAGTTTGCAAAGCCTAAAATATAAGCTTACCTAAAA
TATAAAAAGCTTACCTAAAATATAAAAAACATACTATCTGGCCCTTTACA
GAAGAGTTTGCTGACTCTTCCTTTGTAGAATAGAGCTTAAGAAACAGATA
CTCTACTCTGCTCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCT
CCCCATGGTCTCCCTCTCCCGATGGTCTCCCTCTCCCTCTCTTT CCACGG
TCTCCCTCTCATGCCGAGCCGAAGCTGGACTGTACTG
>allele_c_seq
TGTACATTCCTATGGTTGCATTCCTGCTATAATGCAGAAGGGAATAGTTG
CAACAGAGACCAAATAGTTTGCAAAGCCTAAAATATAAGCTTACCTAAAA
TATAAAAAGCTTACCTAAAATATAAAAAACATACTATCTGGCCCTTTACA
GAAGAGTTTGCTGACTCTTCCTTTGTAGAATAGAGCTTAAGAAACAGATA
CTCTGCTCTCCCTCTCCCTCTCCCTCTCCCCATGGTCTCCCTCTCCCCAT
GGTCTCCCTCTCCCGATGGTCTCCCTCTCCCGATGGTCTCCCTCTCCCTC
TCTTT CCACGGTCTCCCTCTCATGCCGAGCCGAAGCTGGACCTGTACTG
```

Supplementary table 6. QGRS sequences identified and used to predict G4 formation in full length SVA.

DNA strand	Position	End	Length	QGRS	G-Score
Antisense (5'-3')	213	241	28	GGAGACGGGTTTCGCTGTGTTGGCCGGG	12
	281	311	30	GGCCTCCCGAGGTGCCGGGATTGCAGACGG	17
	335	359	24	GGTGCCCAGGCTGGAGTGCAGTGG	15
	592	619	27	GGCCACGACCCCGTCTGGGAGGTGAGG	9
	730	741	11	GGAGGTGGGGG	21
	772	784	12	GGAGGTGAGGGG	18
	879	907	28	GGCCATGATGACAATGGCGGTTTTGTGG	9
	916	929	13	GGGGGAAAGGTGG	19
	1098	1119	21	GGTTAAATGGATTAAGGGCGG	17
	1218	1245	27	GGAAGGCCGCAGGGTCCTCTGCCTAGG	12
Sense (3'-5')	10	31	21	GGGAGAGGGAGAGGGAGAGGG	42
	46	76	30	GGAGAGAAAGGTGCCAGAGGGAGAGTACGG	20
	120	139	19	GGAGGGACGGACTAAGAGG	16
	141	160	19	GGAGTCGGACGGGTCACGG	20
	248	272	24	GGTCGAGGATTGGCGCTCACTAGG	15
	273	289	16	GGTCGGAGCCGGAGGG	19
	380	391	11	GGAGGTGGAGG	21
	397	413	16	GGACGGAACCGGAGGG	19
	431	448	17	GGAGACGGACCGGCGGG	19
	469	488	19	GGGAGACGGACCGACGGGG	15
	510	524	14	GGCGGGCCGGCGGG	21
	554	564	10	GGCGGCGGGG	20
	583	602	19	GGAGACGGGCCGGTGCTGG	20
	630	641	11	GGTCGGCGGGG	19
	658	668	10	GGGGAGGTGG	20
	714	725	11	GGTGGGGCAGG	19
	756	771	15	GGTCGGCGGGGCAGG	20
	786	804	18	GGAGACGGACCGGCGGGG	20
	830	842	12	GGCCGGTGGTGG	20
	1049	1063	14	GGAATGGGGGTTGG	19
	1224	1251	27	GGCGTCCCAGGAGACGGATCCTTTTGG	17
	1282	1310	28	GGAAGGGAGGTGATAGCAGGATACTGGG	19
	1312	1338	26	GGTTTAGGGGGAGACACTCTTTGTGG	9

Supplementary table 7. QGRS sequences identified and used to predict G4 formation in CT-rich alleles.

Allele	Start	Length	QGRS	G-Score
a	4	21	GGGAGAGGGAGAGGGAGAGGG	72
	29	38	GGGTACCAGAGGGAGAGGGAGAGAAAGGTGCCAGAGGG	59
b	10	21	GGGAGAGGGAGAGGGAGAGGG	72
	34	26	GGGAGAGGGGTACCAGAGGGAGAGGG	68
	68	27	GGGAGAGGGAGAGAAAGGTGCCAGAGG	34
c	4	21	GGGAGAGGGAGAGGGAGAGGG	72
	33	26	GGGAGAGGGGTACCAGAGGGAGAGGG	68
	67	26	GGGAGAGGGCTACCAGAGGGAGAGGG	67

Appendix A: Oligonucleotides and PCR cycling conditions

Entry	Application	Genome	Details	Forward 5′ – 3′	Reverse 5′ – 3′	gDNA (bp)	cDNA (bp)	Thermal cycles
1	Amplification of full length SVA for cloning and genotyping	hg19	SVA (full length) at TRPV1 and TRPV3 locus	GATACTCTACTCTGCTCTCCCT	AGATACTAACACAACCTGCCCA	1511	n/a	<u>Touchdown PCR</u> 95 °C 5m 95 °C 30s 65-55°c* 30s 72 °C 1m 35 cycles *decreases by 1 °C across first ten cycles
2	Amplification of ECR for cloning	hg19	ECR adjacent to SVA at TRPV1 and TRPV3 locus	ATTGAGCCATGACCACACC	AGGTTGTGTTAGTATCTGCGG	1196	n/a	
3	Amplification of SVA CT-rich domain Sequencing SVA alleles	hg19	SVA (CT-rich domain)	GCTTGTGGATGCGTGTGTAT	CAGTACAGTCCAGCTTCGG	424	n/a	<u>PCR</u> 95 °C 5m 95 °C 30s 63.6 °C 30s 72 °C 30s 72 °C 1m 35 cycles
4	Amplification of SVA VNTR	hg19	SVA (VNTR)	CCTCCCAAAGTGCAAAGC	CGTTCTCAATGAGCTGTTGG	472	n/a	<u>PCR</u> 95 °C 5m 95 °C 30s 55 °C 30s 72 °C 30s 72 °C 1m 25 cycles
5	Amplification of SVA poly A region	hg19	SVA (poly A region)	CCTATGACCCTGCCAAATCC	AGAGAACAGGGAGGAGGGAT	312	n/a	<u>PCR</u> 95 °C 5m 95 °C 30s

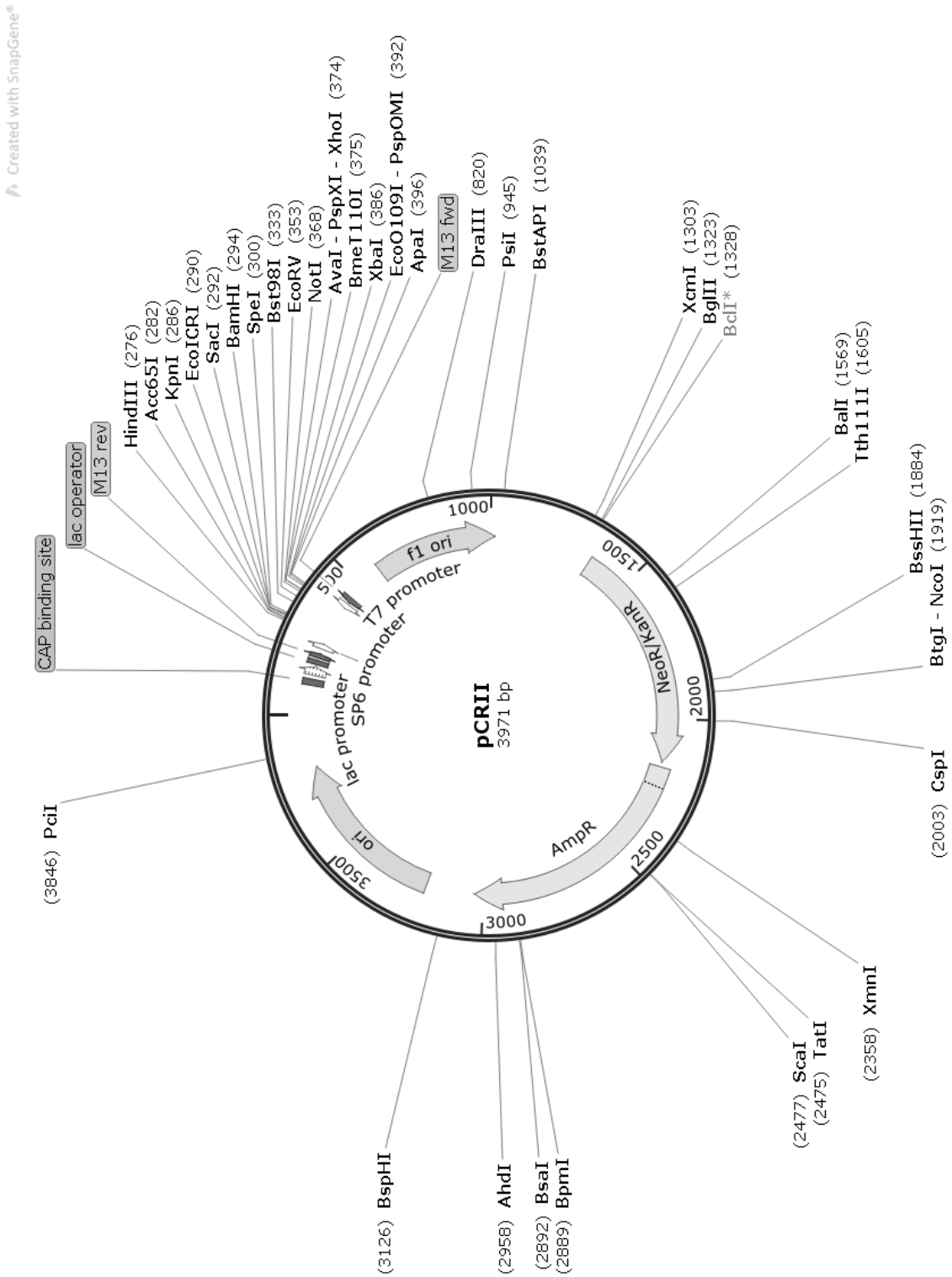
								65 °C 30s 72 °C 30s 72 °C 1m 35 cycles
6	Sequencing cloned inserts in pGL3-P	n/a	RV3 (forward read) GL2 (reverse read)	(RV3) CTAGCAAAATAGGCTGTCCC	(GL2) CTTTATGTTTTTGGCGTCTTCCA	n/a	n/a	n/a
7	CRISPR gRNAs	hg19	gRNA1	TTTTGCAAAAGTACAACCACAGG	n/a	n/a	n/a	n/a
			gRNA2	CAAATTTACTCAGAGCTGGAAGG	n/a	n/a	n/a	n/a
			gRNA3	CTTTCAAATTTACTCAGAGCTGG	n/a	n/a	n/a	n/a
			gRNA4	AGCTGCAAACTTCTGCAAATTGG	n/a	n/a	n/a	n/a
			gRNA5	CCAAAACACCTCTCCTGGTAGGG	n/a	n/a	n/a	n/a
			gRNA6	CCAAAACACCTCTCCTGGTAGGG	n/a	n/a	n/a	n/a
			gRNA7	CTAACACAACCTGCCCAGTCTGG	n/a	n/a	n/a	n/a
			gRNA8	CCCAAAACACCTCTCCTGGTAGG	n/a	n/a	n/a	n/a
			gRNA9	AATGGGCTGGCATAGCTAATTGG	n/a	n/a	n/a	n/a
			gRNA10	TACGTGTGTGTACATTCCTATGG	n/a	n/a	n/a	n/a
			gRNA11	ATAGCTAATTGGCTCTGTGTAGG	n/a	n/a	n/a	n/a
8	Sequencing cloned gRNAs in pSpCas9(BB)-2A-GFP	n/a	U6	GAGGGCCTATTTCCCATGATT	n/a	n/a	n/a	n/a
9	Genotyping modified HEK293 cell lines in CRISPR to assess SVA/ECR deletion	hg19	SVA presence/absence	AACGTGTTTACTGGCTGTGC	AGCAGTGCCTTCTTCATCCT	n/a	n/a	<u>PCR</u> 95 °C 5m 95 °C 30s 60 °C 30s 72 °C 1m 72 °C 5m 35 cycles
10	qPCR of total TRPV1 and total TRPV3 mRNA expression in CRISPR modified cell lines	hg19	TRPV1 (cDNA)	AGCAGTCGATTCTCTCTCTCTC	GCTCCATTTCTTCATCCTTGCT	17026	240	<u>qPCR</u> 95 °C 30s 95 °C 10s 60 °C 30s 72 °C 15s
		hg19	TRPV3 (cDNA)	ACGCACGTCTCCTTCCTTAA	CTCTGGGTTCCGCTTCTACA	740	159	
		hg19	ACTB (cDNA)	CACCTTCTACAATGAGCTGCGTGTG	ATAGCACAGCCTGGATAGCAAC GTAC	599	179	

								40 cycles
11	RT-PCR of TRPV1 and TRPV3 transcript variants in CRISPR modified HEK293 cell lines	hg19	TRPV1 ENST00000572705.1	AGCAGTCGTATTCTCTCTCTCTC	GCTCCATTTCTTCATCCTTGCT	17026	240	<u>Touchdown PCR</u> 95 °C 5m 95 °C 30s 65-55°c* 30s 72 °C 1m 35 cycles *decreases by 1 °C across first ten cycles
		hg19	TRPV1 ENST00000571088.5	AGACCACTCTTCTCCCACAC	GCTCCATTTCTTCATCCTTGCT	4683	149	
		hg19	TRPV1 ENST00000399759.7	GTGTGCAGTATAGATTCAGCGT	GCTCCATTTCTTCATCCTTGCT	3491	211	
		hg19	TRPV1 ENST00000399756.8	ACAGTCTGCTTGGCTCTTCT	GCTCCATTTCTTCATCCTTGCT	517	517	
		hg19	TRPV1 ENST00000425167.6	CTGGGACCGGGAAGATCG	CTCCACCAAGAGCATGTCGT	3927	260	
		hg19	TRPV1 ENST00000576351.5	CTGGGACCGGGAAGATCG	CTCCACCAAGAGCATGTCGT	3927	197	
		hg19	TRPV1 ENST00000310522.5	CTGGGACCGGGAAGATCG	CTCCACCAAGAGCATGTCGT	3927	47	
		hg19	TRPV3 ENST00000301365.8	ACGCACGTCTCCTTCCTTAA	GGACTCCACCATCCCTCAAA	821	243	
		hg19	TRPV3 ENST00000576742.5	ACGCACGTCTCCTTCCTTAA	CTCTGGGTTCCGCTTCTACA	740	159	
		hg19	TRPV3 ENST00000616411.4	ACGCACGTCTCCTTCCTTAA	CTTACATCCAGCCCCAAAGC	1505	927	
		hg19	TRPV3 ENST00000572519.1	ACGCACGTCTCCTTCCTTAA	CAGGCTGGTCTCAAACTCCT	355	355	
12	qPCR in mouse tissues	mm10	LINE-1 (cDNA)	CGTGATCCTAAGACCTCTAGTG	CAGTGGGCAGAGTATTCTC	124	124	<u>qPCR</u> 95 °C 30s 95 °C 10s 60 °C 30s 72 °C 15s 40 cycles
		mm10	TDP-43 (cDNA)	GCGATGGTGTGACTGTAAACT	TCTACCACTTCTCCATACTGAC	1384	147	
		mm10	ACTB (cDNA)	GCCAACCGTGAAAAGATGAC	GTACGACCAGAGGCATACAG	558	104	
		mm10	B2M (cDNA)	CATGGCTCGCTCGGTGAC	CAGTTCAGTATGTTCGGCTTCC	3203	135	
		mm10	GAPDH (cDNA)	AGGTCGGTGTGAACGGATTTG	TGACCTCAACTACATGGTCTACA	617	123	
		mm10	GUSB (cDNA)	CCGACCTCTCGAACAACCG	GCTTCCCGTTCATACCACACC	1833	169	
		mm10	HMBS (cDNA)	ATGAGGGTGATTCGAGTGGG	TTGTCTCCCGTGGTGGACATA	814	134	
		mm10	HPRT (cDNA)	TGACACTGGCAAAACAATGCA	GGTCCTTTTCACCAGCAAGCT	3769	95	
		mm10	PPIA (cDNA)	GGCAAATGCTGGACCAAAC	CATTCCTGGACCCAAAACG	341	149	

Appendix B: vector maps

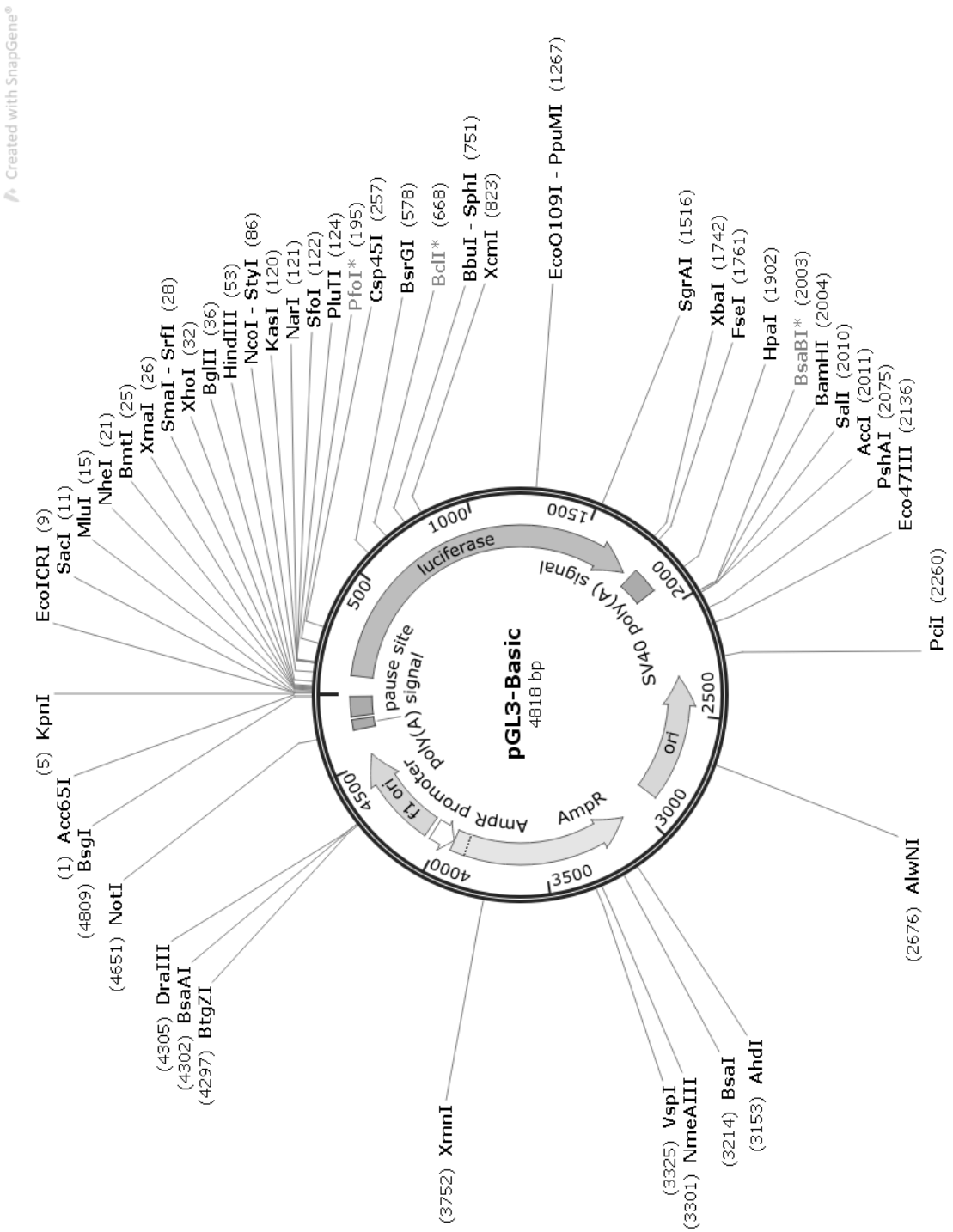
Vector map for pCRII

Sequence available at https://tools.thermofisher.com/content/sfs/vectors/pcrii_seq.txt



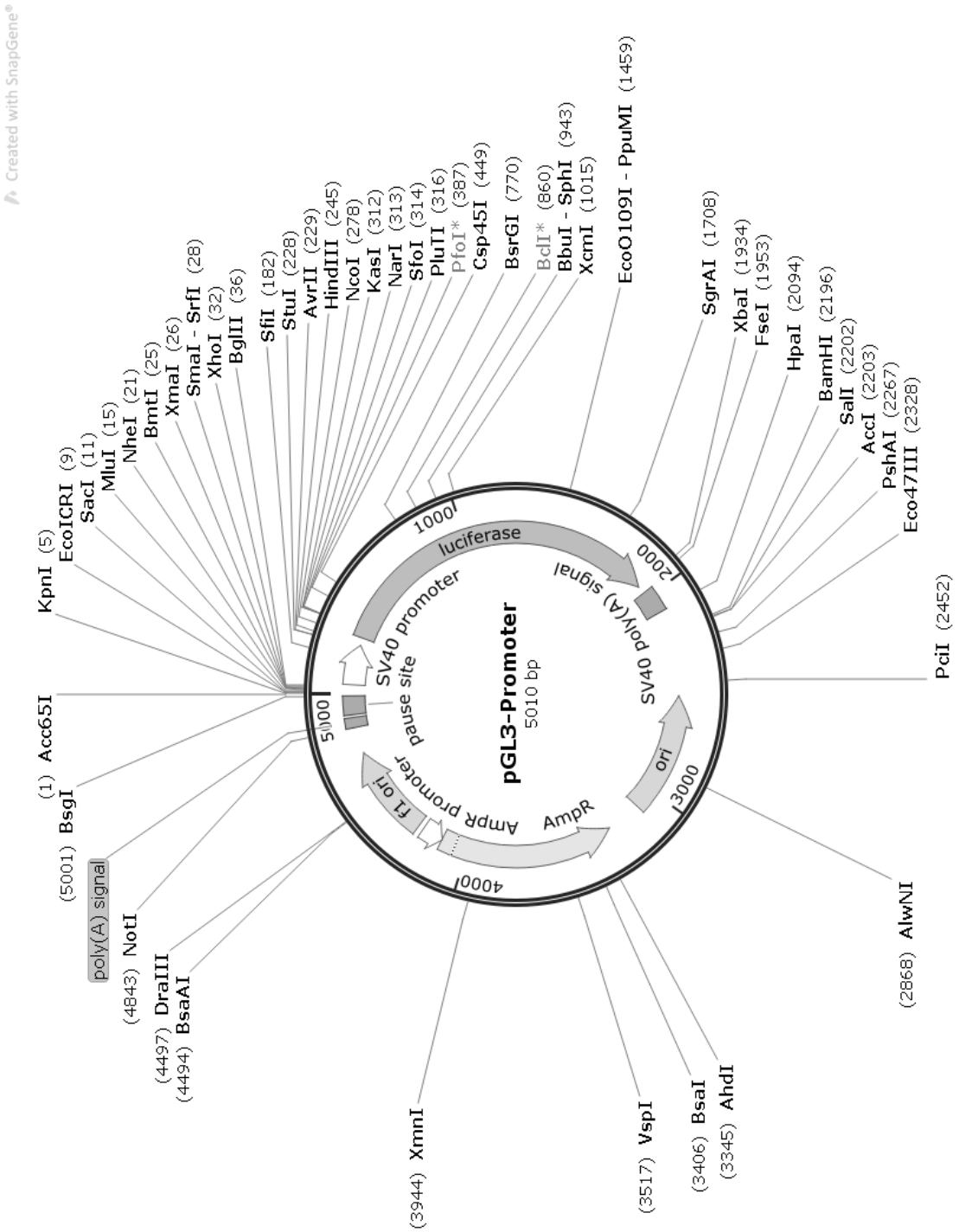
Vector map for pGL3-Basic

Sequence available at https://www.addgene.org/browse/sequence_vdb/2929/



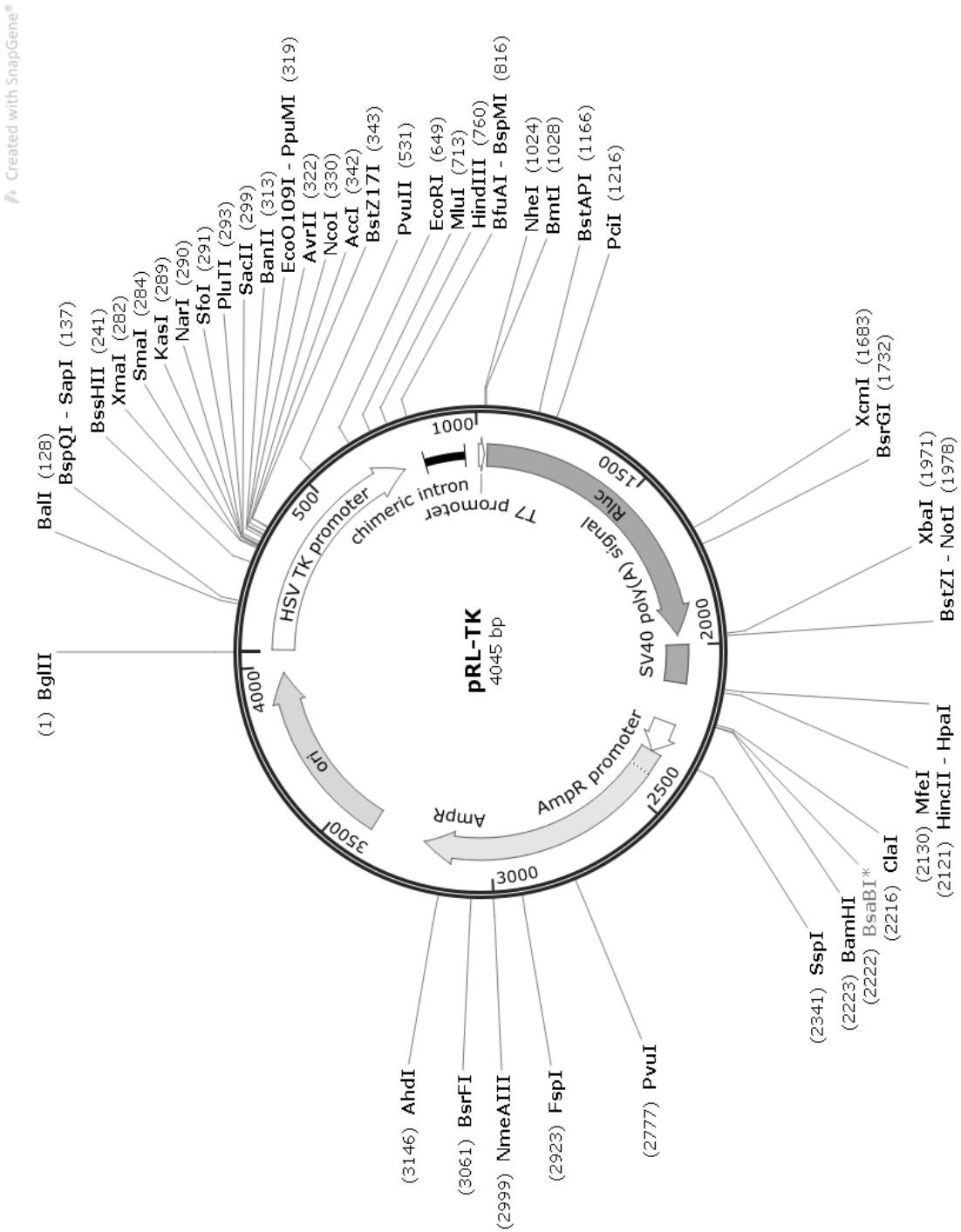
Vector map for pGL3-Promoter

Sequence available at https://www.addgene.org/browse/sequence_vdb/2932/



Vector map for pRL-TK

Sequence available at https://www.addgene.org/browse/sequence_vdb/3950/



Vector map for and pSpCas9(BB)-2A-GFP

Sequence available at https://www.addgene.org/48138/sequences/