UNIVERSITY OF LIVERPOOL

The role of Non-coding DNA in mental health

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

Veridiana Pessoa Miyajima

JANUARY 2023

Acknowledgements

I would like to express my sincerest gratitude to my supervisors, Professor John Quinn and Dr Vivien Jill Bubb, for supporting my Ph.D. research. I also recognise that it would not be possible without their patience, motivation, enthusiasm, and immense comprehension. Your guidance helped me with all obstacles in writing this thesis. I could not have imagined having better supervisors.

Special thanks to Dr Ursula Anderson, for the support, through Frances & Beatrice Anderson Memorial bursary; and for the encouragement and belief in the importance of research.

My sincerest thanks to Maurizio Manca, my 'lab partner', who shared many hours with me in the lab and always had a big smile. Thank you for keeping me hopeful in the grey days.

I am grateful for my many friends from Quinn's Lab, who shared precious time with me, Kate, Alix, Abi, Paul, Kejhal, Roshan, Paulina, Oly, Kim, Jack, Ben, Ana and Emma. Thank you for supporting me, sometimes with knowledge, advice, a chat, or a cup of coffee and a slice of cake to maintain our sanity. Thank you for the moments we spent together; I miss you all.

Also, I would like to thank the research team from the Wirral Child and Health Study (WCHADS), represented here by Helen Sharp and Dr Jonathan Hill and our many collaborators, Dr Lakis Louliglo, Dr Tony Payton, Dr Dan Rujescu.

I could not forget to acknowledge my Lab team and colleagues at Hemoce, especially from the Covid/CDC Labs and the genomic surveillance team I have been working with over these difficult years of the pandemic. You guys have been a great inspiration and good company.

Lastly, I would like to dedicate to my dearest family. Fabio, for constantly pushing me to the limit; Oliver for the tips and help with many tasks; Sarah for always being there to remind me that there is more in life than work; and Bob & Chico for the company. Nadja, my incredible mother, for the continuous support and assistance at difficult times; Sandra and Andre for believing this could be done. Juliana, my beloved baby sister (in memorian), and Valdemar, my wonderful father (*in memorian*) – Hey Dad, I have kept my promise!

Thank you, Eternal Father, for keeping me alive as his mercies endure forever. To God, all glory!

Abstract

Veridiana Pessoa Miyajima

The role of non-coding DNA in mental health

This PhD dissertation examines the role of non-coding DNA in mental health by looking at genetic variations at particular classes of repetitive elements, which were selected based on their potential functional role. One of the most crucial characteristics of non-coding DNA is its ability to regulate gene expression by multiple pathways, with the majority of them yet to be elucidated. Here we have selected genomic regions associated with neurological disorders, for which variable number tandem repeats (VNTRs) and a particular class of repetitive mobile elements (non-LTR retrotransposons) have been shown to account for substantial regulatory effects on gene regulation.

Retrotransposable elements have been demonstrated to act as expression modulators for multiple genes. Here we investigated the contribution of SINEs block within the TOMM40-ApoE-ApoC1 gene locus, particularly an intronic poly-T variant (rs10524523, '523') in TOMM40, and assessed the potential functional effects using reporter gene assays in a neuroblastoma cell line model. Examing the polymorphic poly-T stretch within the SINEs block, we observed that the shortest allele variants (polyT-17) were found to act as repressor regulators, as indicated by our reporter gene constructs with luciferase expression assays. We also observed that the shortest poly-T alleles were significantly associated with diminished language capacity and age-related changes compared to the longer poly-T variants. This supports our hypothesis that TOMM40 (and its clustered gene locus) is a critical regulator of human cognitive ageing and may be able to influence the rate of age-related cognitive decline.

Our second aim was to examine the potential of the uVNTR and dVNTR of the MAOA gene to drive and regulate the expression of the MAOA gene and their respective genetic variation domains. We carried out gene expression work of both VNTRs in a heterozygous neuroblastoma cell line and, using an engineered cells line with the single and double knockout of the VNTRs, compared the endogenous MAOA expression and correlated our findings to cohort samples. We found that the dVNTR was accountable for modulating the gene expression of two distinct coding isoforms differentially.

Finally, We also investigated the combination role of MAOA uVNTR and dVNTR in a cohort of schizophrenia patients and controls. . In the Schizophrenia cohort, we identified an association with suicide attempts in male carriers of the low-activity allele of the uVNTR. Overall, our findings warrant further investigation to the evidence on attributable risk alleles. Also we examined the genetic profiles of MAOA, SLC6A4 and DRD4 VNTRs, since these genes have been associated with behaviour and other neurodevelopmental disorders. We used samples from a child cohort, to analysed the genotypes of VNTRs located within the monoaminergic genes Our results indicated that both VNTRs of MAOA were in LD, and the distribution of genotypes observed agreed with the one reported by the literature.

Publications

*Indicates first co-authorship

- Manca, M.; **Pessoa, V***.; Myers, P.; Pickles, A.; Hill, J.; Sharp, H.; Murgatroyd, C.; Bubb, V.; Quinn, J.P. (2018). Distinct chromatin structures at the monoamine oxidase-A (MAOA) promoter correlate with allele specific expression in SH-SY5Y cells. Genes, Brain and Behavior, 2018, DOI: 10.1111/gbb.12483

- Manca, M., **Pessoa, V***., Lopez, A. I., Harrison, P. T., Miyajima, F., Sharp, H., & ... Quinn, J. P. (2018). The Regulation of Monoamine Oxidase A Gene Expression by Distinct Variable Number Tandem Repeats. Journal of Molecular Neuroscience: MN, doi:10.1007/s12031-018-1044-z

- Payton · P. Sindrewicz · **V. Pessoa** · H. Platt · M. Horan · W. Ollier · V.J. Bubb · N. Pendleton · J.P. Quinn. ”A TOMM40 poly-T variant modulates gene expression and is associated with vocabulary ability and decline in non-pathological ageing”. Neurobiology of aging 2015. doi:10.1016/j.neurobiolaging.2015.11.017

Abbreviations

5-HTT	5-hydroxy tryptamine (serotonin)
5-HTTLPR	Serotonin-transporter-linked polymorphic region
A	Adenine
AA	Amino Acid
AD	Alzheimer's disease
ADHD	Attention-Deficit/Hyperactivity Disorder
ALU	Arthrobacter luteus' retroelement
AluJ	Alu subfamily Jurka
AluS	Alu subfamily Smith
AluY	Alu subfamily Younger
AMP	Adenosine monophosphate
APOC	Apolipoprotein C
APOE	Apolipoprotein E
ATP	Adenosine-5'-triphosphate
AVP	Arginine vasopressin
BDNF	Brain-derived neurotrophic factor
BLAST	Basic Local Alignment Search Tool
BP	Bipolar disorder
Bp	Base pairs
BS	Binding site
C	Cytosine
C-	Carboxy-
CACNA1C	Calcium Channel, Voltage-Dependent, L Type, Alpha 1C Subunit
cAMP	Cyclic adenosine monophosphate
CAS9	CRISPR-associated protein 9
cDNA	Complementary deoxyribonucleic acid
ChIP	Chromatin immunoprecipitation
CI	Confidence interval
CNS	Central nervous system
CNV	Copy number variants
CoREST	Cofactor for REST
CpG	CG dinucleotides
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
CTCF	CCCTC binding-protein
d.f.	Degrees of freedom
dVNTR	Distal Variable Number Tandem Repeat
DA	Dopamine
DAT	Dopamine transporter
DKO	Double knockout
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DNMT1	DNA (cytosine-5)-methyltransferase 1
DNMTs	DNA methyltransferases
dNTP	Deoxynucleotide triphosphate
DRD4	Dopamine D4

ECR	Evolutionary conserved region
EDTA	Ethylenediaminetetraacetic acid
ENCODE	Encyclopaedia of DNA Elements
EZH2	Enhancer of zeste homolog 2
FAD	Flavin Adenine Dinucleotide
FLAM	Free left Alu monomer
FTD	Fronto Temporal Dementia
FXS	Fragile-X syndrome
G	Guanine
G4	G-quadruplexes
GABA	Gamma-aminobutyric acid
GRIN1	Glutamate receptor, ionotropic, NMDA 1
GTF	general transcription factor
GWAS	Genome wide association study
GxE	Gene-environment interaction
H2A/B	Histone H2 subunit A/B
H_2O_2	Hydrogen peroxide
H3	Histone H3
H3K27me3	Trimethylation of lysine 27 in histone H3
H3K4me2	Dimethylation of lysine 4 in histone H3
H3K9me1	Monomethylation of lysine 9 in histone H3
H3K9me3	Trimethylation of lysine 9 in histone H3
H4	Histone H4
HAP1	Near-haploid human cell line that was derived from KBM-7 cells
HD	Huntington's disease
HeLa	Henrietta Lacks clonal cell line
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hg	Human genome
Htt	Huntingtin
HWE	Hardy-Weinberg Equilibrium
In-del	Insertion - Deletion
K562	Human- derived chronic myelogenous leukaemia cell line
Kb	Kilobase
KBM-7	Human-derived chronic myelogenous leukaemia cell line
KO	Knockout
LB	Luria-bertani broth
LD	Linkage disequilibrium
LINE	Long interspersed nuclear elements
LTCC	L-type voltage-dependent calcium channel
LTR	Long terminal repeat
MAO	Monoamine Oxidase
MAOA	Monoamine oxidase A
MAOB	Monoamine oxidase B
MBD	m5CpG-binding domain
MDD	Major depressive disorder
MeCP2	Methyl-CpG-binding protein 2
MeDIP	Methylated DNA Immunoprecipitation
miRNA	microRNAs

mRNA	Messenger ribonucleic acid
N-	Amino-
NA	Noradrenaline
NaCl	Sodium chloride
NCBI	National Centre for Biotechnology Information
ncRNA	Non-coding RNA
NMDA	N-methyl-D-aspartate
NRSE	Neuron restrictive silencing element
NRSF	Neuron restrictive silencing factor
NSC	Neuronal stem cells
nt	Nucleotide
NTR	Number Tandem Repeat
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PD	Parkinson's disease
PIC	Protease inhibitor cocktail / pre-initiation complex
PKA	Protein kinase A
PKC	Protein kinase C
PLB	Passive lysis buffer
PLC	Phospholipase C
Pol	Polymerase II
POLR2A	RNA polymerase II
qPCR	Quantitative PCR
r^2	Squared correlation coefficient
RE1	Repressor element-1
REST	Repressor element-1 silencing transcription factor
RNA	Ribonucleic acid
RNAi	RNA interference
RT-PCR	Reverse transcriptase PCR
SCZ	Schizophrenia
SH-SY5Y	Human-derived neuroblastoma cells
SIN3A	SIN3 Transcription Regulator Family Member A
SINE	Short interspersed nuclear element
S-K-NAS	Human-derived neuroblastoma cells
SLC	Solute carrier gene family
SLC6A4	Solute carrier family 6 member 4
SNP	Single Nucleotide Polymorphism
SORT1	Sortilin 1
SP1	Specificity protein 1
SSR	Simple Sequence Repeats
STin2	Serotonin-transporter-intron-2
STR	Short Tandem Repeat
SVA	SINE-VNTR-Alu element
T	Thymine
TBE	Tris/Borate/EDTA
TBP	TATA-box-binding protein
TE	Tris-EDTA or transposable element
TF	Transcription Factor

TFBS	Transcription factor binding sites
TOMM40	Translocase of outer mitochondrial membrane 40
TR	Tandem repeat
TSG	Tumour Suppressive Gene
TSS	Transcriptional start site
U	Uracil
UCSC	University of California, Santa Cruz
UTR	Untranslated region
uVNTR	Upstream variable number tandem repeat
VNTR	Variable Number Tandem Repeat
VPA	Valproic Acid
XIST	X-inactive specific transcript gene
μ	Mu

Summary

List of Figures

List of Tables

.

Chapter 1

General Introduction

1.1 Thesis overview

This PhD dissertation examines the role of non-coding DNA in mental health by looking at the genetic variations in these repetitive elements. One of the most crucial characteristics of non-coding DNA is its ability to regulate gene expression by a multitude of pathways, with the majority yet to be clarified. Here we have the genomic regions associated with neurological disorders, for which variable number tandem repeats (VNTRs) and a particular class of repetitive mobile elements (non-LTR retrotransposons) have been shown to account for substantial regulatory effects on gene regulation.

First, by investigating the repetitive DNA elements from the retrotransposon class and their function as part of a clustered gene locus, in chapter 3. We focused on the Short Interspersed Nuclear Elements (SINEs), represented by Alu elements forming a SINEs block within the Translocase of Outer Mitochondrial Membrane 40 Homolog (*TOMM40*) gene, as well as SINE-VNTR-Alu (SVA) downstream to the Apolipoprotein C1 (APOC1) gene. These elements have been demonstrated to act as expression modulators for multiple genes. The TOMM40-APOE-APOC gene cluster has been linked with various age-related traits in humans, including non-pathological cognitive ageing, late-onset Alzheimer's disease, and longevity. For the former, in order to gain insight into their functional genomic significance, we examined a polymorphic poly-T stretch within the SINEs block, observing that the shortest allele variants was found to act as repressors in a stepwise manner, as indicated by our reporter gene constructs with luciferase expression assays. Furthermore, we independently observed that the shortest poly-T alleles were significantly associated with diminished language capacity and with age-related changes compared to the longer poly-T variants. This supports our hypothesis that TOMM40 (and its clustered gene locus) is a critical regulator of human cognitive ageing and may be able to influence the rate of age-related cognitive decline in a non-demented elderly cohort.

In chapter 4, we investigated the transcriptional regulation of variable number tandem repeats (VNTRs) located in established monoaminergic pathways, starting by looking at the upstream VNTR (uVNTR) and distal VNTR (dVNTR) of the Monoamine oxidase A (MAOA) gene. The dVNTR was particularly important to our understanding of the function of repetitive DNA in driving the expression of *MAOA*, an X-linked gene and a recognised behaviour-associated locus, where the newly uncovered dVNTR was examined and found to differentially modulate gene expression in our experimental work. We evaluated the concerted combination of these two VNTRs as part of a haplotype for functional genomics and genetic association investigations.

Finally, in chapter 5, we analysed a selection of five referenced VNTRs mapping neurological genes, whose roles are essential for the neurophysiological response to chronic stress and trauma. Monoamine Oxidase A (MAOA), Solute Carrier family 6 member 4 (SLC6A4), and Dopamine Receptor D4 (DRD4) were among these critical genes that also harbour relevant VNTRs. The relevance is due to growing evidence that these markers are predicted to alter physiological responses associated with neurodevelopment problems and various behavioural traits. In addition to evaluating the frequencies of the major and minor allele variants of each VNTR and build the haplotypes for the pairs in linkage disequilibrium (LD) in MAOA and SLC6A4, we also sought to build a panel containing the combinations of major and minor alleles for these VNTRs and estimate the constitution of blocks containing both 'risk' and 'non-risk' alleles according to the literature. This would allow inferring increased likelihood of carrying detrimental combinations of so-called 'risk' alleles, facilitating stratified analysis with behavioural conditions and endophenotypes, thus potentially increasing statistical power to detect true associations.

1.2 Non-coding DNA and the risk underlying mental health.

When the first draft of the human genome was finished in 2001, it surprisingly showed that only around 2% was coding DNA [1]. Therefore, the vast majority of the human genome is composed of non-coding DNA.

Since several studies have aimed to understand how the human genome has adapted to include both conserved stretches of regulatory non-coding DNA (ncDNA) present in many species and human-specific regulatory DNA elements that affect mRNA production. This combined effect of DNA elements enables genetic analysis to investigate the role of these elements, to understand the role in the genome, and the interaction with genes in several loci.

The human brain responds to typical developmental, psychological, and physiological cues and accounts for the challenges faced from prenatal to postnatal development [2-4]. These elements can hold gene mutations, which have been shown to change the regulatory features of the gene and are also linked to the likelihood of certain diseases [5, 6].

First, we must notice the most preserved regulatory areas in the genome and how they may be functionally changed by the most specific and well-researched kind of genetic variation, the single nucleotide polymorphisms (SNPs). This study focuses on human regulatory elements related to neurological conditions, such as variable number tandem repeats (VNTRs) and non-long terminal repeat (non-LTR) retrotransposons [7], which are significantly more substantial blocks of DNA variation.

1.3 Genetic variation in the genome (one size does not fit all).

Over fifty per cent of human DNA is repetitive: DNA whose sequence is the same or similar to other areas of the genome [8]. This polymorphism may happen in tandem and distribute in the form of transposable (mobile) elements, Figure 1.1, resulting in extensive structural diversity across the human genome [9]. This thesis will refer to

variations within the genome, which may differ in size or structure, but all aim to a better understanding of the function of DNA in non-coding regions.

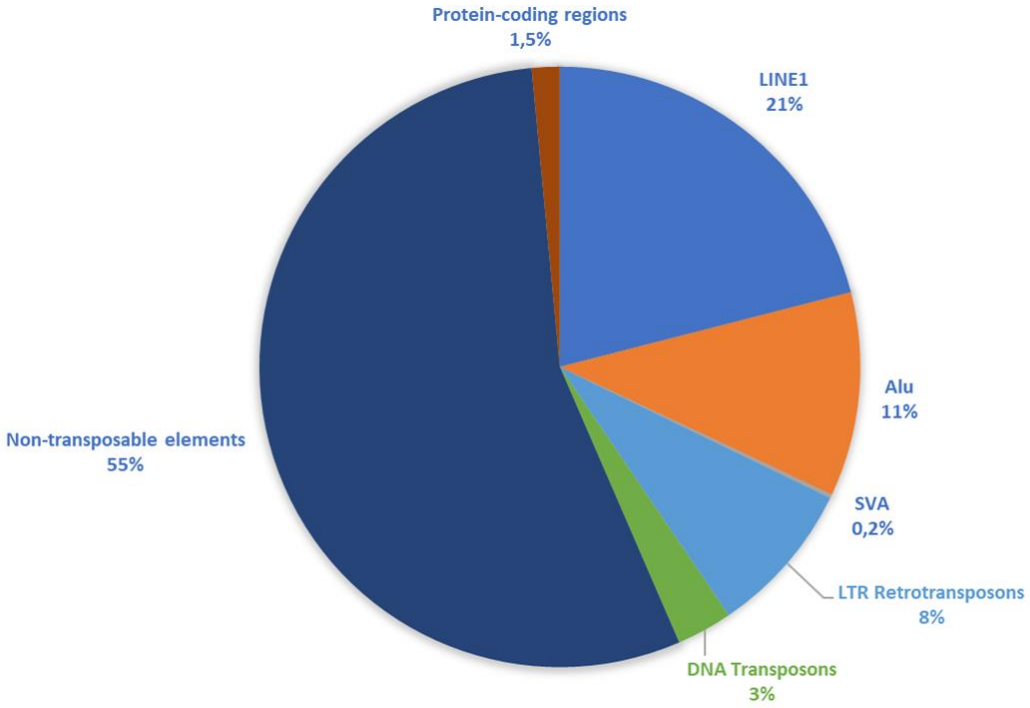


Figure 1.1. Human genome content in percentage (%). Transposable elements (TEs) compromise around 45 % of the human genome, and non-LTR retrotransposons are the most abundant, with Alu, LINE and SVA elements encompassing around 32%.

1.3.1 Single-nucleotide polymorphism (SNP)

The single-nucleotide polymorphism (SNP) is primarily assessed for genomic variation. It has been recognised that it can change the amino acid encoded in a coding exon and modify a protein. However, in non-coding DNA, SNP can act in several roles. For example, it changes transcription factors' affinity, specificity, or interaction since they are sequence-specific DNA-binding proteins. As a result, it can cause a change in the function of a specific DNA element. SNPs are abundant across the human genome. The latest count of validated human SNPs by NCBI (dbSNP build 155, 2021, https://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi) presented

a total of 1,053,623,523 SNPs (1,053 billion), the previous release from the 1000 Genomes Project set out to provide 84.7 million SNPs [10].

Nevertheless, the increase in comparing both releases was basically because of the addition of SNPs with very rare alleles. We have shifted our attention from focusing on DNA variation with the coding protein, where most prior investigations were centred. In genome-wide association studies (GWAS), a substantial number of SNP changes in non-coding DNA were seen to be related to behavioural or mental disorders due to the technique of exploring SNPs more widely as opposed to researching exons solely. [11, 12] Indeed, the GWAS analysis enabled the identification of genes linked to neuropsychiatric illnesses and revealed genetic commonality between the most prevalent mental disorders. [13, 14].

It is more challenging to validate the significance of SNPs in non-coding DNA regions than in coding regions. A regulatory domain, for instance, may be located hundreds of kilobases distant from the ncDNA tagging SNP. The SNP may not be the causal factor but highlights a chromosomal location. An SNP in ncDNA may tag a regulatory domain with hundreds of nucleotides distant from itself (a tagging SNP represents a large section of DNA inherited as one). In many instances, the discovered proteins may collaborate to influence a crucial process underlying well-being and mental health, such as calcium signalling. [6].

Several studies demonstrated the correlation of SNP polymorphisms in neuro disorders, including those associated with depression in BDNF, BICC1 and galanin genes [15-18]. Bioinformatic investigation of this locus using the Psychiatric Genomics Consortium GWAS dataset for schizophrenia revealed that five Evolutionary Conserved Regions (ECRs) in non-coding regions included or were next to genome-wide significant SNPs. [12]. Regulatory DNA often contains epigenetic markers that indicate whether chromatin is active or inactive. Genetic diversity, such as GWAS risk SNPs, might influence epigenetic parameters that affect long-term

regulatory changes in response to stressors. [19]. Epigenetic changes can be implicated in neuropsychiatric disorders. They can occur, including several genes or elements (global) or gene-specific (local) [20, 21]. This also has been confirmed by Nagy et al. showed the association between depression and methylation of astrocytes by global differences observed [22]. Another study has been associated with prenatal and postnatal depression and the local methylation variation at the glucocorticoid receptor gene by Murgatroyd et al. Simplistically, methylation of regulatory regions is considered a repressor of transcription as it interferes with transcription factor binding by limiting the accessibility of specific DNA recognition sequences [23]. The ability to rapidly address genetic variation on near regulatory domains has allowed the development of a significantly better understanding of how the ncDNA SNPs can be mechanistically involved in mental health issues [24].

1.3.2 Variable Number Tandem Repeats (VNTR)

Tandem repeats (TRs) are DNA sequence patterns that arise repeatedly and consecutively at a locus. These components comprise around 3 per cent of the human genome [25]. Variable number-tandem repeats (VNTRs) are DNA sequence patterns repeated in small DNA sequences termed motifs repeated in tandem several times. TRs are sometimes described as short tandem repeats (STRs), simple sequence repeats, and micro and minisatellites: microsatellites are regarded as components having repetitions of 1 to 9 base pairs in length, and minisatellites have around 10-60 base pair motifs [26] The repetitive nature of these sequences makes them susceptible to changes caused by strand slippage and uneven crossing over, which may increase or decrease motif copy number or introduce mutations to motifs [27] Variation in VNTR has been discovered to have an effect on mental health in several studies; related to behaviour [28-30] schizophrenia [31, 32]; and Alzheimer's disease [33]. In addition, there are many correlations between VNTR length and gene

expression [34, 35]. Despite expanding studies on the biological relevance of VNTR variation, our understanding of genetic variation in these sequences falls behind that of non-repetitive DNA. Therefore, further research is needed to understand how genetic variation in VNTR sequences affects mental health. This work will refer to these repeat elements as variable number tandem repeats (VNTRs).

1.3.3 Retrotransposon (RTE)

Transposable elements (TEs) comprise half of the human genome. They are DNA sequences that can move or transpose and are repetitive DNA elements that can mobilise to different locations within the genome [36, 37]. Barbara McClintock discovered TEs in maise in 1950, paving the way for detecting TEs in other organisms and accepting transposition as a widespread phenomenon [38]. TEs are split into two families based on the transposition strategy and intermediates formed during replication: retrotransposons (class I) and DNA transposons (class II), Figure 1.2 [39]. RTEs constitute approximately 42% of the human genome; they replicate through a copy-and-paste mechanism via an RNA intermediate, reverse-transcribed and re-inserted back into the genome [40, 41]. DNA transposons mobilise through a cut-and-paste mechanism via a DNA intermediate, facilitating insertion into new genomic locations [42]. Although initially branded “junk” DNA, TEs are now known to have roles in the regulation and evolution of the genome, as well as contributing to genetic instability and disease progression [42, 43]. Retrotransposons are divided into two major groups: long terminal repeat (LTR) elements, represented by the Human Endogenous Retroviruses (HERVs) and non-LTR sequences that include Long interspersed elements (LINEs), short interspersed elements (SINEs) and SINE-VNTR-Alu (SVAs).

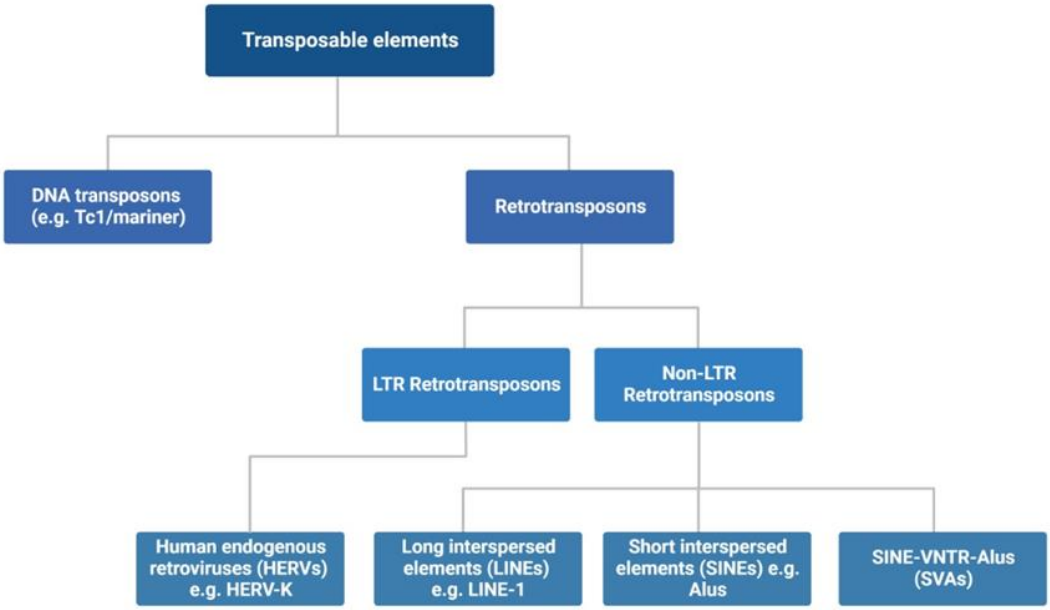


Figure 1.2. Transposable Elements (TE). Representation of the classification divided in retrotransposons (RTE) and DNA transposons. RTE are further divided: 1) long terminal repeat (LTR): HERVs. 2) non-LTR: LINEs, SINEs and SVAs. The figure was created with BioRender.com.

HERVs originated from ancient exogenous retroviruses that infected and incorporated themselves into the genomes of primates roughly 100 million years ago [44]. These provirus leftovers have circumvented host defence systems, incorporating and proliferating throughout the history of the human species [45]. HERVs are composed of 5' and 3' long terminal repeats that enclose the viral genes gag, pro, pol, and env [46, 47]. These genes encode capsid proteins, viral protease, reverse transcriptase, and envelope proteins sequentially. Approximately 8% of the human genome is composed of HERVs, and at least 31 HERV families have been found in humans, with HERV-K being the most recently acquired [45, 48]. HERV-K has been linked to many neurological disorders, such as multiple sclerosis, schizophrenia, bipolar disorder, and amyotrophic lateral sclerosis [49-52]. In addition, demonstrated to be transcriptionally active during embryogenesis and in other conditions [53].

The non-LTR retrotransposons are the only elements known to be mobile in the human genome. LINE-1, SINEs (Alus) and SVAs families can modify gene expression at transcriptional and post-transcriptional levels. They constitute 21%, 13%, and 0.2% of the human genome. LINE-1 is autonomous, as it encodes the ORF1 and ORF2 machinery necessary for mobilisation. Alu and SVA elements are non-autonomous and therefore require the machinery encoded by LINE-1 to transpose [54, 55]. In chapter three, we will focus on the non-LTR retrotransposons.

1.3.4 Long interspersed nuclear element-1 (LINE-1, L1)

LINE-1 comprises roughly 17% of the human genome and is present in over 500,000 copies; however, only around one hundred are active. The bulk of mobilisation events have been attributed to a small number, about 5 to 20% of the active L1; also, only about 40% of L1 are considered very active [56-59]. These 6Kb elements are retrotransposition competent and encode frame proteins: ORF1 is a 40 kDa protein, an RNA-binding protein, and ORF2 is a 150 kDa protein, which has endonuclease and reverses transcriptase activities. An antisense domain that this machinery is required for retrotransposition to occur through a process known as target-primed reverse transcription. Overexpression experiments show that it increases L1 mobility, implying a role in modulating and boosting L1 mobilisation [60]. The non-autonomous elements Alu and SVA can hijack this L1 machinery to translocate and re-insert themselves into the genome [61].

1.3.5 SINEs (Alus)

SINE (Short Interspersed Element) is a class of retrotransposons that constitutes approximately 11% of the human genome. They are primate-specific, with approximately 300 bp in length, and the most abundant mobile element in the human

genome, particularly *Alu* elements, representing more than a million times. It is estimated that *Alu* appeared 65 million years ago (mya), first with the lineage *AluJ* (Jurka), followed by *AluS* (Smith) around 30 mya, and the most recent one, *AluY* (Young), the only one reported to have the ability to move along the human genome [62-65].

Alu elements (SINEs) arose roughly 65 million years ago and are distinctive to primates. These retrotransposons are one of the most successful mobile elements, with over a million copies in humans and accounting for around eleven per cent of the human genome [62]. Alu elements are comprised of left and right monomers produced from the 7SL RNA gene, which is a component of signal recognition particle [66].

[62]. Larsen et al. suggested that Alus may induce neurodegenerative disorders by disturbing the mitochondria of neurons [67]. Chapter 3 demonstrates the involvement of the poly-T variant (rs10524523) part of an Alu element in an intronic region of the Translocase Outer Mitochondrial Membrane 40 Homolog (TOMM40) gene. This has been shown to interact with transcription machinery and differentially modulate reporter gene expression.

1.3.6 SINE-VNTR-Alu (SVA)

SVAs are the youngest non-LTRs. They are present only in primates, with the most recent subclass of the element being unique to humans, dating back between 18 and 25 million years ago (mya) [68, 69]. SVAs are composed of a 5’ hexamer (CCCTCT)n, also referred as CT element, an antisense Alu-like region, a central GC-rich VNTR, a SINE-R domain, and a 3’ poly A termination signal (poly A tail); usually, the length is around 0.7–4 kb in length (Figure 1.3). The SINE-R domain of these retroviral elements is homologous to the 3’ env gene of HERV-K10 [70], indicating their retroviral origin. The number of SVAs is between 2700 and 3000 copies of the human genome, which are divided into subclasses A–F based on their evolutionary age and

the content of the SINE-R domain. SVA-A is the oldest subclass, around 13.6 mya, while SVA-F is the youngest, which emerged at 3.2 mya [17]. An SVA-F1 subclass has also been found, which occurred from an alternative splicing event at the MAST2 locus in which exon 1 of the MAST2 gene was spliced into an intronic SVA element, which was then mobilised [71]. Genome-wide SVA analysis demonstrated that 34% of human-specific SVAs are formed by subclasses A to C, and 65 per cent belonged to the younger subclasses D to F1, from which SVA-Ds were the largest subclass, comprising 44 per cent of the overall SVA content in the human genome [55]. The newest SVA retrotransposons may be especially important for human health and evolution because they have inserted elements that are only found in humans and have the potential to control specific genomic loci [72].

1.4 The VNTRs as transcriptional regulators of gene expression.

VNTR variation can affect the regulation of gene expression. Many of the best-characterised genetic polymorphisms correlating with mental health issues are found in repetitive DNA. These include the VNTRs [26], examples of which have been identified in critical behavioural and mental health-related genes. VNTRs have been demonstrated to be both biomarkers and transcriptional regulators in genes such as the serotonin transporter, the dopamine transporter and monoamine oxidase A [20, 21, 29, 30, 73-76]. In these three examples, the primary DNA sequences of the VNTRs are rapidly evolving such that humans have their own specific VNTR sequences. All three of these monoaminergic genes contain a minimum of two VNTRs that have been demonstrated to act independently and synergistically as transcriptional regulators whose function is further modulated by the repeat copy number within the VNTR [20, 73, 75]. The copy number of the repeat itself is also a biomarker for good mental health and well-being, thus correlating function with phenotype [28-30, 75]. Perhaps not unexpectedly, VNTRs and GWAS SNPs in the

same promoter may act additively or synergistically to regulate gene expression. This is exemplified by one of the promoters of the schizophrenia candidate risk gene, MIR137 [31, 77], where experimentally in vitro, the VNTR in the promoter can support differential reporter gene expression based on the copy number of the repeat within the VNTR, and inclusion of the promoter region encompassing the GWAS SNP can further modulate expression depending on the allele of the SNP present. This illustrates a route to identifying the potential functional significance of non-coding variants in transcriptional or post-transcriptional regulatory mechanisms in areas distinct from the region of the DNA in which the GWAS SNP is found. The rapid evolution of VNTRs has been noted more globally for contributing to primate evolution; analysis in humans and non-human great apes identified that genes with VNTRs have higher expression divergence than those without [78]. The association of VNTRs with gene expression is reflected in the finding that VNTRs are enriched in promoter regions and locations close to transcriptional start sites for mRNA expression [79-81]. Generally, VNTRs have not been analysed as extensively as SNPs. This may be attributed to the requirement to perform PCR to genotype each VNTR target and the inability to accurately identify such regions in the initial short-read whole genome sequencing protocols. Improved depth and coverage in the entire genome sequence combined with the development of bioinformatic programmes such as Expansion Hunter may enhance the association of VNTRs and other repeat variants with neuropsychiatric conditions [82].

1.5 Non-coding DNA in gene regulation and function.

Gene regulation is essential for appropriate development. During development, genes are switched on and off in distinct ways, for example, by differentiating brain cells from liver cells. Moreover, gene regulation also helps cells adapt rapidly to environmental changes. Neurodevelopmental causes are often attributed to conditions such as

schizophrenia and autism [83, 84]. They influence the transcriptome throughout development. They were utilising a transgenic mouse model; early in vivo studies demonstrated how a human serotonin transporter VNTR may variably influence gene expression in crucial parts of the serotonergic lineage depending on the length of the tandem repeat [76, 85]. The co-expression of transcription factors will determine the regulation of regulatory domains during development, as illustrated by the schizophrenia-associated gene CACNA1C, whose expression in development mirrors that of the transcription factor EZH2, a critical regulator of the CACNA1C gene promoter [6]. An argument may be made that abnormalities in the transcriptome at certain moments in foetal development might result in a physical change in neuronal connections that would be more difficult to rectify than transcriptome changes in the adult [86]. Schork et al. find four unique genome-wide significant loci containing variations likely to influence genes expressed during mid-gestation in the developing neocortex. This period is supported by regulatory chromatin active throughout neurodevelopment in foetuses. The results indicated that common genetic variations induce dysregulation of genes which are essential in the regulation during neurodevelopment. It might be associated with a high risk of mental disorders later in life.[87]

Soldner et al. found a Parkinson's disease risk variation in a non-coding distal enhancer element that regulates the expression of α-synuclein (SNCA), a major gene involved in the pathophysiology of Parkinson's disease [88]. The regulation of gene expression can determine a myriad of functions. Several studies have demonstrated that non-coding DNA genetic variations are not only associated with the risk of a condition or disease but also has been confirmed that they can modify the regulatory properties of gene. In cancer, the non-coding alterations identify regulatory mutations that have been implicated in diagnostics, tumour progression, recurrence, and survival [89-95].

Since McClintock discovered transposable elements in maise, the idea that an organism's genome is not a permanent object but prone to change and rearrangement has been revolutionary. It has revolutionised how scientists see genetic patterns of inheritance. Initially, the hypothesis was received with opposition from the scientific community; nevertheless, the importance of transposons soon gained widespread acceptance, and the scientific community began to understand that genome replication did not always adhere to a uniform pattern.[38]

Transposable elements (TEs) play essential roles in the stability and evolution of the genome. They also contribute to human genetic variation and significantly affect human health through copy number variants, structural variants, insertions, deletions, and changes to gene transcription and splicing [96].

1.6 Retrotransposable elements within tomm40-apoe-apoc1 locus in the modulation of gene expression

As life expectancy rises, age-related cognitive impairment becomes a significant issue due to its large societal and economic impacts. Identifying genetic risk factors for cognitive ageing is becoming a significant research subject to avoid or reduce the course of cognitive decline. Recently, Genome-Wide Association Studies (GWAS) have highlighted the significance of the TOMM40-APOE locus in non-pathological cognitive decline and identified several SNPs, such as APOE rs429358 and TOMM40 rs11556505, as loci associated with cognitive ageing [97, 98]. Dementias are the most prevalent manifestation of abnormal cognitive ageing, with Alzheimer's disease (AD) being the most prevalent form. [97, 99]. In contrast to early-onset AD (EOAD), which is defined by a genetic mutation in one of the following three genes: AβPP, PSEN-1or PSEN-2, late-onset AD (LOAD) is more sporadic [97]. The APOE 4 allele of the Apolipoprotein E (*APOE*) gene is the most significant genetic factor in LOAD. [99]. APOE is in a gene-dense region of chromosome 19, between TOMM40 and

APOC1 genes. The association of this particular locus with AD has been well established [98, 100]. In addition to APOE, genetic polymorphism within the adjacent gene, TOMM40, was reportedly involved in the predisposition to AD [101].

Since the TOMM40 gene is 2 kb away from APOE and they share the same linkage disequilibrium block [102], it is hard to find separate functional areas that make people more likely to get Alzheimer's.

Roses and colleagues have shown that the TOMM40 L variation is strongly linked to APOE4 and that the S and VL variants are strongly linked to APOE3. The longer TOMM40 variations have been associated with an increased risk of developing late-onset Alzheimer's disease (LOAD) and an earlier age of disease onset in APOE3 individuals compared to the shorter form [103, 104]. Other studies, however, have either not seen this connection or claimed that the VL variation is linked with a decreased risk of AD [105] [106]. Another study found that the VL variation was related to a reduced risk of Alzheimer's disease, whereas the L variant was associated with an increased risk [107]. This same study discovered that the L variation drastically decreased the likelihood of reaching the age of 100. Several studies have also linked the "523" S variation to enhanced memory and executive function [108-111]. Bernardi et al. found that the L variation is related to an elevated risk of Alzheimer's disease and cognitive impairment among APOE e4 carriers. At the same time, the VL variation is linked to an earlier illness start among APOE e3/e3 carriers [112].

As an intronic gene variant, it is very doubtful that poly-T variants may disturb protein folding that substantially affects the stability of the TOM complex. Nonetheless, the examined literature revealed a strong relationship between the poly-T SNP and various kinds of brain and cognitive deterioration [109, 111, 113, 114]. This showed that more extended versions of 523 are linked to metabolic pathways that may make the brain more sensitive to environmental stressors.

1.6.1 Translocase of Outer Mitochondrial Membrane 40 Homolog (TOMM40)

The *TOMM40* gene is located on human chromosome 19q13.32 and encodes ten exons and nine introns and encodes a component of the Translocase of Mitochondrial Outer Membrane (TOM) complex [115]. TOM is a protein responsible for transporting and sorting proteins across the mitochondrial membrane and is required for energy metabolism, lipid synthesis, cell apoptosis and cellular homeostasis [116]. Genetic association studies have shown a substantial connection between genetic variations at multiple loci and the remarkable human lifespan [117].

The function of mitochondria impacts the performance of metabolic and signalling networks. The activity of mitochondria is tightly regulated, and mitochondrial dysfunction has been linked to disease pathogeneses. TOMM40 has been linked to Alzheimer's disease risk and non-pathological cognitive decline, possibly via the mitochondrial cascade hypothesis, which suggests a hereditary influence on mitochondrial durability and function [118]. Plaque and vascular amyloid deposition have been linked to SNPs in the TOMM40 gene [119]. Various TOMM40 SNPs have been related to non-pathological cognitive ageing [108, 120]; however, none of them has been shown to be functional, and it remains possible that they may be in linkage disequilibrium with another TOMM40-APOE variation.

Between exons 6 and 7 of TOMM40, SINEs, such as Alus, are quite abundant. Six adjacent SINEs (Alu) elements have inserted themselves across TOMM40 intron 6 with little homology identified in other primates but not in non-primate animals. Of particular interest: AluSx, AluYc3, AluJb, AluJo, another AluJb, and FLAM A, which contains the poly-T variation rs10524523 abbreviated as "523", which is an antisense Alu element. The fact that 523 is part of an Alu mobile element monomer inserted in the antisense orientation into TOMM40 intron six has mainly been overlooked as a significant aspect of its relevance. According to its length analysis, 523 has three

variants: short (S) 20 T residues, long (L) 20–30 T residues, and very long (VL) 30 T residues [103].

1.6.2 Alus at intro 6 and TOMM40 stability

The 3′ end of TOMM40 is particularly rich with Alu elements, and over evolutionary time, these have inserted themselves into the germline in both sense and antisense orientations, showed in Figure 1.3. The relative age of each Alu element and the orientation of Alu insertion events play an essential role concerning potential Alu exonisation and deleterious recombination events that can disrupt TOMM40 mRNA transcripts [39, 65, 121].

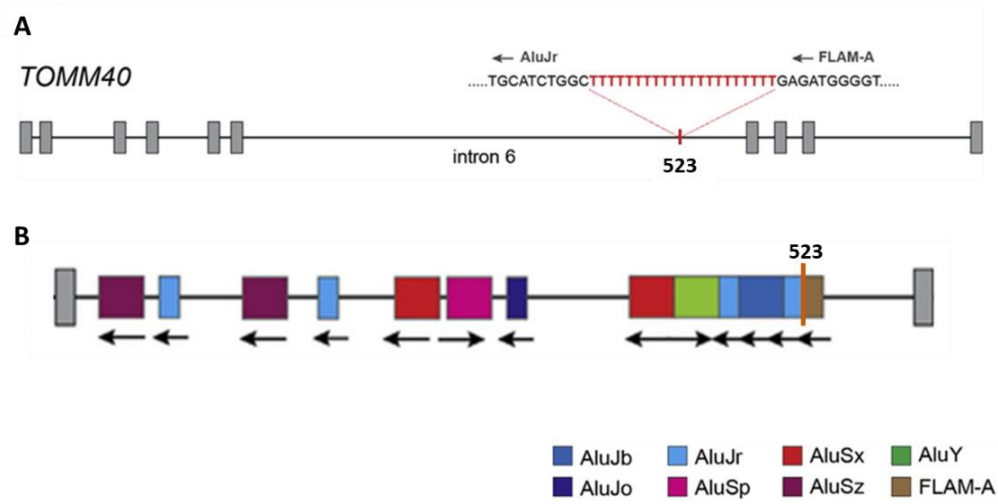


Figure 1.3 Representation of TOMM40 gene with SINEs (Alus) indicating 523 locations. A) TOMM 40 gene with the SINE block at intron 6 (Alu elements) and the poly-T 523. B). Alus of intron 6 with arrows indicating sense and antisense orientation. Adapted from Larsen et al. (2017) [122]

Alus bind to themselves post-transcriptionally when inserted inside a gene in opposing orientations and neighboring each other (Figure 1.4A), as the SINEs block in intron 6 of TOMM40 can result in the development of a duplex stem-loop structure that is maintained by the Alu nucleotide sequence and length [123]. These Alu-based secondary structures fundamentally change the shape of pre-mRNA molecules and

serve as the principal binding site for Adenosine deaminases acting on RNA (ADARs) catalyse adenosine (A) to inosine (I) editing of RNA. The ADAR proteins will bind to the double-stranded pre-mRNA duplex and edit out A to I, recoding pre-mRNAs (Fig. 1.4B). When working in coding areas (directly or indirectly), the translation machinery reads the resultant I residues as guanosine (G), and this process partially explains the extraordinary variety of the human proteome that is not encoded in the original DNA sequence [121, 123]. However, the vast majority of A-to-I editing on Alu elements occurs within pre-mRNA introns and 3′ UTRs, and this has a surprising number of direct effects on gene regulation and function, including the formation of different splice donor and acceptor sites that as a consequence causing Alu exonisation and alternative gene splicing [123, 124].

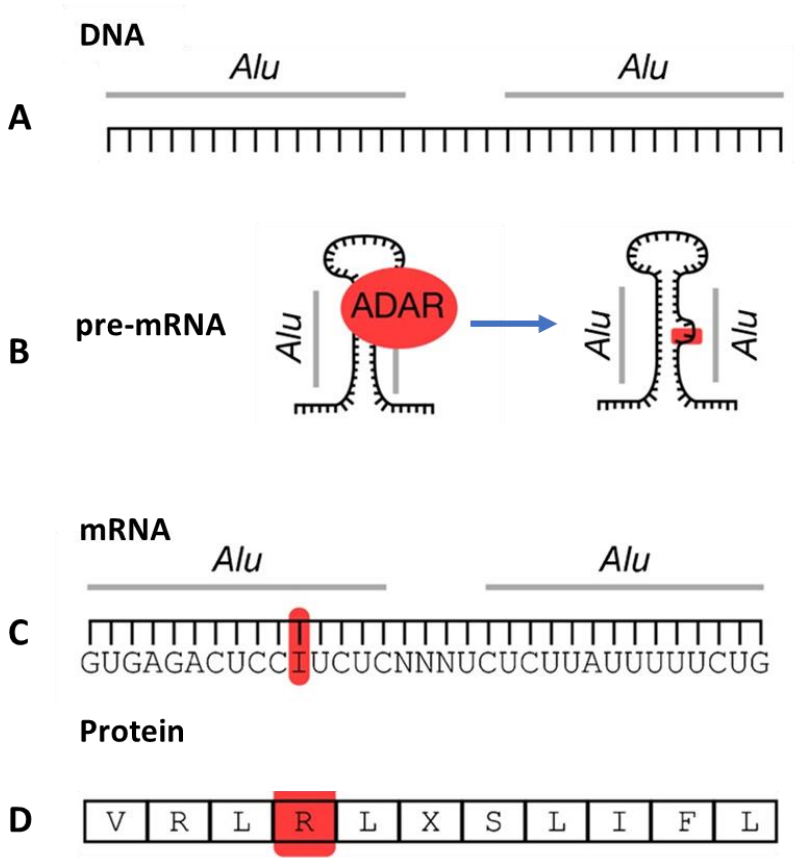


Figure 1.4. Exonization of intronic Alus. A) Adjacent intronic Alu elements may bind to each other within pre-mRNAs. B) Thus, generate a persistent stem-loop secondary structure, the major substrate for A-to-I editing. ADAR proteins bind to pre-mRNA Alu structures. C) The enzyme converts adenosine (A) residues to inosine (I). The translation machinery reads I

residues as G if present in coding areas, which may result in amino acid alterations and different protein conformations. D) The approach also applies to inverted Alus inside 3' untranslated regions, where ADAR editing might lead to nuclear retention of mRNAs. Adapted from Larsen et al. (2018), [124]

Currently, there is a vast number of evidence regarding the interaction of the APOE-TOMM40-APOC1 variations in AD risk is consistent with the complicated role of the APOE region in Alzheimer's disease (AD) pathogenesis [2, 125-128]. Kulminski et al. found that APOE 4- polygenic profiles are variably related to tau but not in plasma amyloid 42, and the modulation of the four alleles' impact by TOMM40 and APOC1. The variations suggest a genetic basis for the distinct involvement biomarkers in the aetiologia of Alzheimer's disease [129]. Still, many questions related to how this locus contributes to others neuro disorders.

Here, we investigated the interaction between the poly-T 523 found in an Alu-rich area of TOMM40 intron 6 [130]. Indeed, we shed light on the expression of the TOMM40 gene, as our data has provided evidence to allow others to elaborate on this discovery of how retrotransposons can contribute to mental health and other disorders.

1.6.3 Human-specific SVA D downstream of APOC1 gene.

A bioinformatics investigation of the genomic region showed the existence of many retrotransposons distinct to primates and humans. 116 SVAs are located on chromosome 19 [131]. An SVA D was identified at about 1750 kb downstream of the APOC1 gene (Figure 1.5). According to conservation studies, this specific SVA D is located on chromosome 20 in chimpanzees. Despite the shift in localization, the SVA D shares 96% of its sequence with human DNA. Gorillas and orangutans have a comparable SVA D (95% identity) on chromosome 14 but not on chromosome 19. According to the UCSC browser, neither the gibbon nor the rhesus macaque genomes

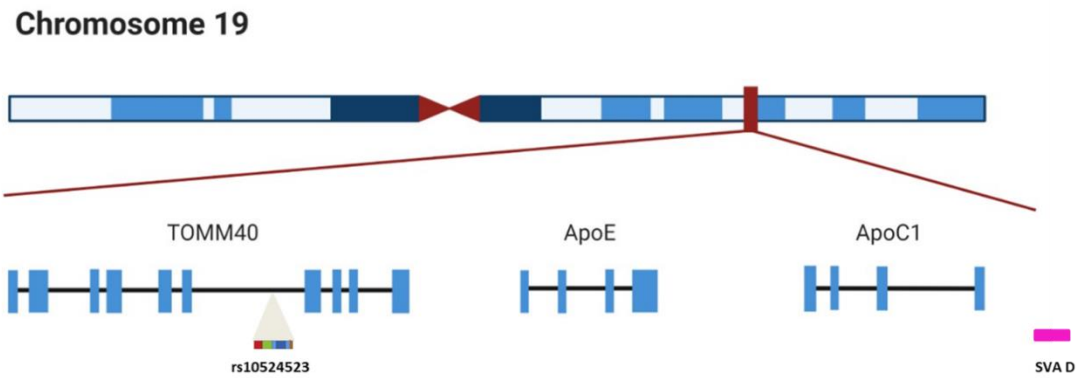


Figure 1.5 Representation of TOMM40-APOE-APOC1 Locus. Displaying chromosome location, at TOMM 40 gene the SINE block at intron 6 (Alu elements), APOE gene and APOC1 with the SVA D downstream.

Therefore, APOC1 SVA D is unique to humans. Previous studies suggest that old-world monkeys, such as the rhesus macaque, possess just the precursor of the VNTR domain present in SVAs, but no full-length SVAs, and that these retrotransposons are specific to humans. [131, 132]. Genomic analysis using data from ENCODE project (available at www. genome.ucsc.edu/ENCODE/) indicates that this particular SVA D is located approximately 5-7 kb downstream of potential active regulatory elements. According to the database, the region is enriched in the H3K27 Acetylation histone mark, often found near active regulatory elements. In addition, the region is rich in transcription factor binding sites and DNAse hypersensitive areas, which are indicative of regulatory elements, such as promoters. However, APOC1 SVA D did not significantly modulate the luciferase gene activity. Future work could involve functional assays using different cell lines or model systems.

1.7 The monoaminergic pathway

The monoaminergic neurotransmitter pathway is crucial to maintaining mental health since it is engaged in several elements of psychological and emotional balance [133]. Several neurological diseases have been linked to key neurotransmitters like monoamine, serotonin (5-hydroxytryptamine or 5-HT), noradrenaline (NE), and

dopamine (DA). The monoaminergic pathway within the human CNS is shown in Figure 1.6 ways, including by inhibiting the reuptake of neurotransmitters like serotonin and noradrenaline, one of the mechanisms of antidepressants' action in the brain.

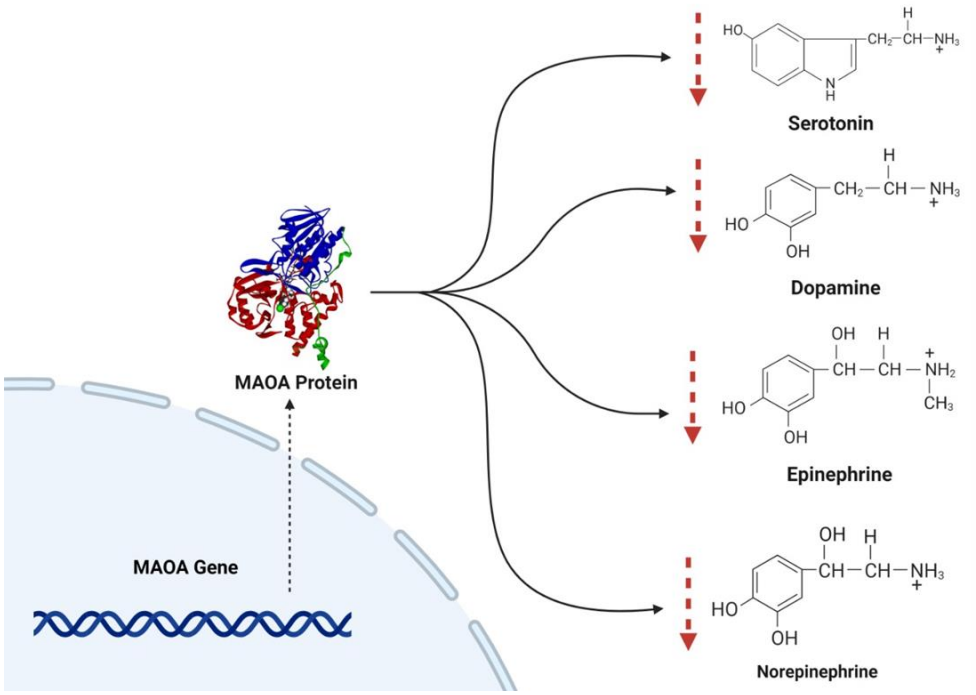


Figure: 1.6 Monoaminergic pathways. Representation of the MAOA protein and critical neurotransmitters, such as serotonin, dopamine, epinephrine and noradrenaline. The figure was created with BioRender.com.

MAOA is expressed in the presynaptic terminals of catecholaminergic neurons, where it selectively degrades and regulates the levels SE, NE, and DA, (Figure 1.7), following reuptake from the synaptic cleft and, as a result, plays an essential role in regulating complex behaviours such as mood, appetite, sleep, cognition, perception, motor activity, temperature regulation, pain control, It is involved in various physiological processes, including blood pressure control, peripheral and central nervous system neurotransmission, and smooth muscle function modulation. The control of the MAOA gene is a crucial factor in determining the monoamine concentration in the central nervous system [134]. Many antidepressants available for treatment work in the monoaminergic pathway in several

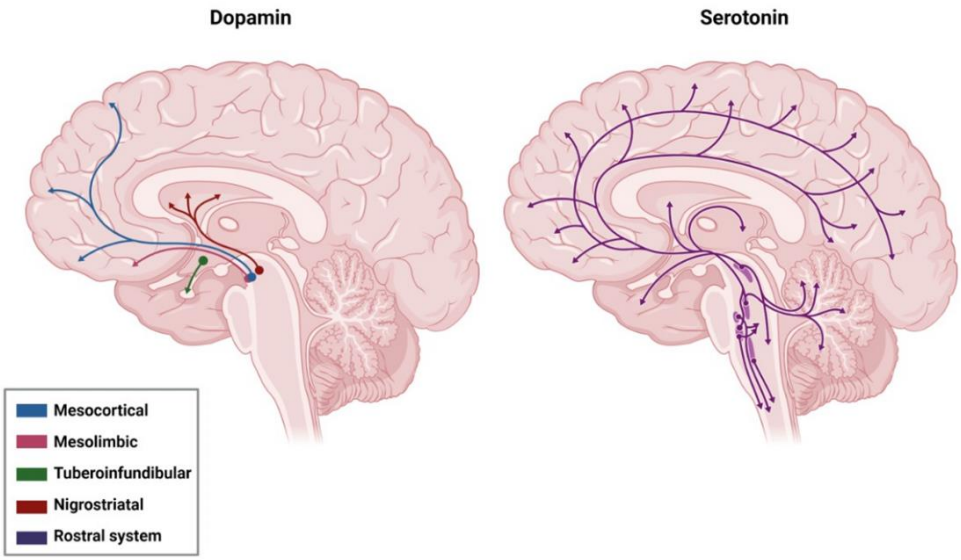


Figure: 1.7. A schematic of the main monoaminergic pathways within the human CNS. Main dopaminergic (left), and serotoninergic (right) pathways within the brain. Left: dopaminergic neurons are concentrated in the midbrain 1) mesocortical 2) mesolimbic 3) tuberoinfundibular 4) nigrostriatal. Right: showing the serotoninergic neurons Rostral system. The figure was created with BioRender.com.

Given the importance of *MAOA* in the metabolism of key neurotransmitters implicated in aggressive behaviour, most notably serotonin (5-hydroxytryptamine; 5-HT) [135, 136], it is not surprising that a large body of research has linked aggressive phenotypes to the MAO system. Thus, human genetic research has linked aggressive behaviour to unique allelic differences in the genes that code for the MAOA [30, 134, 135, 137]. These findings prompted the scientific community and the public to refer to *MAOA* as the warrior gene [138]. On the other hand, pharmacological studies characterised the behavioural changes associated with the administration of selective and nonselective MAO inhibitors (MAOIs), which, interestingly, are widely prescribed to treat a variety of mental disorders, such as anxiety, depression, posttraumatic stress disorder (PTSD), and Parkinson disease [139, 140]. Even though early pharmacological experiments demonstrated a role for Monoamine inhibitors (MAOIs) in the reduction of aggressive phenotypes, the findings from these studies were difficult to interpret due to MAOI side effects and their influence on a variety of unrelated behaviours [140].

Brunner et al. discovery that mutations in the human MAO-A gene result in aggressive phenotypes [141] and that MAOs' primary function is in the metabolism of 5-HT and other monoamine neurotransmitters justify a strong scientific interest in the modulatory role of the MAO system in adaptive and pathological aggression [142, 143]. Nilsson et al. described the involvement of MAOA and environmental risk. Eight have found that the low activity allele of uVNTR is positively associated with the G x E. Another 5 with the low variant also had partial positive interaction and the other three studies demonstrated at least one high-activity allele. The gene's position on the X-chromosome is a second issue with the G x E method for analysing the effect of MAOA. Females have two X chromosomes, while men only have one, and heterozygosity is possible in females but not males. Due to the obscurity of MAOA expression in heterozygous allele carriers, several researchers have chosen only men or excluded heterozygous females from their samples [144]. The serotonin transporter gene SLC6A4 and the 5HTTLPR polymorphism have been analysed for similar unexplained heterozygous effects. If the heterozygous person is expected to exhibit intermediate phenotypic expression between the two homozygous forms, molecular heterosis will not be considered. This molecular heterosis happens when an individual heterozygous for a specific genetic polymorphism has a much larger or smaller influence on a trait than an individual homozygous for either allele [145]. Accumulating data indicates that molecular heterosis is prevalent in humans, may occur in up to fifty per cent of all gene associations, and is essential for genes in the monoaminergic system, such as DRD4 and SLC6A4 [146]. If such a heterosis effect occurs, it will be mitigated by grouping homozygous people with the susceptibility allele in the same group as those with only one susceptibility allele, as opposed to people homozygous for the other non-susceptibility allele. [137].

1.7.1 Gene x environment interaction (G x E)

Gene-environment interaction refers to the consequences of environmental and genetic factors on the health of individuals [134, 147, 148]. It has been thoroughly proven that polymorphic genetic variations may be differentially affected by environmental stresses and act independently on gene expression; consequently, they might be regarded as risk factors for neuropsychiatric diseases or inappropriate behavioural responses (Figure 1.8) [29].

A substantial proportion of phenotypes are determined by variations of the same gene or genomic region, like in Huntington's disease, in which a genetic expansion produces a non-functional protein [149]. Nevertheless, there are occasions where genetic variation cannot explain disorders' onset, the progress of these diseases, or the increase of risk factors to that particular variation. Several disorder onsets are often discordant with Mendelian inheritance patterns [150], diverging within families and other biological factors. It is necessary to comprehend how environmental factors can contribute to the disruption in gene function, potentially causing a long-lasting response to those environment challenges.

Interactions of G x E may influence the likelihood of developing behavioural and neurodevelopmental disorders by causing a differential gene expression [151, 152].

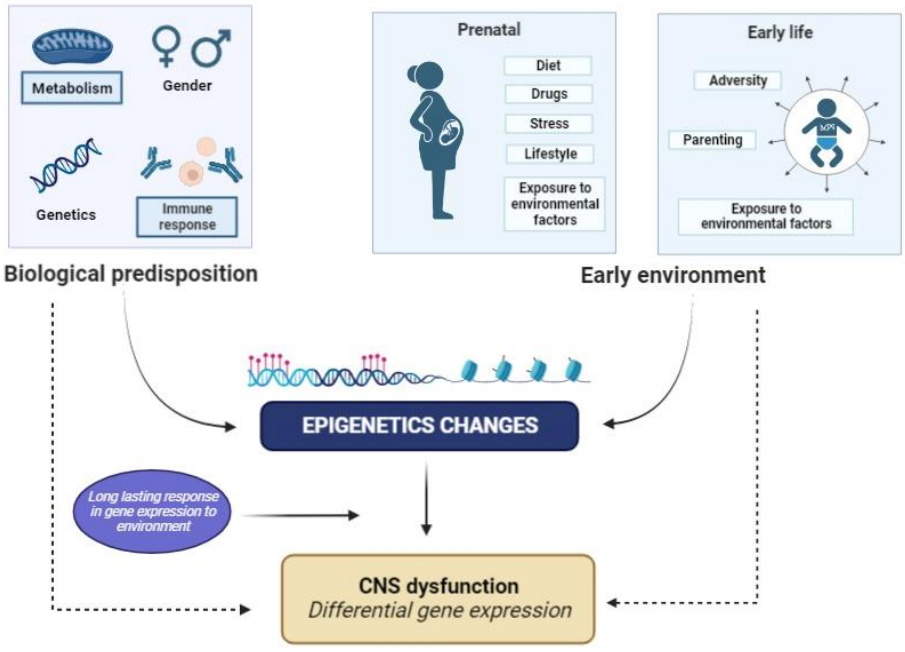


Figure 1.8. Gene-environment interaction. Prenatal and early life exposure to early environmental factors combined with a biological predisposition and other parameters, such as parenting, and adversity, can affect the development of the foetus from the first stages of pregnancy and compromise the health of the offspring later in life. Figure adapted from Quinn et al. [8] and using BioRender.com.

Recently, it has become more relevant to take into consideration the interplay involving G x E factors, as it might explain why certain disorders develop, considering that individuals can have identical genotypes, but if considered the diverse environments, potentially we can observe different responses, it is important to find the patterns of G x E modulations, to understand how it can. impact in several diseases, including mental health [30, 153].

The low activity and high activity variations of the MAOA VNTR that have already been found have been linked to several CNS disorders through case-control studies. In addition, environmental factors have the capacity to alter the adaptability of cells through epigenetic processes, such as DNA methylation, histone modifications, and X chromosome inactivation, that are used by cells to control gene expression to evolve and adapt to their environment [154, 155]. The MAOA gene, which is discussed at length in chapters 4 and 5, is one of the most often quoted as an example of G x E. These epigenetic changes are necessary for normal development, but they can be easily changed by things like immune responses, diet, stress, adversity, parenting and others environment challenges, which can change how genes are expressed, causing not only temporary but permanent changes, (Figure 1.9) [23, 156]

The topic of childhood abuse and stressful life events is complicated and broad. The incidence of abuse may be more significant than its form, intensity, or length. It is unknown why some types of misuse correlate more strongly in some research than others [157]. In addition, there is a great deal of overlap among maltreatment variables. Various maltreatment variables also occur in the context of and intersect with other indicators of low-income family function, such as witnessing domestic violence involving other family members, having an alcohol or drug-abusing parent,

having low socioeconomic status, a low parental educational level, or parental unemployment, living in a socioeconomically deprived neighbourhood [136, 158]

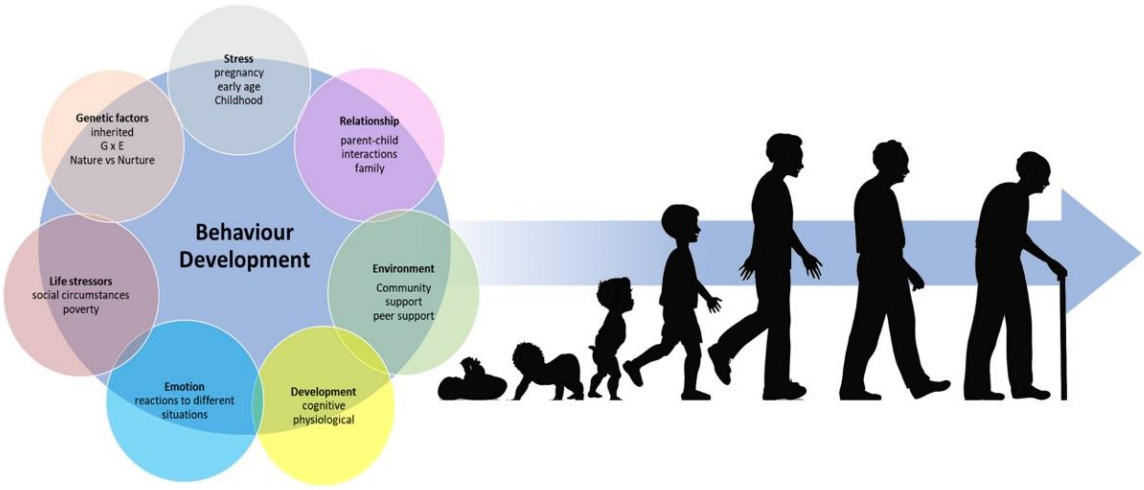


Figure 1.9. Gene x Environment interactions can modulate behaviour development. Early in life, while the brain is still developing, experiences may alter the epigenome, leading to lifelong influence on mental health.

MAOA is one of the most studied brain candidate genes because it is a significant regulator of monoamine neurotransmitters in the brain. It is one of the best-characterised and most cited genes in gene-environment interaction (G x E) studies, especially concerning central nervous system (CNS) disorders [159]. The uVNTR polymorphism, which is in the promoter region of the MAOA gene, has been well-studied and shown to be functional in regulating the MAOA gene. A second polymorphism, the dVNTR, which is 500 bp upstream of the MAOA promoter area, drives the regulation of the MAOA gene through the uVNTR, which will be explained in more detail later in chapter 4 [160].

1.7.2 Monoamine Oxidase A

In 1993, a Dutch family was the first to describe the monoamine oxidase A (MAOA) gene, where it was shown that males with aggressive behaviours had a mutation in

the MAOA gene. The MAOA gene encodes the monoamine oxidase A enzyme, which is required to breakdown neurotransmitters like dopamine, serotonin, and adrenaline [43].

MAOA (UCSC browser, hg38, chrX:43,656,219-43,748,821) is one of the most studied brain candidate genes and has been considered a major regulator of monoamine neurotransmitters in the brain and is one of the best characterised and most cited genes in gene-environment interaction (GxE) studies, especially concerning central nervous system (CNS) disorders [30, 152, 159].

MAOA is expressed in the presynaptic terminals of catecholaminergic neurons, where it selectively degrades and regulates the levels of serotonin (SE), noradrenaline (NE), and dopamine (DA) following reuptake from the synaptic cleft and, as a result, plays an essential role in regulating complex behaviours. Such as mood, appetite, sleep, cognition, perception, motor activity, temperature regulation, and pain control. It is involved in various physiological processes, including blood pressure control, peripheral and central nervous system neurotransmission, and smooth muscle function modulation. The control of the MAOA gene is a crucial factor in determining the monoamine concentration in the central nervous system [161]

MAOA is a gene on the X chromosome; hence, the ratio between men and females is different since males are hemizygous. In addition, it has been suggested that MAOA is part of a locus that escapes X inactivation [162-164]. This may be compatible with the recent finding that XIST, a gene that contributes to the regulation of X chromosome inactivation, is overexpressed in females with significant affective disorders [162]. Therefore, differential X chromosomal inactivation may substantially affect the control of MAOA in females.

This additional degree of difficulty in the modulation of MAOA gene expression in females may explain in part why many studies of the MAOA GxE only report on males, which is attributed to uncertainty regarding the active allele in heterozygous females [135, 165, 166].

Aslund et al. and Nikulina et al. conclude that research in both sexes has led to more complex results about how the uVNTR affects behaviour [167, 168]. For example, ADHD and anxiety only have strong effects in women, while autism, bipolar disorder, and aggressive behaviour only affect men [139, 154].

1.7.3 Upstream variable number of tandem repeats (uVNTR)

In the promoter region of the MAOA gene, an upstream variable number of tandem repeats (uVNTR) consists of a 30-bp motif that may be repeated 2, 3, 3.5, 4, and 5 times [169]. The 2, 3, and 5 repeats are often classified as low-expression variants (MAOA-L). However, the 3.5 and 4 repeat VNTRs have been demonstrated to enhance reporter gene expression by 2- to 10-fold and are identified as high expression variants (MAOA-H) [169].

There is evidence that specific uVNTR mutations function as moderators of the observed connection between various environmental risk factors and child behavioural disorders [28, 30, 135]. However, most of the research identifies the uVNTR as the primary, or single, modulator of MAOA expression and a potential biomarker for stress-related disorders, with the various alleles involved in major depressive disorder, addiction, aggressive conduct [154, 170-172].

1.7.3.1 Distral variable number of tandem repeats (dVNTR)

The second VNTR, designated dVNTR, is roughly 500 bp upstream of the uVNTR and consists of two distinct decamer repeats, CCCCTCCCCG (repeat A) and CTCCTCCCCG (repeat B). There are known genotypes with 8, 9, 10, 11, and 12 repetitions, 9R and 10R being the most prevalent. Similar to the uVNTR, the 9R and 10R vary in transcriptional effectiveness in reporter gene tests, with the 9R being

considered the high activity variant than the 10R and the other genotypes being intermediate [159].

In this study, we also analysed the variation in the haplotype of the MAOA promoter in the general population to identify common haplotypes that may be used to stratify the genetic risk of MAOA polymorphism in mental illnesses. By analysing the semi-haploid HAP1 cell line [173], deleted for either the uVNTR, dVNTR, or both VNTRs, we determined the effect of the two VNTRs, both individually and in combination, on MAOA gene expression and specifically on the two distinct transcription start sites (TSSs) of MAOA's main isoforms. Chapter 4 provides a detailed discussion.

1.7.4 Solute carrier family 6 member 4

The serotonin transporter (5HTT, SERT) is coded by the solute carrier family 6 member 4 (SLC6A4) gene placed on chromosome 17q11.1–17q12, spans 31 kb and comprises 14 exons [174-176]. The 5HTT is necessary for serotonin reuptake (5HT) from the synaptic synapse and is essential for serotonin level regulation in the brain [177]. The role of SLC6A4 in several mental disorders, such as PTSD, depression, and aggressiveness, is well established [178, 179].

The primary molecular focus for selective serotonin reuptake inhibitors (SSRIs) is 5-HTT, which is responsible for serotonin reuptake on the terminals and cell bodies of serotonergic nerve cells [180].

Serotonin plays a crucial role in the monoamine pathway as a neuromodulator, and an imbalance of serotonin can play a role in the pathophysiology of many related neuro diseases. Furthermore, it is crucial to comprehend the function of these genetic variations in the serotonin transporter (SLC6A4).

1.7.4.1 Serotonin transporter-linked polymorphic region (5-HTTLPR)

A variable number of tandem repeats (VNTRs) of insertion–deletion (indel) variation, known as the serotonin transporter-linked polymorphic region (5-HTTLPR), have been identified in the promoter region of the SLC6A4 gene [21]. The polymorphism is located approximately 1 kb upstream of the transcription start site (TSS), hg38 chr17: 30,237,061–30,237,479, and is a polymorphism with two different alleles, with 14 and 16 repeats (R), also known as short (S) and long (L) alleles, which can be an insertion or deletion of the 44 bp [181].

The L and S variations of the promoter polymorphism have been described to have functional differences in regulating the 5-HTT gene's transcription and subsequent 5-HTT availability [144]. The L-allele is associated with adequate serotonin transporter expression, while the S-allele is associated with decreased serotonin transporter expression [182]. It has been hypothesised that it modifies the functioning of the L-allele by altering the expression of the SLC6A4 gene [183]. It has been observed that the long G (L-G) haplotype functions similarly to the S-allele, resulting in a reduction in the expression of the SLC6A4 gene, in contrast to the long A (L-A) haplotype, which increases the expression of the SLC6A4 gene [184]. It has been suggested that the short allele (S) can lead to less transcription of the SLC6A4 gene. Studies suggested that the interaction between the short/short (S/S) allele of the 5-HTTLPR and stress in life was a predictor of major depression (MDD) [11]. One research revealed that individuals with the SS genotype had an increased chance of developing aggressive and violent behaviours, impulsive aggressiveness, neuroticism, criminal behaviour, and wrath [185]. According to the evidence, the S-allele is associated with conduct disorder [159], and antisocial behaviour [144]. This research reveals a relationship between genetic variations of the SLC6A4 gene and mental health illnesses and behavioural difficulties. The S-allele is associated with both aggressiveness and PTSD, indicating that their pathophysiological mechanisms are comparable. The S-allele decreases the expression of the SLC6A4 gene, resulting in serotonin

dysregulation [186]. This disruption in serotonin control is caused by an imbalance in the neuroendocrine response, which is seen in aggressive behaviour.

1.7.4.2 Serotonin transporter intron 2 (STin2)

A second polymorphism in intron 2 of SLC6A4, hg38: chr17: 30, 221,314–30, 221,701, referred to as the serotonin transporter intron 2 (STin2), has been investigated in aggression [187]. This polymorphic region is distinguished by varying numbers of 17-bp repeats, resulting in 9, 10 and 12 repeat variants, with the 10R and 12R alleles being the most frequent [188]. The 12 repeats are designated as the long allele (L), while 9R and 10R are designated as short alleles (S). The short allele is related to decreased expression of the SL6A4 gene. However, the 12R is associated with increased expression of this gene [189]. The findings indicate a correlation between the L allele and aggressive behaviour in those with impulsivity and chronic and pervasive hostility [73]. Recently, Hemmings and colleagues discovered that the 12R variant was associated with decreased reactive aggression but increased perpetration of aggression in a male South African population, indicating that genetic underpinning may be necessary to differentiate between forms of aggression [190]. It is believed that these genetic variations contribute to the development of aggressiveness.

1.7.5 Dopamine Receptor D4 (DRD4)

The DRD4 gene codes for the dopamine receptor 4 (DRD4) protein, which can be found in the CNS in the hypothalamus, frontal cortex, hippocampus, and striatum. The protein moderates neuronal memory development [191]. According to studies, alterations in DRD4 have a detrimental effect on brain function and may increase the chance of developing aggressiveness [192].

A variation in DRD4 spanning exon 3, especially 2-repeats (2R), was shown to be related to delinquency and aggression [193]. The 7-repeats (7R) in exon 3 of DRD4 have been linked to aggressive, externalising, and antisocial behaviour [194, 195]. Recent research [196] has shown an environmental link between traumatic occurrences and the DRD4 gene. The studies of a link between DRD4, aggressiveness, and PTSD show that mutations in the DRD4 gene may impair the normal function of the DRD4 protein, consequently affecting the standard processing of memories [197].

Chapter 2

Material and Methods

Part 1: Material

Part 2: Methods

2.1 MATERIAL

2.1.1 Commonly used Buffers and Reagents

TBE buffer (5x): 108g Tris base (Sigma), 55g Boric acid (Sigma), 5.84g EDTA (Sigma), made up to 2L with distilled water. TBE buffer is further diluted to 0,5x, as required to prepare agarose gel for electrophoresis, as described in section 2.2.5.3.

LB agar: 40 g/L in distilled water (Fluka Analytical), and used as described in section 2.2.5.8, to plate transformed E. coli cells, aim to grow individual colonies to cloning protocols.

LB broth: 25 g/L in distilled water (Fluka Analytical), required to culture clone aimed to DNA isolation in mini and maxi-preparation of plasmid DNA, sections 2.2.5.9.1 and 2.2.5.9.2.

70% Ethanol Molecular Grade: 700 ml of ethanol molecular grade (Sigma) diluted to 1L using distilled water. Required to the protocol of RNA extraction, described in section 2.2.3.

2.1.2 Chromatin Immunoprecipitation (ChIP) buffers

Reagents required to the in vitro ChIP, as explained in section 2.2.15:

TE buffer: 10 mM Tris–HCl, pH 8.0, 1 mM EDTA.

Cell lysis buffer: 50mM Hepes - KOH pH7. 5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5%

NP-40, 0.25% Triton X-100, protease inhibitor cocktail (Sigma Cat No. P8340).

Nuclear lysis buffer: 10mM Tris-HCl, pH8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA, protease inhibitor cocktail (Sigma).

ChIP dilution buffer: 16.7 mM Tris–HCl, pH 8.1, 167 mM NaCl, 1.1%Triton X-100, 0.01% SDS, 1.2 mM EDTA, supplemented with 10 µl/ml 100X protease inhibitor cocktail (Sigma) immediately before use.

Low-salt wash buffer: 20 mM Tris–HCl, pH 8.1, 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA.

High-salt wash buffer: 20 mM Tris–HCl, pH 8.1, 500 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA.

Lithium chloride (Li Cl) wash buffer: 10 mM Tris–HCl, pH 8.1, 250 mM LiCl, 1% Igepal, 1% sodium deoxycholate, 1 mM EDTA.

Elution buffer: 50 mM Tris–HCl, pH 8 1 mM EDTA, 1%SDS, 50 mM $NaHCO_3$, 2. 5 M NaCl.

RNAse: RNAse A: 20 mg/ml. (Sigma)

SDS 10%: Sodium dodecyl sulphate (Sigma) 10% w/v in distilled water.

Proteinase K: Proteinase K solution, 20mg/mL, >600mAU/mL Activity, Ready-to-use (Qiagen).

2.1.3 Drug Treatment Solutions

Cortisone (Sigma): dissolved in ethanol to make a 1 mM stock which was diluted in complete SH-SY5Y tissue culture media to a final concentration of 20nM, 50 nM or 100 nM.

Lithium chloride (Sigma): dissolved in sterile filtered d.H2O to make a 1 M stock.

Which was diluted in complete SH-SY5Y tissue culture media to a final concentration of 1 mM (Hing *et al.*, 2012, Roberts *et al.*, 2007).

Valproic acid sodium salt (Sigma): dissolved in sterile filtered dH2O to make a 1M stock which was diluted in complete SH-SY5Y tissue culture media to a final concentration of 5 mM (Pan *et al.*, 2005, Zhang *et al.*, 2003, Phiel *et al.*, 2001).

2.1.4 Cohorts

2.1.4.1 Wirral child health and development cohort

The Wirral Child Health and Development Study (WCHADS), a prospective epidemiological longitudinal cohort, aimed to study prenatal and infancy origins of emotional and behavioural disorders funded by the Medical Research Council (MRC). The study was introduced to the mothers when they attended an appointment at Wirral University Teaching Hospital for the 20 weeks gestation ultrasound scanning between 12 February 2007 and 29 October 2008. After obtaining written informed consent, research midwives administered questionnaires and subsequently gathered obstetric outcome data from medical records.

Our samples comprehend an intensive sub-group of 316 families stratified from a total cohort of 1233 first-time mothers with live singleton births. All participants have taken part in several assessments before being randomly stratified into sub-sample groups. Groups were defined based on low, medium and high psychosocial risk based on the psychological evaluation done at the recruitment to the study, at 20 weeks of pregnancy, as described by Sharp et al. [198]. As a result, 316 were identified for an additional intensive review (intensive sample).

Saliva samples for nucleic acids extractions were collected during the interview and physiological measures, where mothers and children came into the laboratory for detailed observational, interview and physiological measures.

Wirral socioeconomic conditions range between the deprived inner city and affluent suburbs, but with low numbers of women from ethnic minorities. The socioeconomic

status of the stratified sub-sample was determined using the revised English Index of Multiple Deprivation (IMD). A detailed sample description is given by The English Indices of Deprivation 2004 (Revised); Report to the Office of the Deputy Prime Minister; Neighbourhood Renewal Unit: London, UK, 2004. [199]

All families were given IMD ranks according to the postcode of the area where they lived and assigned to a quintile based on the UK distribution of deprivation. In the extensive sample, 41.8% were in the most deprived UK quintile, consistent with high levels of deprivation in some parts of the Wirral.

Ethical approval for the study was granted by the Cheshire North and West Research Ethics Committee (UK) (reference no. 05/Q1506/107).

2.1.4.2 Manchester Longitudinal Studies of Cognition in Normal Healthy Old Age Cohort.

Genomic DNA samples of 1613 volunteers from The Longitudinal Study of Cognition in Normal Healthy Old Age (The University of Manchester) with documented longitudinal trajectories in cognitive function consists of a large sample of older adults from the North of England, UK; the Manchester and Newcastle Longitudinal Studies of Cognitive Aging Cohorts (Rabbitt et al., 2004). Recruitment took place in Newcastle and Greater Manchester between 1983 and 1992. At the start of the study, 6063 volunteers were available, 1825 men and 4238 women, with a median age of 65 years (range 44 - 93 years).

The studies have collected biennially alternated batteries tests and assessments of demographic, lifestyle, health, cognitive, and emotional health data, as described by [200]

Ethical approval for all projects was obtained from the University of Manchester.

2.1.4.3 Schizophrenia cohort

Genomic DNA samples from 823 patients with schizophrenia (mean age 37.66 years, range 18-71) and 762 healthy controls (mean age 46.27, range 19-72) were kindly provided by our collaborator Prof. Dan Rujescu, Department of Psychiatry, University of Halle-Wittenberg. Subjects were of German or central European descent and provided written informed consent. Schizophrenic patients were selected based on diagnosis under the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) and International Classification of Disease-10 (ICD-10). Detailed medical and psychiatric histories were collected for each patient, including the Structured Clinical Interview for DSM-IV (SCID), to evaluate lifetime Axis I and II diagnoses. Unrelated healthy controls were randomly selected from the general population of Munich, Germany. To exclude any healthy volunteers with neuropsychiatric disorders, both the subjects and their first-degree relatives completed an initial screening process followed by a detailed medical and psychiatric history assessment using a semi-structured interview. Participants that did not meet the exclusion criteria were invited to a comprehensive consultation, including the SCID I and SCID II, to validate the absence of any lifetime psychotic disorder. All participants provided written informed consent following a detailed and extensive description of the study, which was approved by the local ethics committee of Ludwig Maximilians University, Munich, Germany, and carried out following the ethical standards outlined in the Declarations of Helsinki.

2.1.5 Human cell lines

Cells were cultures as described in section 2.2.8, and treated as required to transfection section 2.2.9 and 2.2.10, reporter gene assay, 2.2.11 and mRNA expression analysis 2.2.12.

*2.1.5.*1 SH-SY5Y

The human-derived neuroblastoma cell line was obtained from the *American Type Culture Collection* (*ATCC, CRL-2266).*

2.1.5.2 S-K-NAS

Human neuroblastoma cell line, CRL-2137, from the European Collection of Cell Culture (ECACC).

2.1.5.3 HAP1

HAP1 is a near-haploid human cell line that was derived from a male with chronic myelogenous leukaemia (CML) designated cell line KBM-7 (Horizon) [173]

2.1.6 Cell culture media

2.1.6.1 Complete media for SH-SY5Y cells

Earle's modified Eagle's medium (Sigma) and HAM's F12 (Sigma) at a ratio of 1:1, supplemented with 10% foetal bovine serum (Sigma), 1% 200 mM L-glutamine, 1% 100 mM sodium pyruvate and 100 U/ml penicillin/ 100 µg/ml streptomycin.

2.1.6.2 Complete media for S-K-NAS cells

Dulbecco's Modified Eagles medium with 4500mg glucose/L (Sigma) supplemented with 1% (v/v?) non-essential amino acid solution (Sigma), 100 U/ml penicillin and 100 µg/ml streptomycin and 10% foetal bovine serum (Sigma).

2.1.6.3 Complete media for HAP1 cells

Iscove's Modified Dulbecco's Medium (IMDM) (GIBCO), supplemented with 10% foetal bovine serum (Sigma) and 100 U/ml penicillin/ 100 µg/ml streptomycin.

2.1.6.4 Cryoprotective freezing medium.

90% foetal bovine serum (Sigma), 10% DMSO (Sigma).

2.1.6.5 Cryoprotective freezing medium for HAP1.

Medium A: Iscove's Modified Dulbecco's Medium (IMDM) (GIBCO) + 20 % foetal bovine serum (Sigma).

Medium B: Iscove's Modified Dulbecco's Medium (IMDM) (GIBCO) + 20 % foetal bovine serum (Sigma) + 20 % DMSO (Sigma).

2.1.7 PCR primers

Table 2.1: PCR primers used for genotyping, gene expression profiling, ChIP 2.2.15 and sequencing.

Gene	Region	Forward (5'- 3')	Reverse (5'- 3')	Application *	Size (bp)	Template	Annealing (°C)
SLC6A4	5HTT-LPR	ATGCCAGCACCTAACCCCTAATGT	GGACCGCAAGGTGGGCGGGA	1	419	DNA	55
DRD4	VNTR Exon 3	CGTGTGCTCCTTCTTCCTAC	GGGTCTGCGGTGGAGTCT	1	398	DNA	55
MAOA	uVNTR	GAACGGACGCTCCATTCGGA	ACAGCCTGACCGTGGAGAAG	1and 4	324	DNA, cDNA	61
MAOA	dVNTR	GGGTTAAGCGCCTCAGCTTG	CAAGAGTGGACTTAAGGAAGCAG	1	365	DNA	TD 65 - 55
SLC6A4	STin2	GGATTTCCTTCTCTCAGTGATTGG	TCATGTTCCTAGTCTTACGCCAGTG	1	388	DNA	TD 65 - 55
APOC	SVA D	CGTGAGCCACATCGTCTG	CTAAGGCCGGGCTCTGTG	1 and 2	989	DNA	TD 65 - 55
TOMM40	SINEs	TGCTTTGCTCTGTGCTGTCT	TGCACTGAGGTCCAAGTCTG	2	1856	DNA	60
TOMM40	alu1AI	CTCGATCGTGGTGGATTTCT	CCTTTTCAAGCCTCAGGATG	2	1660	DNA	60
TOMM40	alu1B	CTCGATCGTGGTGGATTTCTC	GAATTGGGGAGGAAGAGAGTG	2	1567	DNA	60-62
TOMM40	MiRs	CACAGCTGAGCACTCGATTC	GTTCCCTACCGCTGAGGTC	2	391	DNA	60
TOMM40	T40_Mir3	CTCCCAAAGTGCTGGGATTA	GACCAGTGTGCCTGAGATGA	2	167	DNA	60
TOMM40	FlamA 41T/ 17T	CCTCCAAAGCATTGGGATTA	CATTGGGGAGGAAGAGAGTGG	2	249	DNA	60

TOMM40	Alu3Mir3	GGAGCATCTGATGGGTGTTT	GACCAGTGTGCCTGAGATGA	2	480	DNA	60
TOMM40	polyT	TGCTGACCTCAAGCTGTCCTC	GAGGCTGAGAAGGGAGGATT	2	184	DNA	60
TOMM40	FlamA_T	CCTCCAAAGCATTGGGATTA	ATTGGGGAGGAAGAGAGTGG	2	248	DNA	60
ACTB	ACTB	CACCTTCTACAATGAGCTGCGTGTG	ATAGCACAGCCTGGATAGCAACGTAC	3 and 4	199	DNA, cDNA	63
MAOA	Exon I	CGGGTATCAAAAGAAGGATCG		4	298	cDNA	61
MAOA	Exon IIA		GCATCAGAGGAAAGCAGCTCCTGG				61
MAOA	Exon III	TACGTAGATGTTGGTGGAGCT		4	440	cDNA	61
MAOA	Exon VI		GGGCACCACTCGGATATTCT				61

2.1.8 Primers for cloning

Table 2.2: Reporter gene and expression constructs generated for use in *in-vitro* luciferase assays, as related in section 2.2.5.

Name	Primers for amplification (5' – 3')	Vector	Orientation	Application
pT40_al1_FW	AA**ACGCGT**CTCGATCGTGGTGGATTTCTC	pGL3P	Forward	Luciferese
pT40_al1_RV	TTTCTAGAGAATTGGGGAGGAAGAGAGTG			
FlamA_T_iso_FW	CGAGCTCTTACGCGTGCTAGCCTCCAAAGCATTGGGATTA	pGL3P	Forward	Luciferese
FlamA_T_iso_RV	AGATCGCAGATCTCGAGCATTGGGGAGGAAGAGAGTGG			
T40polyT iso FW	CGAGCTCTT**ACGCGT**GCTAGTGCTGACCTCAAGCTGTCCTC	pGL3P	Forward	Luciferese
T40polyT iso RV	AGATCGC**AGATCT**CGAGGAGGCTGAGAAGGGAGGATT			
Tomm40 SINEs FW	CGAGCTCTT**ACGCGT**GCTAGTGCTTTGCTCTGTGCTGTCT	pGL3P/pGL3C	Forward	Luciferese
Tomm40 SINEs RV	AGATCGC**AGATCT**CGAGTGCACTGAGGTCCAAGTCTG			
APOC SVA D FW	CGAGCTCTT**ACGCGT**GCTAGCGTGAGCCACATCGTCTG	pGL3P	Forward	Luciferese
APOC SVA D RV	AGATCGC**AGATCT**CGAGCCTAAGGCCGGGCTCTGTG			

Note: FW: Forward and RV: Reverse. Underlined sequences indicate vector overlapping, bold indicates the sequence of restriction enzymes RE sites of Insertion sequences, orientation is for direct cloning into the specified vector.

2.1.9 ChIP grade antibodies

Table 2.3. ChIP grade antibodies used in the human SH-SY5Y cell line, described in section 2.2.15.

Antibody	Host Species	Species Reactivity	Immunogen	Company and Catalogue Number
Anti-H3	Rabbit (Polyclonal)	Human, mouse, rat	A synthetic peptide corresponding to the carboxyterminal of human Histone H3.	Abcam, 1791
Anti-H3K9me3	Rabbit (Polyclonal)	Human, mouse, rat	Synthetic peptide derived from within residues 1-100 of Human H3, trimethylated at lysine 9	Abcam, 8898
Anti-RNA Pol II CTD phospho Ser5	Rat (Monoclonal)	Human	Synthetic peptide containing the RNA pol II C-terminal domain (CTD)	Active Motif, 61085
Anti-REST	Rabbit (Polyclonal)	Human, mouse, rat	GST fusion protein corresponding to residues 801-1097 of full-length human REST/NRSF	Millipore, 07-579
Normal rabbit IgG	Rabbit (Polyclonal)		Unconjugated antibody not directed against any known antigen. Used as a non-specific IgG control	NEB, 2729
Anti-CTCF	Mouse (Monoclonal)	Human	His-tagged recombinant protein corresponding to human CTCF	Millipore, 17-10044
Anti-Nucleolin	Mouse (Monoclonal)	Human	Human nucleolin protein from Raji cell extract.	Abcam, ab13541
Anti-hnRNPK	Rabbit (Polyclonal)	Human, mouse	A synthetic peptide corresponding to a sequence from the C-terminus of isoform an of human hnRNP K.	Abcam, ab70492
Anti-CNBP	Goat (Polyclonal)	Human, rat	Synthetic peptide: GESGHLARECTIE, corresponding to the C-terminal of Human CNBP.	Abcam, ab48027

2.2 Methods

2.2.1 Saliva sample collection

The saliva samples were collected in the Oragene DNA and RNA container (DNA Genotek, ON, Canada) until the saliva volume reached the indicated level marked on the collection tube. The storage and subsequent extraction were performed following the protocol outlined in the manufacturers' instructions to the DNA. RNA was extracted following the validated protocol for RNA extraction with Trizol.

2.2.2 gDNA Purification from whole saliva samples

The sample was weighted and subtracted 5.66g, corresponding to the average weight of the kit collection before adding the saliva. Volumes used for the purification were calculated according to the sample collection and followed the procedure described in the manufacturer's protocol. DNA was rehydrated by adding 500µl of Nuclease-free water and incubating at room temperature overnight to allow the pellet to dissolve, followed by the vortex. DNA concentrations were determined by the ratio 260nm/280nm using a NanoDrop 8000 Spectrophotometer. Samples were stored at -20°C.

2.2.3 RNA extraction from whole saliva samples

Samples were transferred to a 15ml Falcon tube and homogenised using an 18G needle syringe to allow cells to be pelleted, followed by a centrifugation step at 4000 rpm/4°C /10 minutes. Added 200µl Trizol and used a syringe with a 21-25G needle to break down the cells. Leave at room temperature for 10 minutes. In the sequence added, 40µl chloroform was to the Falcon tube mix by vortex and centrifuge at maximum speed/4°C/10 minutes. Removed the upper layer and added an equal volume of Isopropanol molecular grade, mixed by vortex and centrifuge at maximum speed/4°C/20 minutes. Removed supernatant and pellet

washed with 70% molecular grade ethanol. Pellet was resuspended in 50µl RNAse free water and stored at -80°C.

2.2.4 Primers

Polymerase chain reaction (PCR) primers were designed using the online primer designer software Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi), which generates a list of suitable PCR primers for amplification of the sequence of interest based on appropriate melting temperatures, GC% content and potential dimerization and hairpin formation. In general, primers were designed to be 18-25 nucleotides in length, have a melting temperature of 50-65°C and a GC-content between 40-60%. Primer specificity was determined using the *In-Silico PCR* and *Pick Primers* tools available from the UCSC Genome Browser (http://genome.ucsc.edu/index.html) and National Center for Biotechnology Information (NCBI) (http://www.ncbi. nlm.nih.gov/tools/primer-blast/). Primers were purchased from Eurofins MWG Operon and are listed in Tables 1 and 2

2.2.5 General Cloning Methods

2.2.5.1 PCR primer design for direct cloning into commercial vectors

Primer sequences are shown in *section* 1.8. Each primer had a minimum length of 15bp; appropriate restriction sites (present in the multiple cloning site of the chosen vector but absent in the target sequence) were added to the 5' end of the forward and reverse primers so that they were incorporated at the ends of the DNA sequence following PCR amplification. A random sequence of 6 bp was also included in the 5´ of the restriction sites to ensure efficient DNA cleavage by the restriction enzymes. Primers used for direct cloning are listed in Table 2.1. and PCR methods are outlined below in *section* 2.1.2.

2.2.5.2 PCR using a proof-reading polymerase

Phusion High-Fidelity DNA Polymerase (NEB) was used in PCR to amplify DNA targets. The Phusion DNA Polymerase master mix is outlined below:

Table 2.4: Reagents used for the PCR with Phusion DNA Polymerase.

Component	Volume (n=1)	Final Concentration
5X Phusion HF buffer	4 µl	1X
PCR nucleotide mix (10 mM of each dNTP)	1 µl	200 µM
Forward primer (10 µM)	1 µl	0.5 µM
Reverse primer (10 µM)	1 µl	0.5 µM
Phusion DNA polymerase (2 U/µl)	0.2 µl	0.4 U
Nuclease free water	X µl	-
DNA template	X µl	-
Final volume	20 µl	-

PCR reactions were performed in a peqSTAR 2X (peqlab) or SimpliAmp (Thermo Fisher) thermocycler. The amount of DNA template used varied between primer sets depending on the abundance of the target for amplification. Thermal cycling conditions were as follows: initial denature at 98°C for 30 seconds, followed by 25 cycles of denaturing, annealing and extension at 98°Cfor 10 seconds, X°C for 30 seconds and 72 °C for 30 seconds, respectively, with a final extension cycle at 72 °C for 10 minutes. The annealing temperature of each primer set was optimised using a gradient PCR and is listed in Table 2.1, along with primer sequences and expected product sizes. PCR products were analysed by agarose gel electrophoresis.

2.2.5.3 Analysis of DNA using agarose gel electrophoresis

DNA fragments from PCR reactions or restriction digests were analysed by gel electrophoresis on 1-2.0% agarose, a percentage subject to the fragment size, gels were supplemented with GelRed (1:10,000 dilution) (Cambridge Bioscience) or 0.5 µl per 10 ml of ethidium bromide (Sigma 10 mg/ml) and compared against a 100 bp or 1 Kb DNA ladder (Promega). The voltage (standard running conditions of 5 V/cm) and time for which the gel was run was dependent on the size of the fragments. The DNA was visualised using a UV transilluminator (BioDoc-it Imaging System).

2.2.5.4 Recovery of DNA from agarose-gels

Following separation by gel-electrophoresis, DNA fragments of the expected size were extracted from the agarose gel and column-purified using the QIAquick Gel Extraction Kit (QIAGEN), following the manufacturer's instructions. The purified DNA was eluted in a 30 µl Elution Buffer.

2.2.5.5 Restriction digest and DNA purification

Restriction enzyme digests were used either to create specific nucleic acid overhangs for ligation or as a diagnostic tool for determining the presence and/or orientation of inserts. Restriction enzymes were purchased from Promega, and digests were performed using the following reaction components:

Component	Volume (n=1)
Nuclease free water	X µl
10X Buffer	2 µl
Acetylated BSA (10 µg/µl)	0.2 µl
DNA (1 µg)	X µl
Restriction Enzyme (10 U/µl)	0.5 µl
Final volume	20 µl

Recommended buffers for optimum enzyme activity were used, digestions were incubated at the appropriate temperature for the enzyme activity for 1-4 hours and fragments run on agarose gels to visualise their size 2.1.3.

2.2.5.6 Ligation - T4 DNA Ligase

Purified DNA fragments were ligated into the appropriate vectors at an insert: vector molar ratio of 3:1 using the following ligation calculation:

$$Insert\ (ng) = \frac{vector\ (ng) \times size\ of\ insert\ (kb)}{vector\ size\ (kb)} \times ratio \left\{ \frac{insert}{vector} \right\}$$

For the ligation reaction, 1 µl (100 ng) of vector and the appropriate volume of the insert was added to 1 µl ligase 10X Buffer (Promega) and 1 µl T4 DNA Ligase (1-3 U/µl) (Promega) and made up to a final volume of 10 µl with nuclease-free water. The reaction was incubated at room temperature for 4 hours or at 16°C overnight.

2.2.5.7 Gibson Isothermal Assembly

Gibson Isothermal Assembly is a high-efficiency DNA end-linking technique using three enzymes to join two or more sequences of blunt-ended or 3' overhang dsDNA in only 1-hour reaction. A total of 0.02–0.5 pmols (weight in ng) x 1,000 / (base pairs x 650 daltons) of DNA fragments (blunt-ended or 3' overhang) were used when assembling 1 or 2 fragments into a vector and 0.2–1.0 pmoles of DNA fragments when assembling 4-6 fragments, with a 4-fold excess of the insert to vector. DNA fragments were diluted in nuclease-free water to a volume of 10µl and mixed with 10µl of 2X Gibson Assembly Master Mix (NEB Cat. No. E2611S) to a total volume of 20 µl, in a 50 µl PCR tube on ice, followed by 1 hr incubation at 50°C.

2.2.5.8 Transformation of chemically competent E. Coli cells

As detailed in the manufacturer's protocol, the Gibson assembly ligation mix was transformed into XL Gold ultracompetent cells (Agilent). Briefly, 100 µl aliquots of cells were thawed on ice and placed in pre-chilled 14 ml falcon tubes. 4 µl β-mercaptoethanol was added to each aliquot and incubated on ice for 10 minutes, swirling gently every two minutes. 2 µl of ligation mixture was added to an aliquot of cells, swirled gently to mix, and then incubated on ice for 30 minutes.

After incubation, cells were heat shocked in a water bath at 42 ºC for 30 seconds and incubated on ice for 2 minutes. Next, 0.9 ml NZY+ broth (pre-heated to 42 ºC) was added to each, and tubes were set at 37 ºC for one hour, shaking at 225 rpm.

Following one hour of incubation, 200 µl of each transformation mix was plated on LB agar plates supplemented with 100 µg/ml ampicillin and incubated overnight at 37 ºC.

2.2.5.9 Isolation of plasmid DNA from bacterial cultures

2.2.5.9.1 Mini preparation of plasmid DNA

To test plasmid DNA for inserts following molecular cloning, a small-scale preparation of DNA (up to 20 µg) was undertaken. Individual colonies grown on LB agar plates were transferred to 5 ml LB broth supplemented with 100 µg/ml complemented with the appropriate antibiotic and cultured overnight at 37°C on a shaker at 225 rpm. According to the manufacturer's guidelines, DNA was isolated from the resulting bacterial culture using the QIAprep Spin Miniprep Kit (QIAGEN). Purified plasmid DNA was eluted in 50 µl nuclease-free water into 1.5 ml microcentrifuge tubes and either stored at – 20 ºC or directly used for restriction enzyme digest to determine the presence or absence of an insert. To restriction enzyme digestion to check for the correct size and orientation of the insert.

2.2.5.9.2 Maxi-preparation of plasmid DNA

A Plasmid Maxi Kit (Qiagen) was used to obtain high yields of plasmid DNA from transformed bacteria of greater purity than that generated from mini preparations for use in downstream applications such as *in vitro* reporter gene assays.

The bacterial cultures were set up by adding 200 µl of 5 ml miniprep culture *section 2.1.11.1* to 100 ml of LB broth supplemented with the appropriate antibiotic and grown overnight at 37°C with shaking at 225 rpm, to generate an adequate amount of bacteria for extraction of the plasmid DNA. DNA purification was carried out according to the manufacturer's instructions for high-copy plasmids. The resulting DNA pellets were resuspended in 150 µl EB buffer, and the concentration was determined using a Nanodrop 8000 before storing for later use at – 20 ºC.

2.2.6 Sequencing

Sanger Sequencing confirmation of plasmids with successfully cloned inserts was carried out externally by Source Bioscience Life Sciences. The samples and primers were supplied as required by the company.

2.2.7 Glycerol stocks

Glycerol stocks of transformed bacteria were made for long-term storage. 1.4 ml of the overnight culture was transferred to a microcentrifuge tube and pelleted by centrifugation at 8,000 rpm for 3 minutes at room temperature. The supernatant was removed, and the pellet was resuspended in 0.5 ml of sterile 15% glycerol (v/v in LB broth) and transferred into a cryovial. This was then immediately frozen at -80^{o}C.

2.2.8 Cell Culture

2.2.8.1 Culturing of SH-SY5Y, S-K-NAS, HAP1, K562 cells

Human cell lines; SH-SY5Y neuroblastoma-derived, S-K-NAS neuroblastoma-derived, HAP1 and K562 both chronic myelogenous leukaemia-derived were maintained in culture media outlined in *section 1.6* at 37°C, 95% air, 5% CO_2, in T175 tissue culture flasks until 70-80% confluent. To passage cells, media was removed, and the cells were washed down with 1X sterile PBS (Sigma) pre-warmed to 37°C. Following removal of the PBS, 5 ml of pre-warmed 1X trypsin (Sigma) was washed over the cells and then removed, and the cells incubated at 37°C for approximately 3 minutes or until the cells began to detach from the bottom of the flask. To neutralise the trypsin, cells were washed down in 10 ml of pre-warmed complete tissue culture media and disaggregated into a single cell suspension through pipetting. Between 1-2 ml (approximately 1-2.4 million cells depending on cell type) of the cell, the suspension was then transferred into a new T175 flask with 40 ml of the appropriate media for that cell line. Cell lines were tested for mycoplasma infection every six months using the MycoAlert Mycoplasma Detection kit (Lonza).

2.2.8.2. Cell counts with a haemocytometer

To determine the number of cells per ml of media, cell counts were performed using a haemocytometer. Cells were passaged as described in *section 2.2.8.1* up to the cells being washed down with 10 ml of media. Before the coverslip was placed onto the counting surface of the haemocytometer, both parts were washed with 70% ethanol. On the centre of the counting surface of the haemocytometer, there are 25 squares (5x5) bounded by three parallel lines each containing 25 smaller squares (5x5). To perform the cell count, 20 µl of the cell suspension was introduced under the coverslip and the counting surface was visualised under a light microscope on the 10X objective. The number of cells within the 25 larger squares bounded by three parallel lines was counted including cells touching the top or left-hand borders of the 25 squares and excluding those in contact with the bottom or right-hand border. This area corresponds to 0.1 mm^3, therefore, the number of cells was multiplied by $1x10^4$ (10,000) to give the number of cells in 1 cm^3 which is the equivalent of 1 ml. This gave the number of cells per ml of media used for calculating the density at which the cells were seeded.

2.2.8.3 Freezing cells for storage in liquid nitrogen

For long-term storage, cell lines were frozen in a cryoprotective freezing medium, described in *sections 1.6.5 and 1.6.6* in liquid nitrogen. The cells were grown in T175 flasks until 70-80% confluent and then passaged as described in *section 2.2.1*, except in place of the usual cell culture media, the cells were resuspended from the surface of the flask using 10 ml of cryoprotective freezing medium with continuously pipetting to ensure homogeneity. The cell suspension with the cryoprotective freezing medium was split across cryovials up to a volume of 1.8 ml in each. The cryovials were immediately placed into a Mr Frosty with isopropanol at - 80˚C for 24 hours and then transferred to liquid nitrogen.

2.2.9 Transient Transfections

SH-SY5Y, K562 cells were seeded into 6-well or 24-well plates at approximately 60.000 /400,000 cells per well according to the superficies area and transfected with 1 µg of the plasmid of interest using the TurboFect (Thermo Scientific) transfection reagent following the manufacturer's guidelines. Cells were incubated for 48 hours post-transfection and then lysed, and the Dual-Luciferase Reporter Assay (Promega) was performed following the manufacturer's protocol.

2.2.10 Co-transfection assays

To analyse the effect of over-expression of a transcription factor on reporter gene activity, cells were seeded in 24-well plates at approximately 100,000 cells per well and transfected with 1 µg reporter plasmid DNA and 10 ng TK renilla as internal control, using TurboFect (Thermo Scientific) transfection reagent following the manufacturer's guidelines. Transfected cells were lysed 48 hours post-transfection and reporter gene activity was measured using the Dual-Luciferase Reporter Assay System (Promega).

2.2.11 Reporter gene assay

2.2.11.1 Cellular lysis

At 48 hours post-transfection, tissue culture media was removed from the cells and the cells were washed with 1X PBS. For cellular lysis, 100 µl of 1X passive lysis buffer (PLB) was added to each well of the 24-well plate and the plate was incubated at room temperature on a rocking platform for 15 minutes. 20 µl of the cell lysate was then transferred to an opaque 96-well plate for analysis.

2.2.11.2 Measurement of reporter gene activity

The appropriate amount of luciferase assay reagent II (LARII) and Stop and Glo reagent was prepared for the number of measurements required and allowed to reach room temperature. The opaque 96-well plate containing the cell lysate was placed into a Glomax 96 Microplate Luminometer (Promega) which had been set up under default settings for dual-luciferase reporter gene assays for experiments using two injectors. The injectors were first flushed with distilled water, 70% ethanol, distilled water and air to thoroughly clean them and then primed with the luciferase reagents (LARII in injector 1 and Stop and Glo in injector 2) before the Promega dual luciferase program was run, which measured the bioluminescence from the reaction catalysed by the firefly and *renilla* luciferase enzymes. The LARII was added first to measure the bioluminescence produced by the reaction catalysed by the firefly luciferase protein and then the Stop and Glo quenches this reaction and is used to measure the bioluminescence from the reaction catalysed by the *renilla* luciferase protein.

2.2.11.3 Statistical analysis

Using the measurements recorded for the activity of the two co-transfected reporter gene constructs, the activity of the constructs across the different wells can be accurately compared as the renillin is used for normalisation to control for experimental variability caused by differences in transfection efficiencies. Fold changes in firefly luciferase activity (normalised to *renilla* luciferase activity) supported by the reporter gene constructs over the pGL3 controls were calculated and significance was determined using one-tailed *t*-tests. Significance was scored as follows */# $P<0.05$, **/## $P<0.01$, ***/### $P<0.001$. For each transfection, a minimum of n= 4 was used.

2.2.12 mRNA expression analysis

2.2.12.1 In vitro RNA extraction

Total RNA was extracted using TRIzol reagent (Invitrogen) following the manufacturer's instruction. instructionsP1 and SH-SY5Y cells were plated into 6-well plates at approximately 400,000 cells per well and incubated for 24 hours. The media from each well was removed and 1 ml of TRIzol was added per well and pipetted up and down to lyse the cells. The cell lysate/TRIzol mix was added to a nuclease-free microcentrifuge tube and incubated for 5 minutes at room temperature. To each sample 0.2 ml of chloroform (was added, shaken vigorously by hand for 15 seconds and then incubated at room temperature for 2-3 minutes. Samples were then centrifuged at 12,000 x g for 15 minutes at 4^{o}C for phase separation into three layers: a colourless aqueous upper layer containing the RNA, a middle interphase layer and a lower red organic layer containing the DNA and protein. The upper colourless layer (approximately 500 µl) was carefully removed and transferred into a new microcentrifuge tube and 0.5 ml of isopropanol (molecular grade) to each 1ml of TRIzol reagent was added to each sample and incubated for 10 minutes at room temperature. Samples were then centrifuged at 12,000 x g for 10 minutes at 4^{o}C and the resulting supernatant was removed leaving behind the RNA pellet.

To pellet, the RNA, 1 ml of 75% ethanol was added and the sample vortexed and centrifuged at 7,500 x g for 5 minutes at 4^{o}C. The supernatant was removed and the pellet air dried for 5 to 10 minutes before being resuspended in 20 µl of nuclease-free water and incubated on a heat block at 55^{o}C for 10-15 minutes. The RNA samples were kept on ice for quantification and first-strand cDNA synthesis steps or stored at -80^{o}C for later use.

2.2.12.2 Measurement of RNA concentration by spectrometry

RNA was quantified using a Nanodrop 8000. The Nanodrop was set to the RNA setting and calibrated with nuclease-free water (the solvent the RNA was diluted in). A 1.5 µl aliquot of the RNA sample was loaded onto the pedestal and the absorbance was measured. The amount of UV light absorbed at 260 nm by nucleic acids is dependent on their concentration. The Nanodrop measures the optical density (OD) of the RNA and then calculates its concentration (an OD_{260nm} of 1 equals an RNA concentration of 40 µg/ml). The Nanodrop was also used to assess the quality of the RNA by measuring the 260/280 and 260/230 ratios; expected values for high-quality RNA are approximately 2.0 and 2.0-2.2, respectively. The RNA was then stored at -80^{o}C.

2.2.12.3 First strand cDNA synthesis

cDNA was synthesised from total RNA using the GoScript Reverse Transcription System (Promega) following the recommended manufacturer's protocol. For each sample in the experiment the same amount of RNA was used in the reverse transcriptase reaction, combined in a PCR tube with the following components:

RNA (up to 5 µg)	X µl
Random Primers (0.5 µg/reaction)	1 µl
Nuclease free water	X µl
Final volume	5 µl

The mixture was denatured at 70^{o}C for 5 minutes and then cooled on ice. The following reverse transcription mix was added to the RNA, random primers and nuclease-free water and made up to a final reaction volume of 20 µl:

Component	Volume (n=1)	Final Concentration
Nuclease free water (to a final volume of 15 µl)	X µl	-
GoScript 5X reaction buffer	4 µl	1X
$MgCl_2$ (25 mM)	4 µl	5 mM
PCR nucleotide mix (10 mM of each dNTP)	1 µl	0.5 mM
Recombinant RNasin Ribonuclease inhibitor (40 U/µl)	0.5 µl	1 U/µl
GoScript Reverse Transcriptase	1 µl	-

The reaction mixtures were incubated at 25^{o}C for 5 minutes to allow primer annealing and then incubated at 42^{o}C for 60 minutes for the extension step. The reverse transcriptase was inactivated by heating the reaction to 70^{o}C for 15 minutes. The cDNA was diluted appropriately (if 2 µg RNA was converted, a 1:20 dilution was made) using nuclease-free water and stored at -20^{o}C.

2.2.12.4 Semi-quantitative PCR analysis of mRNA expression

For analysis of gene expression, cDNA generated as previously described was amplified using GoTaq DNA polymerase (Promega) following the manufacturer's guidelines. The GoTaq Flexi DNA Polymerase reaction mix was:

Component	Volume	Final Concentration
5X Green GoTaq Flexi buffer	5 µl	1X
$MgCl_2$ (25 mM)	4 µl	4 mM
PCR nucleotide mix (10 mM of each dNTP)	1 µl	0.4 mM
Forward primer (20 µM)	0.25 µl	0.2 µM
Reverse primer (20 µM)	0.25 µl	0.2 µM
GoTaq DNA polimerase (5U/µl)	0.25 µl	0.05 U/µl
Nuclease free water	X µl	-
cDNA template (1:10 dilution)	1 µl	-
Final volume	25 µl	-

The PCR was performed in a thermocycler: peqSTAR 2X (peqlab) or SimpliAmp (Thermo Fisher). The annealing temperature of each primer set was optimised using a gradient PCR and is detailed in Table 1. The amount of cDNA template used varied between primer sets depending on the abundance of the target for amplification. In general, 1 µl of a 1:20 dilution of cDNA generated from 2 µg of RNA was used for target genes, and a 1:200 dilution of cDNA was used for reference genes. Standard thermal cycling conditions were as follows: incubation at 95˚C for 5 minutes, followed by 25 or 35 cycles (for reference and target genes, respectively) of 95˚C for 30 seconds, 57-65 ˚C for 30 seconds and 72 ˚C for 30 seconds, with a final cycle at 72˚C for 10 minutes. Samples were kept at 4˚C before gel electrophoresis (*section 2.2.5.3*) or at -20˚C for long-term storage.

2.2.13 Bioinformatic Analysis

2.2.13.1 Evolutionary Conserved Regions (ECR) Browser

Conservation of transcription factor consensus binding sequences was addressed based on the ECR Browser (http://ecrbrowser.dcode.org/) using the following parameters: minimum matrix conservation (similarity between the consensus binding site for a transcription factor and a potential binding site in the query sequence), 70%; minimum number of homologous sites (the minimum number of sites of which a matrix is built), 4; factor class level (the classification of transcription factors in the TRASNFAC database is hierarchical and includes six levels, from a family of transcription factors to splice variants), 4; and similarity of the sequence to the matrix, 1. The conservation profiles of the TOMM40 gene were determined using data from UCSC and ECR browser (http://ecrbrowser.dcode.org/)

2.2.13.2 HapMap Genome Browser

SNP genotype data for genomic regions of interest corresponding to individuals from the CEPH trios of European origin were downloaded from the HapMap Genome Browser, release #28 (August 2010, NCBI build 36, dbSNP b126), which can be accessed at http://hapmap.ncbi.nlm.nih.gov/. The data download was used to further Linkage disequilibrium analysis.

2.2.13.3 Haploview

Genotype data were uploaded into Haploview 4.1 (www.broad.mit.edu/mpg/haploview/), freely available software for measuring linkage disequilibrium (LD), defining haplotype blocks and identifying haplotype tagging SNPs (htSNPs) [201]. Under the standard Linkage format, htSNPs were identified using the pairwise-tagging function (r2 threshold, 0.8). SNPs were filtered to include only those with a minor allele frequency (MAF) of greater than 5% within a Caucasian population.

LD analysis was performed using the D-prime (D') statistic, which is derived from the earliest measures of disequilibrium, termed D**.** D quantifies disequilibrium as the difference between the observed frequency of a two-locus haplotype (combination of alleles at adjacent loci on a single chromosome) and the frequency it would be expected to show if the alleles were segregating at random. Adopting the standard notation for two adjacent loci — *A* and *B*, with two alleles (Aa and Bb) at each locus — the observed frequency of the haplotype that consists of alleles *A* and *B* is represented by *PAB*. Assuming the independent assortment of alleles at the two loci, the expected haplotype frequency is calculated as the product of the allele frequency (*P*) of each of the two alleles, or *PA* × *PB*. Therefore, one of the most straightforward measures of disequilibrium is $D = PAB - PA \times PB$, which states the linear relationship between a given pair of markers, where a D' value of 1 represents complete LD. The squared correlation coefficient (r^2), used for htSNP

analysis and, therefore, a measure of LD, is determined by dividing D' by the product of the four allele frequencies. When $r^2 = 1$, this indicates that two markers have equal allele frequencies and are in complete LD (D' = 1).

2.2.13.4 National Center for Biotechnology Information (NCBI)

Sequence alignments were performed using the basic local alignment search tool of nucleotide database (BLASTN) [202], available at NCBI *(http://blast.ncbi.nlm.nih.gov/ Blast.cgi).* BLAST finds regions of local similarity between sequences by making comparisons of nucleotide or protein sequences to sequence databases and calculating the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

2.2.13.5 UCSC Genome Browser

Bioinformatic analysis of humans was performed using the UCSC Genome Browser, assembly Hg19 and Hg38 of the human genome using ENCODE data on the USCS genome browser (http://genome.ucsc.edu/) [203].

2.2.14 Genotyping

2.2.14.1 Genotyping analysis of Variable Number of Tandem Repeats (VNTRs)

PCR reactions were set up amplified using GoTaq DNA polymerase (Promega) or Reddy Mastermix (Thermo Scientific), following the manufacturer's guidelines and with primers Forward and Reverse and annealing temperature detailed in *section 2.1.7*.

Components	Volume	Final concentration
template DNA	2µl	10ng
5x green goTaq® flexi Buffer	5µl	1x
MgCl2 Solution	4µl	1mM
PCR Nucleotide Mix	1µl	0.1mM each dNTP
Upstream primer	0.5µl	5pmol
Downstream primer	0.5µl	5pmol
Go Taq® DNA polymerase	0.25µl	0.625U
Nuclease-free water to	up to 25µl	

The PCR was performed in a thermocycler: QB-96 (Quanta Biotech) or peqSTAR 2X (peqlab). The annealing temperature of each primer set was optimised using a gradient PCR and are detailed in Table 2.1. The amount of DNA template used varied between primer sets depending on the abundance of the target for amplification. Standard thermal cycling conditions were as follows: incubation at 95°C for 5 minutes, followed by 25 or 35 cycles (for reference and target genes, respectively) of 95°C for 30 seconds, 57-65 °C for 30 seconds and 72 °C for 30 seconds, with a final cycle at 72°C for 10 minutes. Samples were kept at 4°C before gel electrophoresis, detailed in *section 2.2.5.3* or at -20°C for long-term storage.

2.2.14.2 Genotyping analysis of dVNTR of MAOA gene

PCR reactions were set up Go Taq® DNA polymerase (Promega). The master mix is:

Components	Volume	Final concentration
template DNA	2µl	10ng
5x green goTaq® flexi Buffer	2.5µl	1x
MgCl2 Solution	2µl	1mM
PCR Nucleotide Mix + Deaza*	0.5µl	0.1mM each dNTP
Upstream primer	0.25µl	5pmol
Downstream primer	0.25µl	5pmol
Go Taq® DNA polymerase	0.125µl	0.625U
Betaine Solution	2.5µl	1M
DMSO	0.625µl	
Nuclease-free water to	12.5µl	

*Deoxyribonucleotide triphosphate (dNTP) and 7-Deaza-dGTP, was prepared as below:

dNTPs Mix	%	Concentration
−N7-dGTP, 7-Deaza-dGTP	12.5	1.25mM
dATP	12.5	12.5mM
dCTP	25	25mM
dGTP	25	25mM
dTTP	25	25mM

The touchdown thermal cycling conditions were:

Step	Number of Cycles	Temperature	Time
Initial denaturation	1	95ºC	2'
Denaturation		95ºC	20"
Annealing	10	65ºC - 55ºC	20"
Extension		72ºC	30"
Denaturation		95ºC	20"
Annealing	35	55ºC	20"
Extension		72ºC	30"
Final extension	1	72ºC	5'
Soak		4ºC	∞

PCR products were analysed using both agarose gel electrophoresis, **2.2.15.5** described in *section 2.2.5.3* on a 2% agarose gel incorporating GelRed™ (Biotium). To validate the protocol, the PCR product was also analysed by capillary electrophoresis (ABI 3130, Life Technologies), as described in *section 2.2.15.5*. And they were using the Genemapper V4.0 (Life Technologies) for the genotype calls. The results from both methods were analysed blind from each other. PCR products were analysed using agarose gel electrophoresis or QiaXcel system, described in *sections 2.2.5.3 and 2.2.15.7.*

2.2.14.3 Genotyping analysis of DRD4 ex3, STin2 and uVNTR of MAOA gene.

The VNTRs DRD4ex3, STin2 and uVNTR PCR reactions were set up with Go Taq® DNA polymerase (Promega). The master mix were:

Components	**Volume**	**Final concentration**
Components	Volume	Final concentration
template DNA	2µl	10ng
5x green goTaq® flexi Buffer	2.5µl	1x
MgCl2 Solution	2µl	1mM
PCR Nucleotide Mix	0.5µl	0.1mM each dNTP
Upstream primer	0.25µl	5pmol
Downstream primer	0.25µl	5pmol
Go Taq® DNA polymerase	0.125µl	0.625U
Betaine Solution	2.5µl	1M
Nuclease-free water to	12.5µl	

Touchdown thermal cycling conditions were used to the DRD4ex3, STin2, as described in section 2.2.14.1. To the uVNTR, the thermal cycling condition is described below. PCR products were analysed using agarose gel electrophoresis or QiaXcel system, described in *sections 2.2.5.3 and 2.2.14.7.*

Step	Number of Cycles	Temperature	Time
Initial denaturation	1	2’	95ºC
Denaturation		20"	95ºC
Annealing	35	20"	61ºC
Extension		30"	72ºC
Final extension	1	5'	72ºC
Soak	1	∞	4ºC

2.2.14.4 Genotyping analysis of 5HTTLPR

Genotyping of 5HTTLPR were performed using Reddy Mastermix (Thermo Scientific) following the manufacturer’s guidelines with primers Forward and Reverse detailed in *section 2.1.7*. Furthermore, using the touchdown cycling conditions detailed in section 2.2.14.2. PCR products were analysed using agarose gel electrophoresis or QiaXcel system, described in *sections 2.2.5.3 and 2.2.14.7.*

2.2.14.5 Genetic Analyser capillary electrophoresis

The MAOA distal VNTR was genotyped on an ABI PRISM 3130XL Genetic Analyser (Applied Biosystems) capillary electrophoresis platform followed DNA amplification was performed using 5 ng genomic DNA following the protocol described in *section 2.2.5.5* and the human MAOA distal VNTR primer set detailed in Table 1. The reverse primer was synthesised with a 5′-terminal 6-carboxyfluorescein (FAM) (Eurofins MWG Operon). For analysis on the 3130XL platform, 2 µl of PCR product was subjected to capillary electrophoresis against an internal size standard (GeneScan 500 ROX) (Applied Biosystems) following the manufacturer's protocol. Following manufacturer protocol, the appropriate run module was selected for a 36cm array length and the POP-7 Polymer (Applied Biosystems).

Fragment analysis was performed using the GeneMapper Software version 4.0 (Applied Biosystems).

2.2.14.6 Gel electrophoresis

The same PCR product of section 2.2.15.1 was run using gel electrophoresis to validate the product sizes in agarose gel. The PCR products were separated on a 2% agarose gel, as described in section 2.2.5.3. Expected fragment sizes were 365 ±10 bp. When applicable, duplicate samples were tested, and negative controls were included on each 96-well plate. Genotyping was performed blind to age and gender. Statistical analysis for genotype data is detailed in section 2.2.18.2.

2.2.14.7 QIAxcel Advanced System

The *MAOA* dVNTR and µVNTR, *DRD4, STin2, 5HTTLPR* VNTRs PCR products were analysed on a QIAxcel Advanced System (Qiagen), which is a benchtop multichannel electrophoresis platform. DNA amplification was performed using 5 or 10 ng genomic DNA following the protocol described in section 2.2.14.1 to 2.2.14.3 and the human VNTR primers and HUWE SVA E CT set detailed in Table 1. For analysis on the QIAxcel Advanced System platform, 20 µl of PCR product was subjected to capillary electrophoresis following the manufacturer's protocol. The appropriate run method was the 0M500 for all VNTRs using the QIAxcel DNA High-Resolution Gel kit, the choice of which was determined by the range of the fragment sizes and resolution. The QIAxcel DNA high-resolution gel cartridge is designed for high-resolution 3bp to 5 bp size differences. The gel cartridge can separate fragment sizes ranging from 15 bp to 10 kb. For the runs, the 15/600bp alignment marker was chosen in association with the 25/500bp ladder.

Fragment analysis was performed using the Qiaxcel integrated software version 4.0 (Qiagen) and validated using gel electrophoresis. For gel electrophoresis, PCR products were separated on an agarose gel by the fragment size, as described in section 2.2.5.3.

The range of expected fragment sizes is shown in table 1. Controls for each genotype were included on each 96-well plate. Genotyping was performed blind to age and gender. Statistical analysis for genotype data is detailed in section 2.2.17.2.

2.2.15 Chromatin Immunoprecipitation (ChIP)

2.2.15.1 In vitro ChIP

Cells were grown to 80% confluence in T175 flasks and treated for 1 hour under one of the following conditions: basal (untreated), 1 or 5 mM Valproate. Samples were processed following methods described by Murgatroyd et al. (2012). ChIP buffers are listed in section 2.1.2. Immunoprecipitation was performed using ChIP grade antibodies detailed in Table 2.3. PCR analysis of the immunoprecipitated chromatin samples was performed using primers detailed in Table 2.1.

For each immunoprecipitation (IP), 5 µg of the sheared chromatin was made up to 250 µl using ChIP dilution buffer, supplemented with 10 µl/ml 100X PIC and incubated overnight at 4°C on a rotating wheel with antibodies raised against histone H3, H3-TriMetK9, NRSF, hnRNPK, CNBP, Nucleolin, SP1 and CTCF. The protein–DNA complexes were added to 40 µl DynabeadsTM (Thermo Scientific), which were first pre-cleared by washing twice with 1 ml ChIP dilution buffer supplemented with PIC; the second wash step was left for 2 hours, and then incubated on a rotating wheel for 1 hour at 4°C. The magnetic Dynabead™/protein–DNA complexes were captured by placing the tubes in a magnetic rack for 1 minute to separate the beads from the solution and the

supernatant discarded. Bound magnetic beads were subjected to 5 minutes of washing steps (performed in a 4°C cold room) with rotation to remove non-specific DNA and proteins associated with the Dynabeads™. The beads were washed with 1 ml of low-salt wash buffer, followed by high-salt wash buffer, LiCl wash buffer and finally TE buffer. The immune complexes were eluted by adding 50 µl of an elution buffer containing 50 µg/ ml proteinase K to the magnetic bead/protein-DNA complexes and the supernatant was transferred to a new microcentrifuge tube and mixed at 65°C for 2 h to reverse the cross-linking and release the protein-bound DNA. The samples were then incubated at 95°C for 10 min to denature the proteins and inactivate the proteinase K, and the DNA recovered from the sample by spin column purification using the Wizard® SV Gel and PCR Clean-Up System (Promega) eluted in a volume of 20 µl. The DNA was quantified using a Nanodrop 8000 and analysed by PCR section 2.2.13.2.

2.2.15.2 Methylated DNA Immunoprecipitation (MeDIP)

Methylated double-stranded DNA was isolated from genomic DNA samples using the CpG Methyl Quest DNA Isolation Kit (Merck Millipore), following the manufacturer's instructions. Briefly, 300 ng (recommended concentration) of genomic DNA, which has been sheared into the appropriate fragment size using a Bioruptor® (Diagenode), with 10 cycles (30 sec ON/OFF) to obtain a 500bp average size DNA. The sheared DNA was incubated for 1 hour at room temperature on a rotating wheel with 5 µl of pre-cleared CpG MethylQuest glutathione paramagnetic beads, which were pre-coupled to a GST (glutathione-S-transferase protein)-MBD (methyl binding domain) fusion protein that specifically binds methylated double-stranded DNA. Methylated sequences bound to the CpG Methyl Quest fusion-protein/bead complex were subjected to wash steps and the supernatant containing the non-methylated DNA was kept for comparison. The methylated DNA was eluted from the beads by heating the samples at 80°C for 10

minutes with mixing in 100 µl TE buffer. The sample was separated from the beads by placing the tubes in a magnetic rack, and the supernatant was transferred to a new microcentrifuge tube, and the beads were discarded. Samples were subjected to PCR analysis and then stored at -20°C.

2.2.16 Statistical Analysis

2.2.16.1 Clump analysis

Significance-testing of allele frequency and genotype data for the MAOA distal VNTR and µVNTR between cases and controls of the schizophrenia cohort (*Materials section xxx*) was performed using *Clump 24* analysis software which can be accessed from http://www.davecurtis.net/dcurtis/software.html. The *Clump* program assesses the significance of the departure of observed values from the expected values using a Monte Carlo-based approach. It does this by performing repeated simulations (10,000) to generate contingency tables (2 x N) with the exact marginal totals as the one under consideration and counting the number of times the randomly simulated data achieves a chi-squared value associated with the actual table. The *Clump* software also generates a novel chi-squared value (T4) by 'clumping' columns together into a new two-by-two table in a way which is designed to maximise the chi-squared value. It directly tests the hypothesis that several alleles are more common among the cases than among the controls [204]. The *Clump* program generates four test statistics using Monte Carlo methods to evaluate the significance of chi-squared values by assessing how often the observed value produced is exceeded by chance from the randomly generated simulated datasets. The four tables are as follows:

T1: The raw 2-by-N table supplied by the user.

T2: The original table with columns containing small numbers 'clumped' together.

T3: The most significant of all 2-by-2 tables obtained by comparing each (non-rare) column of the original table against the total of all the other columns.

T4: A 2-by-2 table obtained by 'clumping' the columns of the original table to maximise the chi-squared value.

2.2.16.2 Hardy-Weinberg Equilibrium (HWE)

As a measure of quality control in our VNTR association study, we tested our genotype data for departure from HWE, as this can be used as an indicator of genotyping errors and population stratification. For analysis of HWE, we used the *Hardy-Weinberg equilibrium calculator* (http://www.oege.org/software/ hwe-mr-calc.shtml), which implements the Pearson chi-square (X^2) test statistic to assess the goodness-of-fit of the observed genotype frequency against the expected under HWE [205].

Chapter 3

Gene expression and genetic risk factors within non-coding regions

The work in the following section has been published in the Journal Neurobiology of aging (DOI: 10.1016/j.neurobiolaging.2015.11.017)

3.1 Introduction

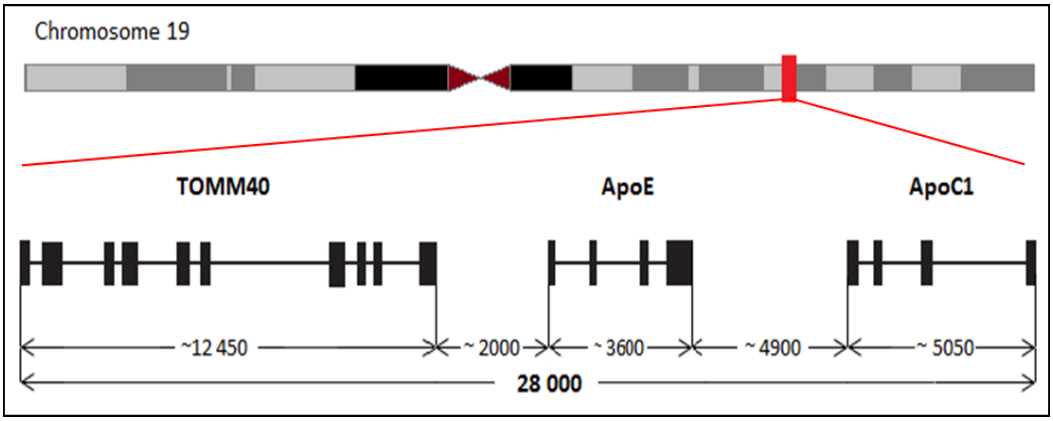
Dementia can cause loss of cognitive function, affecting vital domains, such as memory, understanding, thinking, orientation, learning and judgment. It also can trigger neuropsychiatric symptoms, irritability, anxiety, agitation, depression, and even hallucinations. Alzheimer's disease (AD) is by far the most common form of dementia worldwide, representing over 60-70% of all cases, with the number of affected individuals surpassing 55 million [206]. The ever-increasing life expectancy of humankind only adds to the fact that AD constitutes one of the greatest public health challenges of our century.

There is very compelling evidence that AD has a strong genetic basis with the most robust associations with the genes encoding Apolipoprotein E (*APOE*) and Translocase of Outer Mitochondrial Membrane 40 Homolog (*TOMM40*), referred to as *TOMM40-APOE* locus [207], which corresponds to an 18 kb region within chromosome 19 (Figure 3.1).

TOMM40 encodes a protein of the translocase of the outer mitochondrial membrane (TOM) group, which is involved in transporting and sorting proteins across the mitochondrial membrane [131]. Disruptions of the TOM complex may lead to undesirable mitochondrial stress responses and impaired mitophagy (a mitochondrial quality control mechanism), with detrimental consequences to the integrity of the neurological system [67]. Due to its close genomic proximity, TOMM40 is in linkage disequilibrium with APOE and interestingly with another related gene in the vicinity, the Apolipoprotein C1 (APOC1), spurring further investigative interest of this gene cluster (*TOMM40-APOE*-APOC1) and their functional interactions. Hence, this work also examines the potential involvement of APOC1 in modulating AD risk. Since APOC1 relates directly to the cholesterol transport and metabolism, whose activity extends beyond the regulation of plasma level of lipids, the gene is likely to impact on inflammation, infection and immunity

processes, crucial to both cardiovascular homeostasis and neurological functioning. Like APOE, APOC1 is also a member of the apolipoprotein family and a plausible candidate gene for AD studies. Despite their potential gene-gene interactions, by acting on specific lipoprotein receptor and inhibiting binding mediated by APOE, *APOC1* may also impair cognitive function independently to *APOE* expression [208].

Figure 3.1. TOMM40-ApoE-ApoC1 locus. Representation of chromosome 19 and the relative position of TOMM40, ApoE and ApoC1 genes. The black rectangle represents an exon, whilst the number corresponds to the fragment size in bp.



As mobile elements, retrotransposons are capable of replicating by reverse transcription (RT), recruiting an intermediate RNA to reinsert into a new genomic location [67]. Since Alu elements have been linked with the onset of neurodegenerative disorders, for instance via disturbance of the neuronal mitochondria homeostasis [67], this chapter focuses on *in silico* mapping and study of retrotransposons within the TOMM40-APOE-APOC1 gene locus. Of particular attention is the investigation of a SINE-rich genomic position, consisting of several Alu elements between exons 6 and 7 of the *TOMM40* gene. This region has attracted attention with a particular interest in a block of six adjoining SINEs, namely: AluSx - AluYc3 - AluJb – AluJo - AluJb and FLAM_A. The latter encompasses the poly-T variant rs10524523 (hereafter '523') that has been the

subject of growing investigations. Furthermore, this investigation has been extended to include *APOC1*, an adjacent gene to APOE, in order to address the contribution of a SINE-VNTR-Alu type D (SVA_D) within a human-specific locus, which is located downstream to APOC1 [130].

3.2 Aims

1. To address the contribution of retrotransposable elements within the TOMM40-ApoE-ApoC1 locus and to identify how it can impact the modulation of gene expression.

2. To examine a poly-T variant (rs10524523, ‘523’) located at intron 6 in TOMM40 and whether it may influence gene expression.

3. Identify and characterise retrotransposable elements in association with cognitive ageing.

4. To assess the potential functional effects of the polymorphism of poly-T ‘523’ using reporter gene assays in an S-K-NAS neuroblastoma cell line model.

3.3 Results

Retrotransposable elements within TOMM40-ApoE-ApoC1 locus in the modulation of gene expression and its association with cognitive ageing.

3.3.1 Bioinformatic analysis of the TOMM40-ApoE-ApoC1 locus.

For the purpose of comparison and consistency with experimental work already concluded, and in line with the other sections of this work, the *in silico* analysis presented in this chapter was conducted using the hg19 version of the human genome, available on the UCSC genome browser (http://genome-euro.ucsc.edu/index.html). To analyse the conservation profile here of the *TOMM40-ApoE-ApoC1* gene cluster, the mainstream UCSC and ECR browsers were employed to explore this locus at chromosome 19, whose confirmed that these genes were interspersed with several regulatory elements, particularly non-long terminal repeats (LTRs) of the LINE and SINE classes, with several of them duplicated (Figure 3.1). Examination of the intron 6 region of TOMM40 using the Repeat Masker track option of the UCSC browser, revealed a SINE-rich block structured by several Alu elements. Zooming in the area of interest of the study, it was possible to confirm little conservation of the SINE block across several species, including primates. Interestingly, most of this region, including the polymorphism contained within the T-rich stretch, has been found to be human-specific. Furthermore, the SVA D transposon, present only in humans, was located in close proximity to the APOC1 gene, approximately 1.7 kb downstream (Figure 3.2).

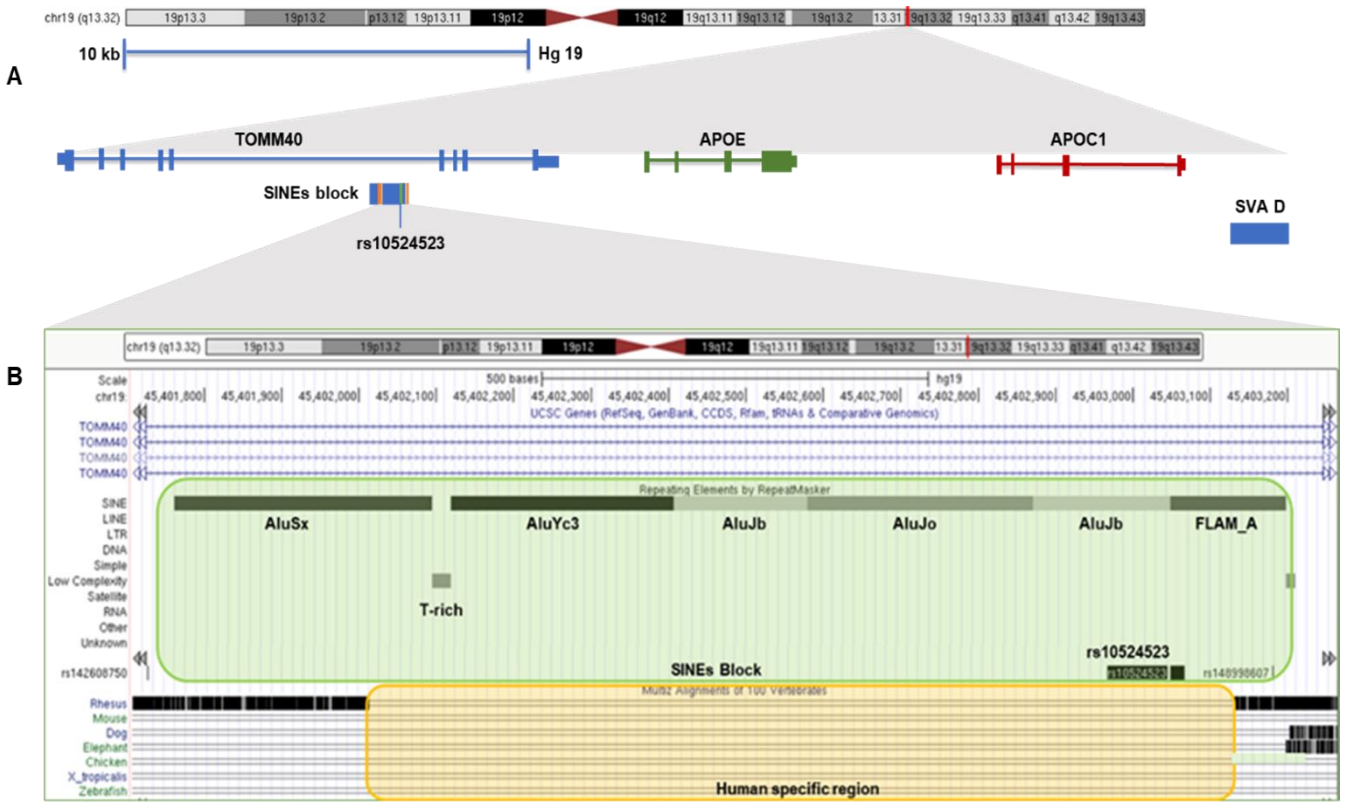


Figure 3.2. Schematic representation of the TOMM40-ApoE-ApoC1 locus and SINEs block. A) Simplified diagram of the TOMM40-APOE-APOC1 gene locus within Chr19q13.32, depicting the exon-intron structure and the presence of the SINE block alongside the key poly-T polymorphism rs10524523 within intron 6 of TOMM40, as well as the SVA_D downstream position to APOC1 (approximately 1.7Kb); **B)** Magnified illustration of the SINE block using the UCSC browser (tracks labels are located on the left hand side of the image). The Alu blocks are highlighted in green, and within the region of interest each Alu element is individually identified, including FLAM_A alongside rs10524523. The conservation plot indicates little conservation of this block across species.

3.3.2 The SINE block and sequence polymorphism rs10524523 within *TOMM40*

TOMM40 was mapped to chr19:45,394,477-45,406,946 (hg19), spanning 12,470 bp with a total of 10 exons. Within the intron 6 of this gene, a SINE block spanning 1,451 bp was found at chr19:45,401,766-45,403,217, whereas the variable length poly-T sequence polymorphism rs10524523, namely '523' (chr19:45403049-45403083), presented a T-stretch of 35 bp according to the reference genome of the UCSC browser. The marker has been mapped towards the 3' end of the SINE block (Figure 3.3). This SINE block region contains other two T-rich stretches located in the positive strand, with the first comprising 25 bp and the other 11 bp, both potentially polymorphic. Conversely, Alu domains frequently possess long poly-T stretches or T-rich regions, which are more commonly positioned in the negative strand. The genomic location of the 6 Alu elements found in the SINE block and their respective strand positions are given in Table 3.1.

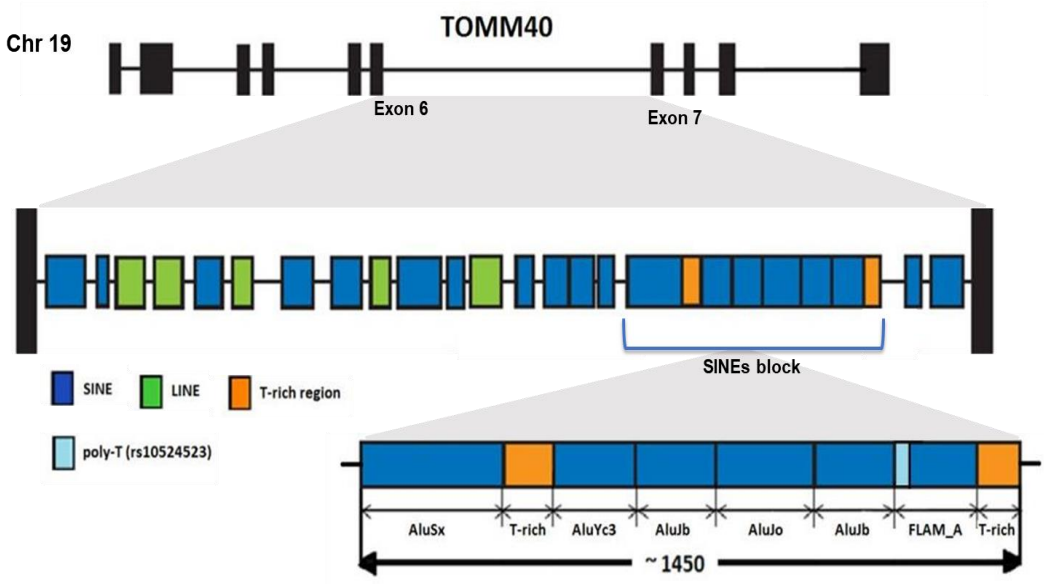


Figure 3.3. Retrotransposon elements between exons 6 and 7 of TOMM40. Representation of LINE (green rectangle), SINE (blue rectangle) and T-rich region (orange rectangle). Further down the figure shows a magnified representation of the genomic structure of the SINEs block within intron 6. The sky blue rectangle corresponds to the poly-T ("523") within the FLAM_A element.

Table 3.1. Genomic location of Alu elements within the SINEs block of TOMM40, respective length and strand orientation. Source: UCSC genome browser, hg19.

SINEs	Start	End	length (bp)
AluSx	45401762	45402094	333
AluYc3	45402120	45402407	288
AluJb	45402408	45402579	172
AluJo	45402580	45402871	292
Alujb	45402872	45403048	177
FLAM_A	45403049	45403198	150

Within the clustered genes *TOMM40-ApoE-ApoC1* there are other non-LTR elements, such as LINE, SINE and SINE-VNTR-Alus (SVA). The distinct structure of these non-LTR retrotransposons is demonstrated in Figure 3.4. Those 3 types of retrotransposons combined constitute approximately 1/3 in length of the 3.2 billion nucleotides of the human genome [36, 131].

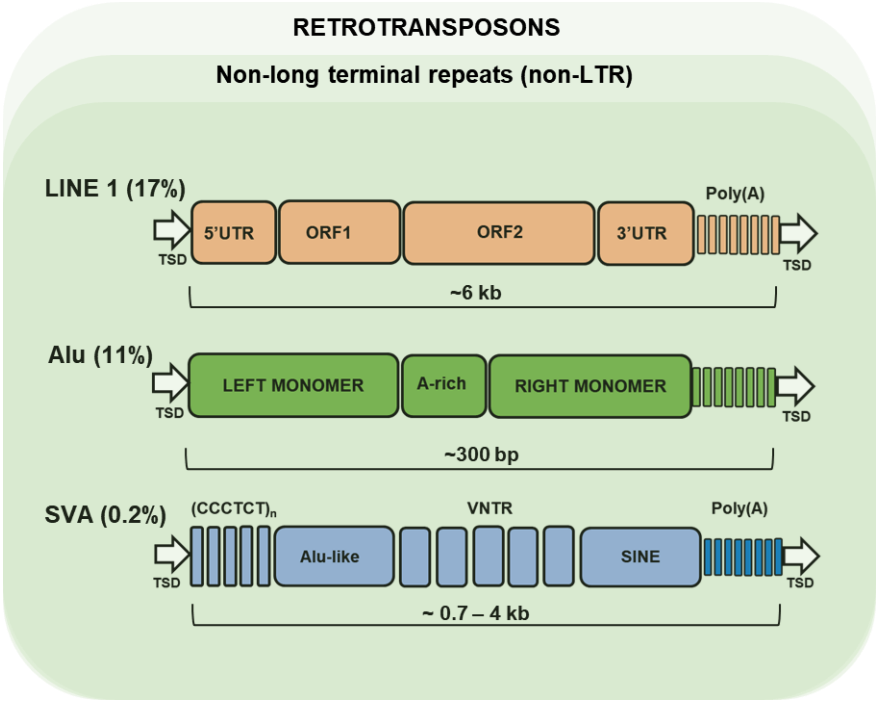


Figure 3.4. General representation of 3 classes of non-LTR human retrotransposons. The non-LTR retrotransposons are structurally distinct, but all share a poly-A tail. LINE-1 (L1) or the autonomous LINE-1 element contains two open reading frames, ORF 1 and 2, as well as typical 5' and 3' UTRs and finally a poly-A tail. A full length L1 spans around 6 kb. On the other hand, Alu comprises two monomers (namely left and right) linked by an A-rich linker sequence (A5TACA6) and a poly-A tail; whose overall length is significantly

smaller (~300 bp). SVA ranges between 700 bp and 4 kb and is a composite element made up of (CCCTCT)n hexamer simple repeat region in the 5′ end, followed by an Alu-like region that is generally composed by two antisense Alu fragments, a variable number of tandem repeat (VNTR) region, a SINE region and a poly(A) tail next to a putative polyadenylation signal (AATAAA). All non-LTR elements are flanked by small target site duplications (TSDs indicated by arrows), typically ranging between 5 and 10 bp.

3.3.3 The SVA D candidate localised downstream to the APOC1 gene

SVAs, another particular focus of this project, are non-autonomous, primate-specific and the evolutionary youngest retrotransposons, constituting 0.13% of the genome. They are divided into A to F subclasses based on their SINE region and estimated age[209, 210]. Some subtype members of A-D, and subclasses E, F and F1 are not shared with other primates and are, therefore, human-specific [103].SVAs are characterised by their complex organisation, as shown in Figure 3.4, and consist of 5' CCCTC-mers, pentamer repeats, followed by an Alu-like sequence, a GC-rich VNTR and a SINE, the latter typically an Alu sequence [211].. This makes their *in vitro* investigation challenging, frequently requiring the development of custom sets of oligonucleotides and protocol adjustments.

As previously highlighted, the human TOMM40/APOE/APOC1 locus is located at position 19q13.32, in a gene-rich cluster. Located within ~5 kb of APOE, APOC1 is a small gene extending 5.0 kb and containing 4 exons, whose encoded transcript generates a precursor of 83 amino acid [212]. Adjacent to APOC1, there is a pseudogene located downstream named APOC1P1, with their intergenic position containing regulatory elements, including a non-LTR element of the SVA D class, which is a subject of investigation by this project.

Since SVAs are believed to be trans-mobilised by the protein machinery of LINE-1 [213], coupled with previous *in vitro* and *in vivo* work by our group reinforced the basis that some SVAs can indeed be relevant transcriptional regulators [209].

Hence, it is plausible that the SVA D as mentioned earlier may possess a regulatory role in which a polymorphism may drive differential expression of the TOMM40-APOE-APOC1 gene complex, particularly of the latter one.

3.4 Functional work and measuring luciferase expression directed by '523' reporter gene constructs

Although original genetic mapping of individual human genes at 19q13.32, as well as at several other gene-dense locations, was achieved by *in vitro* studies on hybrid murine-humanised cells, the existence of clustered genes, such as seen for TOMM40-APOE-APOC1, underpins a biological basis for a concerted regulation of these genes [61]. Given that a wide range of genetic variability within the TOMM40-ApoE-ApoC1 locus has been implicated with age-related cognitive decline, we sought to conduct functional genomic work on candidate retrotransposons likely to drive the expression of these clustered genes. Hence, a luciferase reporter gene assay was carried out, focusing on the length polymorphism within the TOMM40 FLAM_A segment containing alternative alleles of the rs10524523. For the SVA D located downstream to the ApoC1 gene, the experiments employed two well-established human neuroblastoma cell lines: SK-N-AS and SH-SY5Y, both widely employed to represent the neuronal function.

As mentioned previously, the genomic coordinates to the SINEs block were employed to gather target sequences and then primer pairs were computationally designed to specifically amplify the SINE block containing rs1052452 within *TOMM40*, using human genomic DNA control as a template. Due to the complexity of this locus, a series of optimisation steps was conducted to achieve specific amplification through the use of an enhanced protocol combining extended touchdown PCR (TD-PCR) cycles with the use of a robust PCR mastermix containing a proofreading polymerase suitable for high GC content (Phusion

polymerase, NEB). The TD-PCR products were cloned into pGL3 promoter (pGL3p) and pGL3 control (pGL3c) vectors (Promega) and named TOMM40/pGL3p and TOMM40/pGL3c respectively, as previously described in the previous section.

3.4.1 The fragment variant of rs10524523 (poly-T stretch '523)

Using the primers described in Table 1 in section 2.1.7, a human genomic DNA template was amplified employing the optimised protocol detailed in section 2.2.5.2, adopting the following cycling conditions, as described below in Table 3.2.

Table 3.2: **Enhanced cycling conditions using TD-PCR amplification for the three variant alleles of '523'.** These PCR products were subsequently used for the cloning protocol.

	PCR Steps	Temperature (°C)	Time	Number of Cycles
	Initial denaturation	95°C	30 seconds	
Cycling conditions Stage 1	Denaturation	95°C	10 seconds	**10 cycles**
	Annealing	68°C - 58°C (-1°C/per cycle)	30 seconds	
	Extension	72°C	1 min 30 sec	
Cycling conditions Stage 2	Denaturation	95°C	10 seconds	**25 cycles**
	Annealing	58°C	30 seconds	
	Extension	72°C	1 min 30 sec	
	Final extension	72°C	5 minutes	
	Holding	4°C	∞	

The sequence was a repetitive region with high CG content, and our approach was to use the touchdown PCR. TD-PCR is a modification of conventional PCR. Through the use of a higher annealing temperature In the first stage and sequentially decreasing it by 1°C in every successive cycle for the first 10 PCR cycles, the aim is to attain a high specific amplification even at the expense of lower

yields. Therefore, the aim is to maximise specific amplification of the target region to overcompete unwanted amplifications in the first initial cycles, Figure 3.5. That stage facilitates the generation of amplicon products of the desired target, as it delivers the free 3'-OH ends to the Taq DNA polymerase. These techniques are proposed to avoid the competition of non-specific binding after the amplification of the first template, increasing the likelihood of amplifying the sequence of interest, which subsequently will work as a more competitive template for the subsequent PCR cycles of stage two.

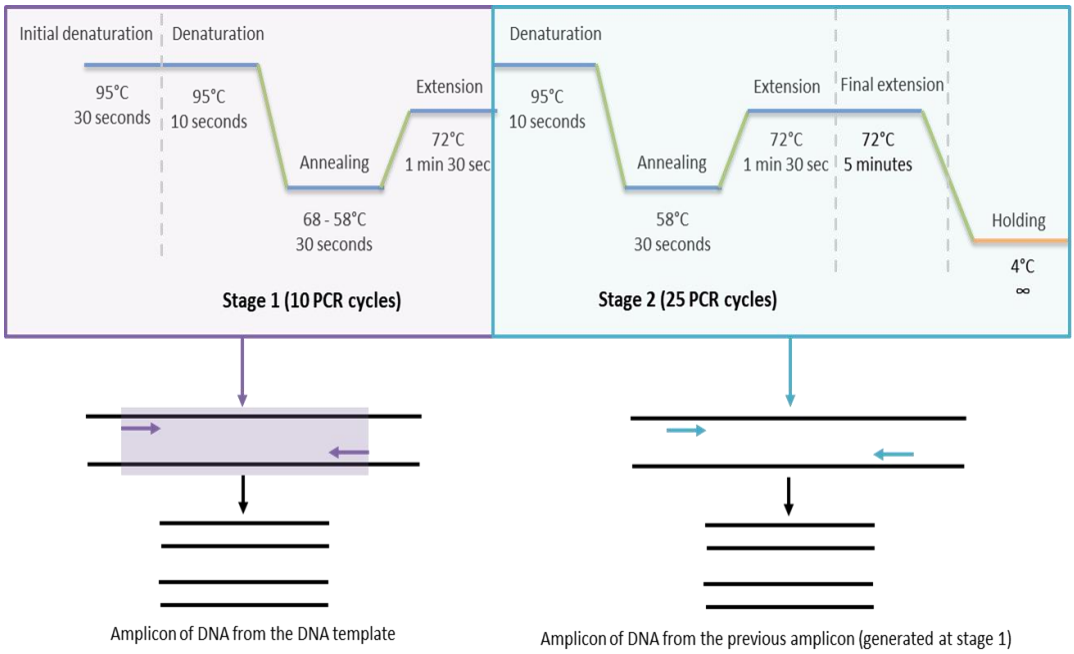


Figure 3.5. Schematic representation of TD-PCR. Shows the stage one and e two of TD-PCR amplification cycles to generate amplicons for difficult genomic targets, such as the polymorphism within the SVA block.

To prevent background contamination arising from the amplification process, PCR products were run in an agarose gel and the target fragment size in the range of 160-190 bp selected for a set of genomic samples predicted to contain the three reported allele variants for this polymorphism. The size-selected purified fragments from the *TOMM40* gene poly-T region and FLAM_A were sent for Sanger sequencing to confirm their exact length and composition, confirming the existence of 3 distinctive length allele variants which were named as following: i) FLAM_A17T; ii) FLAM_A34T; iii) FLAM_A41T, where the numbers given to each allele reflecting the nucleotide length of their T-stretch. As previously stated, we attempted a stepdown approach of the whole SINEs block to rationalise the most promising segments for the functional genomic evaluation. In addition to data sharing collaboration to facilitate the characterisation of the key alternative alleles and guide selection of the genomic samples required for the functional work, this involves extensive *in silico* custom design and evaluation of a number of primer sets to achieve targeted PCR amplification of several position flanking: i) the whole SINEs block; ii) Individual Alu/poly-T rich elements; iii) the positions in the vicinity of the poly-T '523 at AluJb and FLAM_A/ T-rich element in the 3' end of the block (Figure 3.6). Selected PCR primer sets for cloning were revised for the inclusion of a specific isothermal tag oligosequence to introduce a flanking sequence containing a compatible 6bp-restriction enzyme site in the backbone of the vector, as described in section 2.1.8.

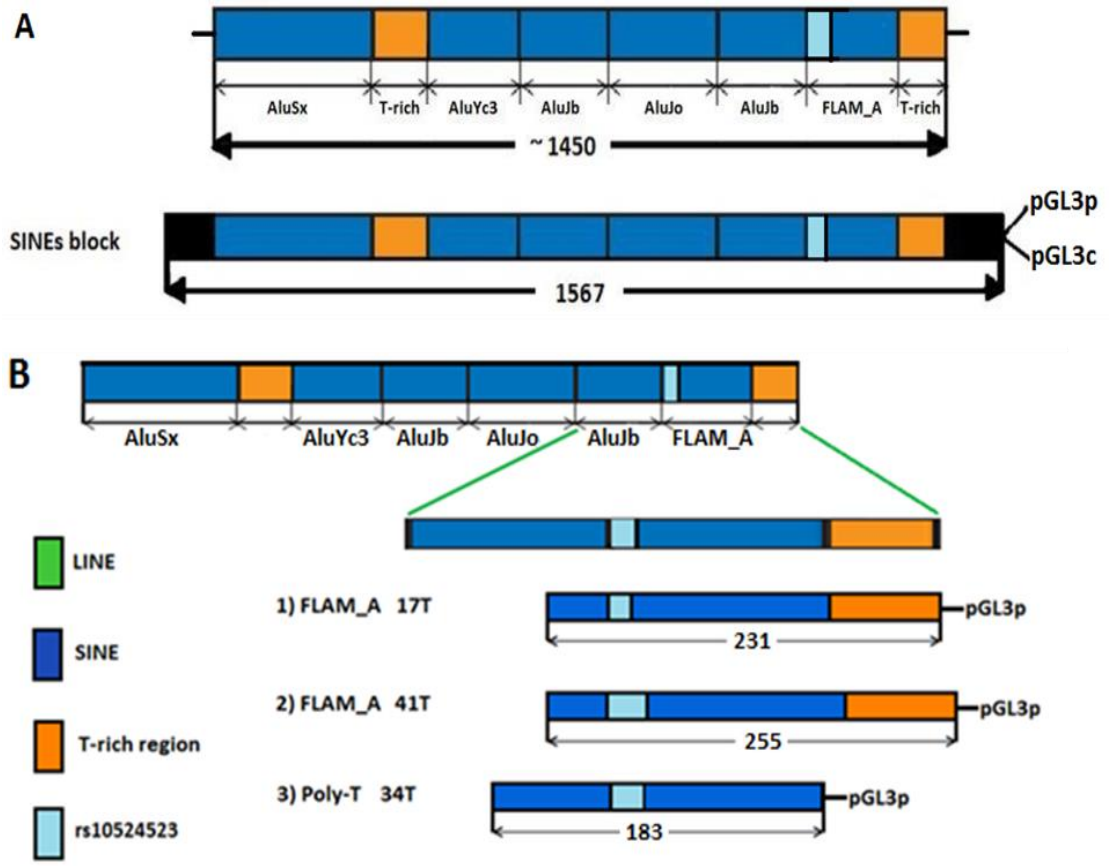


Figure 3.6. Representative illustration of the genomic structure of the SINEs block within TOMM40 and amplified segments that we employed for the characterisation and in the reporter gene constructs. A) Representation of the genomic structure of the SINEs block within TOMM40 spanning 1,450 bp (according to the in-silico analysis of a reference genome) and the trimmed amplified fragment of one of the selected samples (following restriction enzyme digestion) spanning 1,567 bp), which was inserted into the pGL3 promoter (pGL3p) and control (pGL3c) vectors for the initial functional work. B) Schematic representation of the alternative inserts containing the '523' (sky blue) alleles that were cloned into pGL3p vectors. Numbers correspond to the fragment size in bp.

3.4.2 Cloning of the fragments into pGL3 control and promoter vector

The amplified TD-PCR products were purified using a Wizard Clean-up kit (Promega). Prior to the isothermal assembly, the pGL3 promoter (pGL3p) vector was digested with *SmaI* overnight at room temperature. The cut pGL3p vector and the appropriate fragments were mixed in a pmoles ratio of 1:3. The isothermal assembly was performed using Gibson Assembly Master Mix (NEB) and was done according to the manufacturer's instructions except for the incubation time, which was reduced to 15 minutes. The fragment containing the whole SINEs block (six Alu/FLAM_A and two T-rich regions), was amplified and cloned into pGL3p and pGL3 control (pGL3c) vectors. The transformation was performed using *E. coli* cells, XL-10 Gold ultracompetent cells (Agilent Technologies), using a volume of 2 µl of the 1:4 diluted isothermal assembly products and followed the manufacturer's protocol, except for the following: 30µl aliquots of the cells and 1.2 µl of β-mercaptoethanol were used, instead of the suggested 100 µl and 4 µl respectively.

3.4.3 Selection of the Plasmid DNA

Purification of plasmids DNA from transformed *E. coli*, described in section 2.2.5.8, was achieved with 1ml of the 3ml overnight culture using a miniprep purification system (Qiagen). The purified DNA was used to carry out a double restriction enzyme digestion, with *BglII* and *MluI*, overnight at room temperature.

Following the double digestion to determine the presence or absence of the insert, an aliquot of the resulting product was loaded into a 1.2% gel agarose containing ethidium bromide and subject to electrophoresis in 1X TBE with visualisation of the bands in a UV transilluminator. The miniprep cultures containing plasmid DNA of interest were then selected to set 100ml Maxi-prep bacterial cultures (Qiagen) to meet the yield of purified plasmid DNA necessary to perform the report genes assays. The DNA extracted was quantified by spectrophotometry (Nanodrop 8000, Thermo Scientific), and then normalised before storage at -20°C.

3.4.4 Sanger sequencing of the allele variant inserts for the poly-T stretch ‘523

For quality control purposes and to confirm size, content and the allele variant of the inserts selected, Sanger sequencing was carried out by Source Bioscience Sequencing (Rochdale, UK). The sequencing primers selected were from the conserved promoter position of pGL3p vector and are outlined in Section 2.5.12. The results obtained from the inserts were then used to generate consensus sequences, which were grouped for each of the 3 allele variants documented for the T-stretch ‘523 polymorphism.

It is important to note that, although we did observe 3 poly-T variant alleles, in contrast with the original report, we did not detect a poly-T variant allele with 35 repeats. Despite resequencing work, we only found a common allele with 34 repeats, which we considered our reference (34T) for the purpose of this work. The allele variants were named according to the number of poly-T repeats observed, as following: i) 34T (Reference); ii) 41T (Long); iii) 17T (Short). The consensus sequences from our inserts were then subject to multiple alignments against references sequences curated by the UCSC genome browser, including flanking sequences to match our inserts. Parallelisation and visualisation of the sequences were carried using the publicly available tools MAFFT v.7 (with the FFT-NS-1 preset, speed oriented) and Jalview v.2.11.2.5, respectively. The consensus sequences of the inserts, encompassing the polymorphic poly-T stretch, were aligned and compiled below (Figure 3.7). Interestingly, other than the additional poly-T repeats in the sequence, only the 41T variant allele also contained 3 additional SNPs in its downstream sequence: an A>G transition (+2bp), a T>G transition (+10bp) and a G>C transition (+37bp), seemingly in linkage disequilibrium with this allele.

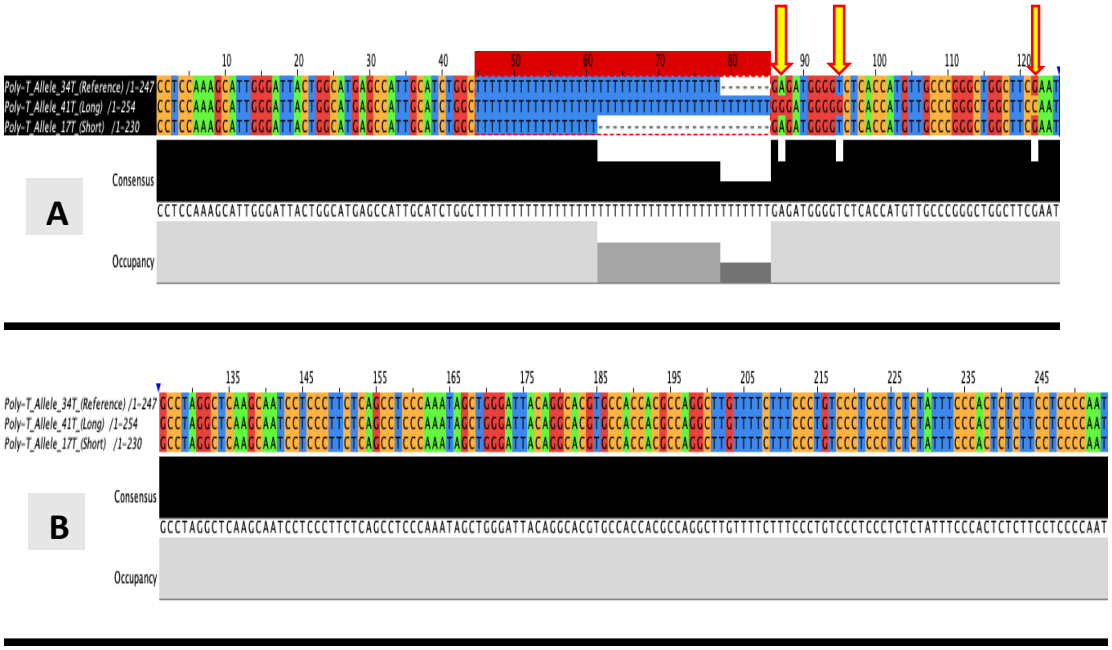


Figure 3.7. Alignment of the consensus sequences compiled from the allele variant inserts of this study, encompassing the poly-T stretch '523'. A) Diagrammatic scheme of positions 1 to 125 of the inserts analysed: The segment denoted by a top red bar (positions 45-85) refers to the polymorphic poly-T stretch and shows the polymorphic sites of the alleles 34T (used as reference sequence), 41T and 17T. Of note, only the 41T variant allele contained 3 additional SNPs in its downstream sequence (indicated by yellow arrows): an A>G transition (+2bp), a T>G transition (+10bp) and a G>C transition (+37bp). B) Diagrammatic scheme of positions 126 to 254 of the inserts analysed shows no apparent polymorphism between the variant alleles.

3.5 Transfection of reporter gene constructs and luciferase assay with the TOMM40 inserts poly-T17, poly-T34 poly-T41T.

Next, we sought to validate the potential regulatory activity of POLY-T17, POLY-T34 and POLY-T41 *in vitro* using the SK-N-AS neuroblastoma cell line as a model. The three selected constructs were cloned into the pGL3P luciferase reporter vector and assayed for regulatory function 48 hours post-transfection. The corresponding plasmids' DNA containing the inserts of interest, were transfected into the human neuroblastoma cell line SK-N-AS cells (ECACC# 94092302), which were maintained as described in section 2.6.1.

Prior to transfection, cells were counted, levels of confluence checked in, and then plated out in 24-well plates at an estimated density of $6x10^4$ cells per well, as

detailed in section 2.6.2. Transfection was performed approximately 24 hours post-plating with 1µg of the appropriate reporter gene construct and 10 ng of internal control construct TK renilla utilizing a standardised kit, TurboFect (ThermoFisher Scientific) following manufacturer's instructions and using serum-free media for the transfection.

3.6 TRANSCRIPTIONAL REGULATORY ACTIVITY OF THE FLAM/POLY-T CONSTRUCTS (POLY-T17, TOMM40/ POLY-T34 AND TOMM40/ POLY-T41

Gene expression was measured through dual luciferase reporter assays using the GloMax luciferase luminometer to measure fluorescence. Expression from the pGL3P vector results in the production of firefly luciferase. Luminescence was assessed using a Glomax 96 Microplate Luminometer (Promega) to obtain the average fold of Firefly Luciferase activity. The data generated was then normalised to internal control the pRL-TK Renillia luciferase vector, described in 2.5.1 and 2.5.2.

3.6.1 Measuring of reporter gene activity by luciferase shows expression directed by '523' reporter gene constructs.

The Dual-Luciferase Reporter Assay (Promega) was performed 48 hours post-transfection cells, as detailed in section 2.5.1. Luminescence was measured employing a Glomax 96 Microplate Luminometer (Promega). We performed duplicate experiments, each time with four replicates per construct. Statistical one-way ANOVA with Bonferroni t-test correction for the three FLAM/poly-T constructs (POLY-T17, POLY-T34 and POLY-T41) and standard t-test for SINE block, in pGL3p and pGL3c were performed to determine the significance of the fold change of the reporter gene constructs over the pGL3p, or pGL3c alone. Significance was scored as follows: *P<0.05, **P<0.01, ***P<0.001, as detailed in Table 3.3.

The changes in the mean values among the POLY-T17, POLY-T34 and POLY-T41 groups were more significant than would be expected by chance. Following adjustment for standard deviation of the duplicates, we still observed statistically significant differences with the inserts, when compared to the baseline expression of the reporter gene alone (p-value = <0.001) - Table 3.3.

A	Source of Variation	DF	SS	MS	F	P
	Between Groups	3	19,305.4	6,435.1	42.0	**<0.001**
	Residual	28	4,285.3	153.1		
	TOTAL	31	23,590.8			

B	Comparison to baseline	Diff of Means	t	P	Significant
	POLY-T17 vs. pGL3p	- 52.19	8.44	<0.001	Yes
	POLY-T34 vs. pGL3p	- 30.07	4.86	<0.001	Yes
	POLY-T41 vs. pGL3p	+ 9.83	1.59	0.369	No

C	Comparison between groups	Diff of Means	t	P	Significant
	POLY-T17 vs. POLY-T34	- 22.12	3.576	**0.008**	**Yes**
	POLY-T17 vs. POLY-T41	- 62.02	10.027	**<0.001**	**Yes**
	POLY-T34 vs. POLY-T41	- 39.90	6.451	**<0.001**	**Yes**

Table.3.3 Statistical analysis of the poly-T constructs (investigated groups) with the pGL3p control group set as the baseline comparator. A) One-way analysis of variance (ANOVA) following normality test (Shapiro-Wilk p=0.78), equal variance test (p = 0.13), with the result showing significant variation attributable to the investigated groups containing the poly-T constructs (p-value <0.001); **B)** Pairwise comparison between investigated groups versus baseline Control Group (pGL3p) with the groups T40/FLAM17 and T40/FLAM34 showing very significant association after Bonferroni t-test correction (p<0.001); **C)** Pairwise Comparison between investigated groups with T40/FLAM17 vs. T40/FLAM34 versus, T40/FLAM17 vs T40/FLAM41 and T40/FLAM34 vs T40/FLAM41 showing a negative trend towards the shorter alleles with pairwise comparison significant associations for all three groups, even after Bonferroni t-test correction (p=0.008, p<0.001, p<0.001; respectively).

3.6.2 Functional activity of the intron 6 SINEs block in reporter gene analysis

The reporter gene constructs generate expression (measured by luciferase activity), which is driven by a heterologous SV40 promoter in the human neuroblastoma cell line SK-N-AS. As shown in Figure 3.8 A, the intron SINE, containing the 34T variant (correspondent to the T40/FLAM34), significantly reduced the reporter gene expression when compared to the minimal promoter (pGL3p) alone ($p<0.001$ and $p<0.01$). Corroborating with the repression was observed in the pGL3p vector. The fragment was also analysed when inserted into the pGL3c vector (which indicates a 7-8 fold increase in intrinsic activity compared to pGL3p). The results for pGL3c were consistent with the results observed for pGL3p and demonstrated that the 17T SINE block (correspondent to the T40/FLAM17 construct) significantly reduced the expression of luciferase when compared with the pGL3c vector alone ($p<0.01$ and $p<0.001$) (Figure 3.8 B).

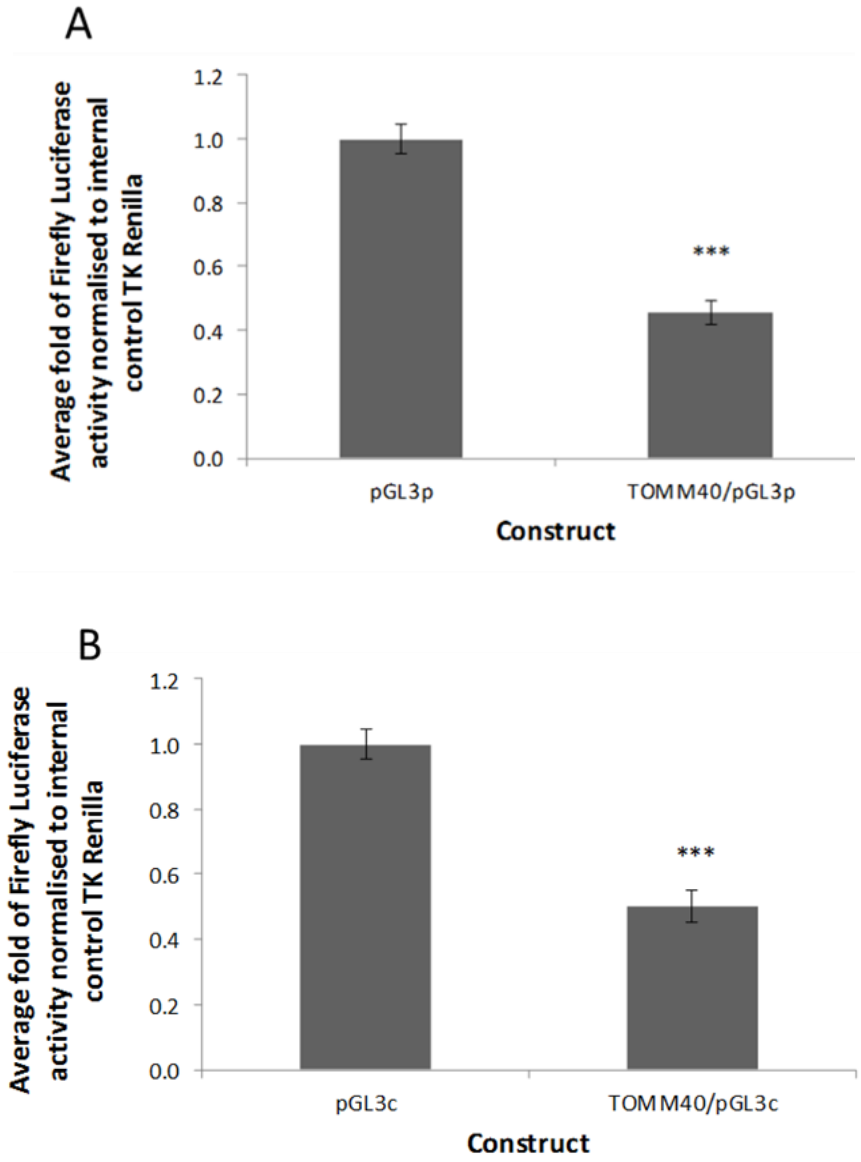


Figure 3.8. Expression analysis using a reporter gene construct containing the SINEs block showing significant downregulation activity in the SK-N-AS human neuroblastoma cell line model. A) The average fold activity of the SINEs block of TOMM40 versus the minimal SV40 promoter vector alone (pGL3p). **B)** The average fold activity of the SINEs block of TOMM40, over the pGL3 control vector (pGL3c), which in addition to the SV40 promoter, also includes an SV40 enhancer. T-tests were performed to measure the significance of the average fold and were scored as follows after Bonferroni correction: *P<0.05, **P<0.01, ***P<0.001.

3.6.3 Functional activity of the FLAM_A with the rs10524523 variants in reporter gene analysis

To address if a functional difference was supported by responses to the high and low activity poly-T allele variants, we compared FLAM_A17T and 34T, which only differ in the number of Ts in each construct (Figure 3.9). The FLAM_A17T showed the most significant ability to alter the reporter gene activity and significantly repressed the levels of reporter gene expression when compared to the minimal promoter alone ($p<0.001$). In contrast, FLAM_A 41T did not affect transcription. A significant reduction in luciferase activity was also observed for Poly-T 34T (p-value, <0.01 and <0.05), however, the repression was not as great as that observed for FLAM_A 17T. Our results would be consistent with a model in which the shorter the '523' variant, i.e. the lower the number of thymine residues within the FLAM_A region, the greater is the ability to modulate transcription.

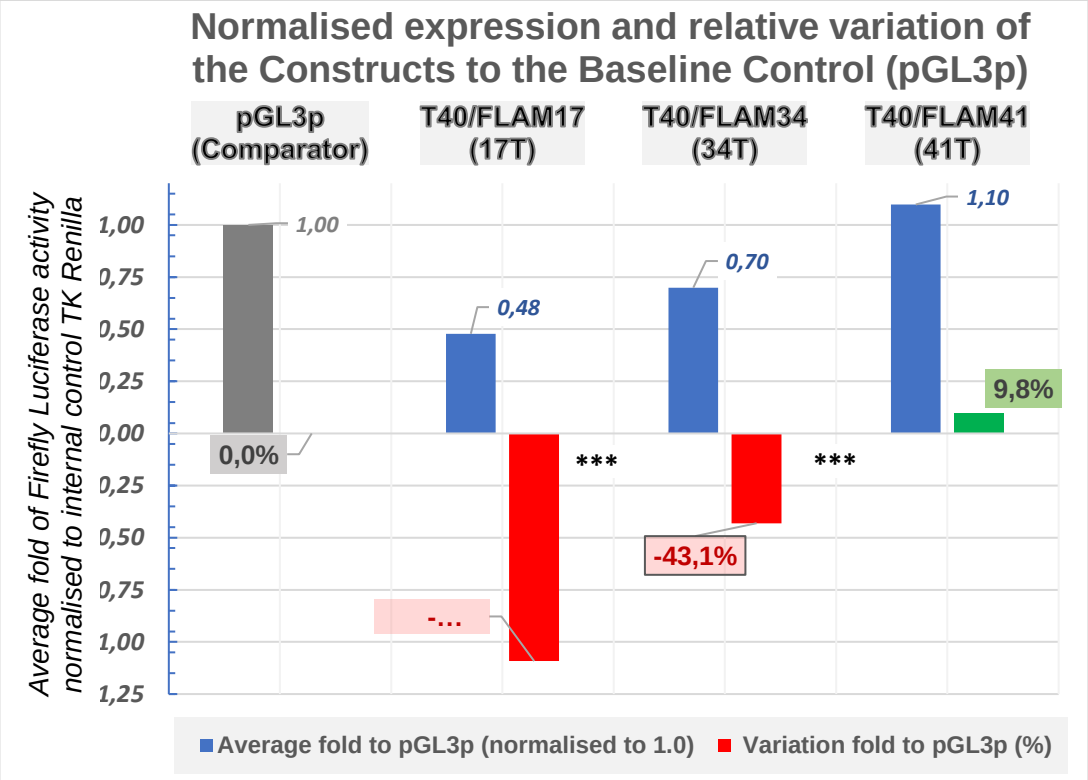


Figure 3.9. *In vitro* gene expression of T40/FLAM constructs compared to pGL3p in an SK-N-AS neuroblastoma cell line model. One-tailed t-tests were conducted to determine significant fold-change differences between the three reporter gene constructs against the minimal pGL3 promoter/pGL3 control alone. Significant alterations were found

with both T40/FLAM17 and T40/FLAM34 ($p<0.001$***), but not with T40/FLAM41 ($p>0.05$). Average firefly luciferase activity for each construct was normalised to the internal control TK Renilla as a standard procedure using a minimum of 4 replicates (n=4). Statistically significant associations are depicted with asterisks, using the following convention: *$p<0.05$, **$p<0.01$, ***$p<0.001$.

3.7 Associations with a longitudinal cognitive cohort study

3.7.1 Study cohort

Samples and data used for the genotypes and analyses were obtained from the Dyne Steel Cohort (DSC), University of Manchester Longitudinal Studies of Cognition in Normal Healthy Old Age, as described in section 2.1.4.2. The data was obtained in collaboration with Dr Antony Payton from the Centre for Integrated Genome Medical Research (CIGMR), University of Manchester.

3.7.2 Cognitive measurements and longitudinal analysis models

Through this partnership with CIGMR, data were shared from a comprehensive, collaborative analysis already undertaken on cognitive assessment of the cohort, as described by Payton et al., 2015 [130]. Using principle components analyses, common factors were derived from key cognitive domains, i.e. memory, vocabulary ability, fluid intelligence (executive functioning) and processing speed for each individual recruited from either sites, Manchester and Newcastle, as described by Rabbitt et al., 2004 [200]. Both male and female participants were included and examined, with their scores standardised separately.

Longitudinal growth curve models were estimated that took the 0 points on the age scale, as age 70 and measured variation on that in units of 10 years. Cognitive measurements with serial data collection occurred up to four occasions, with intervals of approximately five years between them and spanning over a period of 15 to 20 years.

The fixed part of the models that described the overall pattern of change in the

construct for the sample included linear and quadratic age terms. In addition, to account for possible artefactual improvement due to practice effects (a single-step function), each test score was allowed to increment between the first and subsequent occasions in which each test was taken. Individual differences were accounted for by allowing subject-specific random effects for the intercept, describing the within-sample variation in performance at age 70, and for a linear growth term, describing the individual differences in the trend of cognitive decline over the follow-up period. The random effects were assumed to be bivariate normally distributed, and the models were estimated using established generalized linear and latent mixed models (gllamm) by maximum likelihood using adaptive quadrature (www.gllamm.org). Incomplete data observations were included under the assumption that the missing scores were missing at random, allowing attrition to be selective for age, sex, and observed test scores. Subject-specific intercepts and linear trends were estimated using empirical Bayes' methods. This can be used to fully describes the model where a single repeated cognitive test was available. Different tests allowed different means, scales, and error variances, but the tests were assumed to reflect a single underlying construct with a common trend, also known as a common factor g-score.

3.7.3 Genotyping and analysis of polymorphic poly-T variant '523'

The poly-T variant, rs10524523 ('523'), in the *TOMM40* gene was genotyped by the Roses laboratory using a method described previously by Linnertz et al. and Hayden et al. [110, 214].

Briefly, the genotypes of *TOMM40*' 523' poly-T were determined based on length variation. The '523' region of each genomic DNA sample was PCR-amplified using a fluorescently labelled primer. Genotypes were determined on an ABI 3730 DNA

Fragment Analyzer for allelic size discrimination using the Gene Mapper v4.0 software (Applied Biosystems).

The '523' alleles were assigned according to the length/size of their PCR product. Roses et al. [211] established the convention for determining three groups of based on the number of poly-T repeats displayed (Table 3.4 A). Hence, a combination of up to 6 genotypes is possible according to the proposed tri-allelic system: i) Short (S, with 19 or less poly-T repeats); ii) Long (L, with between 20-29 poly-T repeats); and iii) Very Long (VL, with 30 or more poly-T repeats) (Table 3.4 B**).**

Table 3.4. TOMM40' 523' poly-T length variation and alleles combination. **A)** TOMM40' 523' poly-T polymorphism length variation grouped into in three categories according to the number of poly-T repeats: Short (S); Long (L); and Very Long (VL). **B)** Possible genotype combinations based on the proposed tri-allelic system for the poly-T stretch '523' locus.

A

Allelic Group	poly-T repeats
Short (S),	≤19
Long (L)	20-29
Very Long (VL)	≥30

B

Genotypes possible
S/S
S/L
S/VL
L/L
L/VL
VL/VL

3.7.4 Allelic and genotypic analysis of '523' with memory and age-related decline

The minor allele frequencies of the "523" alleles were as following: S = 0.42; L = 0.13; and VL = 0.45; whilst genotype frequencies were: S/S (261, 16.2%); S/L (213, 13.2%); S/VL (626, 38.8%); L/L (27, 1.7%); L/VL (162, 10.0%) and VL/VL (324, 20.1%). Linear regression results are detailed by Payton et al. [114]. Summarising, the Cross-sectional regression analysis indicated that vocabulary ability (a measure of crystallised intelligence) was the most strongly affected by '523', with

the short variant associated with a lower ability. The significance was even more pronounced when homozygous individuals for the short allele were compared to those homozygous for the very long allele ($p < 0.007$). Conversely, using generalized linear and latent mixed (gllam) models it was observed that the vocabulary ability decline rate was faster for those who were homozygous for the very long allele than those who were homozygous for the short allele, though the significance was marginal ($p = 0.05$) and a non-significant trend was observed for memory decline ($p = 0.1$). No significant associations were observed for fluid intelligence or processing speed. Nevertheless, the direction of effect, where the S variant was correlated with the reduced capacity and slower rate of decline, was consistently detected for all tasks.

The '523' homozygous S and VL variants were also analysed in homozygous APOEε3 participants, but this failed to strengthen significance. We considered this study an effort to reproduce earlier findings and consequently did not required to adjust for multiple testing.

3.7.5 In silico evaluation of the human-specific SVA D downstream to the ApoC1 gene.

As previously described in section 3.3.3, we conducted an in silico analysis on chromosome 19 for the clustered genes TOMM40-APOE-APOC. While identifying 116 SVA elements, a particularly SVA D located at 19q13.32 (i.e. chr19:45424392-45425280) has drawn significant attention, since it was located approximately 1,750 bp downstream to the *APOC1* gene (see Figure 3.10). This SVA D does not include the characteristic CCCTCT-mer leading sequence and is displayed by the USCS browser as a fragmented track (Figura 3.11); which contains labels for the rearranged DNA sequence starting at position Chr19:45424340: SVA D followed by a a tetranucoletide tandem repeat stretch (TTTA)n, again as SVA D containing

the central part of the sequence, followed by an AluSx element and lastly a FLAM C.

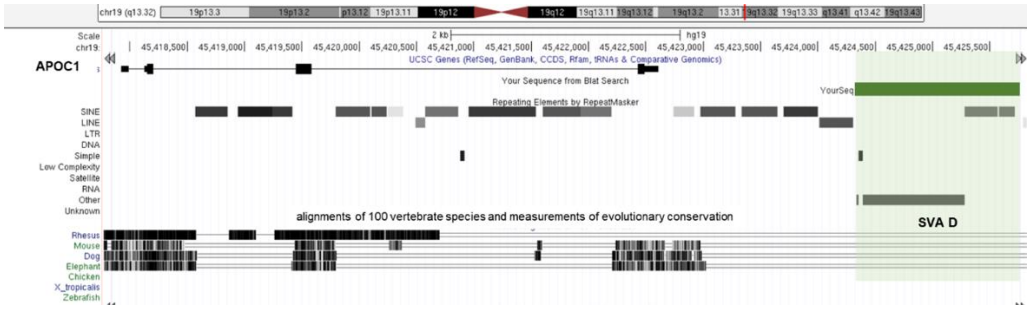


Figure 3.10. A snapshot from the UCSC browser at 19q13.32 (hg19) focusing on APOC1 and the SVA D downstream, which was selected for the study. The image displays the coordinates within chromosome 19, the labels for the repeat maskers to indicate both the LTR and non-LTR elements, the conservation track for comparative genomics. Note that APOC1 gene maps a SINE-dense location and has an SVA D located 1.75 Kb downstream. As shown by the multiple alignments of 100 vertebrate species and measurements of evolutionary conservation, the SVA D is located at a human-specific locus and indeed the vast majority of the intergenic region up to the neighbouring APOC1 pseudogene (APOC1P1)

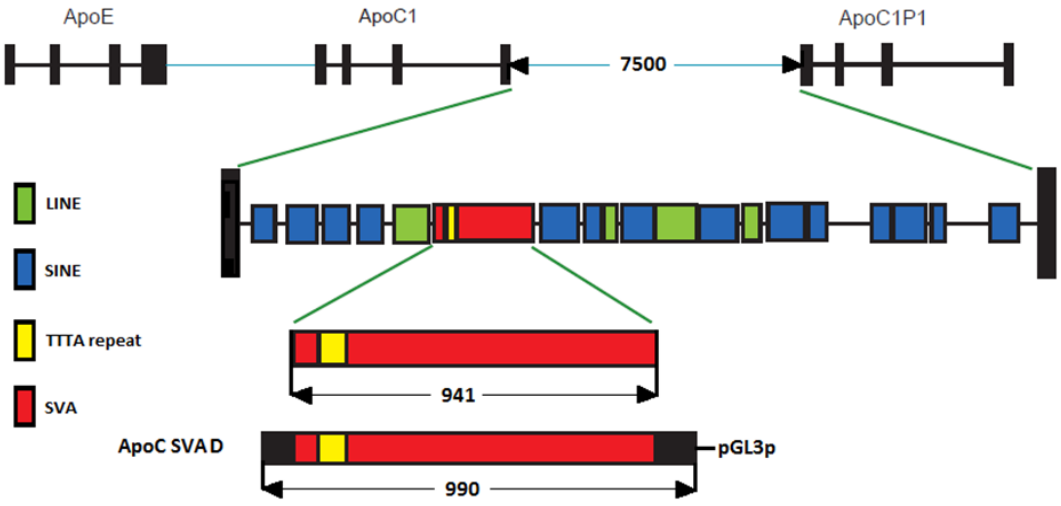


Figure 3.11. Generated construct for ApoC1 SVA D. Diagram showing the genetic structure of the region between ApoC1 and ApoC1P1 genes and the relative relationship of the generated construct. Numbers correspond to the fragment size in bp. Rectangles represent different genetic components: green – LINE, blue-SINE, yellow-(TTTA)n repeat and red-SVA.

Conservation analysis indicates that this particular SVA D has analogous sequences in other primares, but they are localised elsewhere in other chromosomes. For instance, in chimpanzees it maps chromosome 20 displaying 96% of sequence similarity, whereas in gorillas there is 95% of identity, but is located in chromosome 14. Furthermore, the other primates (orangutans, gibbons or rhesus macaque) do not seem to possess SVA D elements, according to in silico analysis with the UCSC browser. ApoC1 SVA D in chromosome 19 is, therefore, human-specific. Our findings agree with existing data suggesting that old-world monkeys, such as rhesus macaque, possess only the precursor of the VNTR domain found within SVA. Therefore, they lack full-length SVAs, and these retrotransposons do seem to be hominid-specific. [215]

Genomic analysis using data from the ENCODE project (available at www. genome.ucsc.edu/ENCODE/) indicates that this particular SVA D is located approximately 5-7 kb downstream to other active regulatory elements. According to the marking of the H3K27 Acetylation histone track, this position is enriched with active enhancers and therefore associated with higher activation of transcripts (as depicted in Figure 3.12). In addition, the area also is also predicted to contain several transcription factors binding sites and DNAse hypersensitive areas, which is also indicative for the presence of regulatory elements, including promoters.

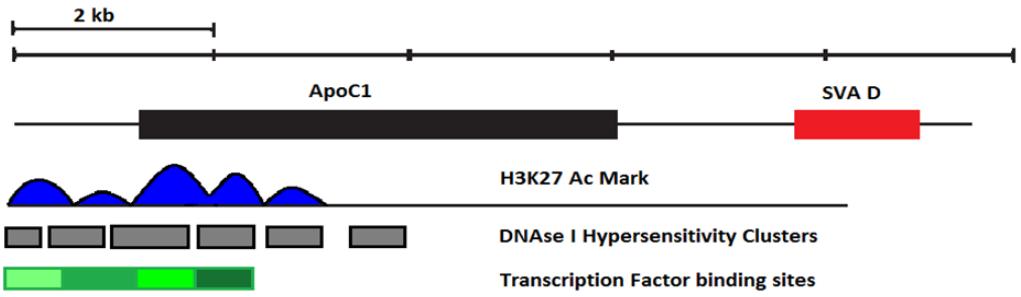


Figure 3.12. ApoC1 - SVA D located approximately 5kb downstream of potentially active regulatory regions. Data from ENCODE project suggest the presence of the H3K27Ac mark, DNAse I Hypersensitivity Clusters and Transcription Factors binding sites in area 5-7 kb upstream of the SVA D.

3.7.6 Functional activity of ApoC1-SVA D.

The reporter gene analysis was conducted to explore the capability of this candidate retrotransposon to account for differential gene expression of APOC1. First, we analysed the *APOC1*-SVA D fragment by performing a luciferase reporter gene assay in SH-SY5Y cell lines. This work included another previously generated SVA D construct obtained from the promoter region of the PARK7 gene, which was made available for this experiment since both SVAs (including the one from this study) were inserted into pGL3 promoter vectors and we planned to use the PARK7 SVA D as a comparator group to replicate its repressive activity found with SK-N-AS ($p<0.01$) in SH-SY5Y cells as well ($p < 0.001$). Results are shown in Figure 3.13. Although the orientation of the PARK7 SVA D in this reporter gene construct was unknown, the SVA D-driven downregulation of luciferase activity observed in SK-N-AS cells was also observed when the SVA D was tested in reverse orientation [8].

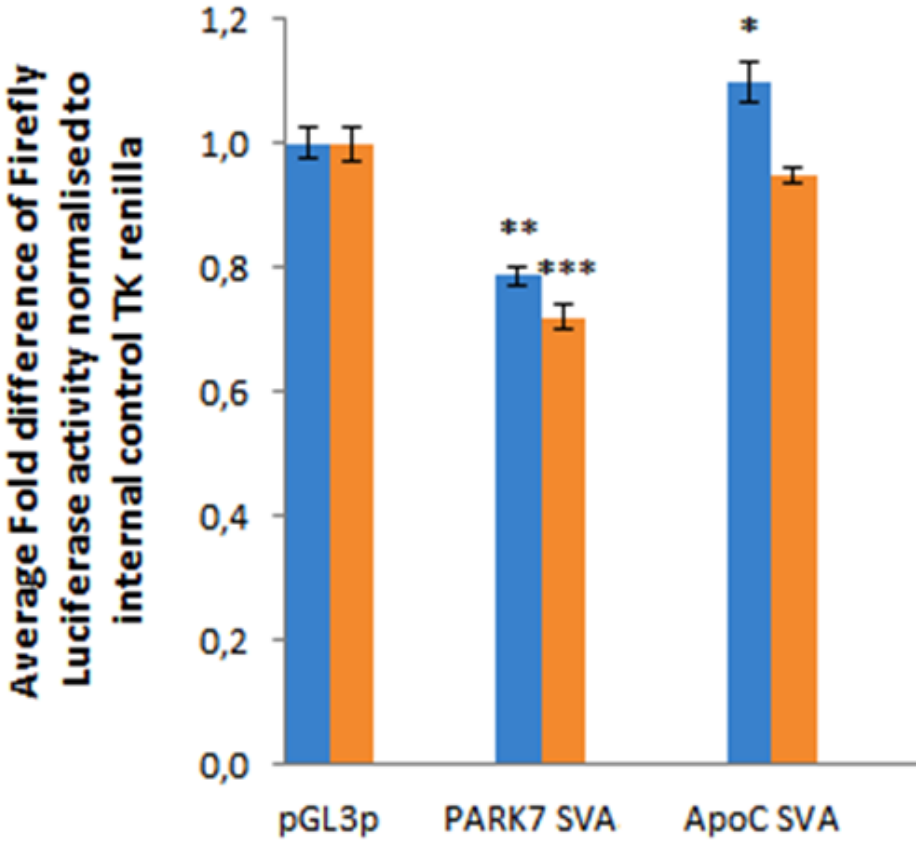


Figure 3.13. The average fold difference of firefly luciferase activity normalised to the internal control showing that the SVA D from ApoC1 resulted in a discrete variation to the pGL3p control alone (Blue bars indicate work done with SK-N-AS cells, whereas the orange bars with SH-SY5Y). The average fold activity of PARK7 SVA D and ApoC1 SVA D compared to the pGL3p alone. Our research group previously examined the results with PARK7 SVA D in SK-N-AS cells; we used this construct as our comparator control and addressed its effect in the alternative SH-SY5Y cell line. PARK7 SVA D could downregulate transcription activity by reducing luciferase reporter gene expression levels when compared to the pGL3 promoter alone (**$p < 0.01$ and ***$p < 0.001$).

In contrast, ApoC1 SVA D did not significantly modulate the luciferase gene activity. One of the duplicate results shows that the retrotransposon is a weak enhancer compared with the minimal promoter ($p<0.05$). The other suggests that the SVA_D does not affect the reporter gene expression. Future work could potentially involve functional assays using different cell lines or model systems,

such as chick embryos, as it has been demonstrated for an SVA_D in the FUS promoter [216].

3.8 APOC SVA D Genotyping with Manchester Samples

To investigate the existence of polymorphisms within the SVA D and subsequently assess their potential allele frequency, we sought to customise, develop and validate a custom genotyping assay for this locus through the testing of human DNA sample sets from the Dyne Steel Elderly Cohort Study, the University of Manchester Longitudinal Studies of Cognition in Normal Healthy Old Age, as previously described in sections 2.1.10 and 3.7.1[200].

3.8.1 Analysis of the methylation status of the SVA D at the APOC1 locus

Due to the complex composition of this SVA D, with repetitive repeat motifs and a high GC content, in addition to an expected size of nearly 1kb (~989 bp), it was very challenging to custom design assays, optimise cycling conditions and consistently amplify this locus. Reproducible PCR amplifications were achieved to analyse potential SVA polymorphisms after a series of protocol adjustments using genomic DNA samples from the Manchester cohort. A hot start DNA polymerase amplification system (GoTaq DNA Mastermix, Promega) and a high concentration of betaine solution were used to successfully amplify with custom-designed assays, as previously outlined in section 2.2 and Table 2.1.

Gel electrophoresis of the PCR products was run using an extended time and adjusted conditions for larger fragments (>1kg) and then visualised with a UV transilluminator (Figure 3.14). However, the fragment sizes made it challenging to define slight variations in length, i.e. size differences inferior to 10% of the expected product sizes. To further attempt discrimination of polymorphic variants, twelve samples with seemingly variable sizes were selected and analysed using capillary electrophoresis (2100 Bioanalyzer, Agilent) for the data acquisition of

electropherograms in order to size and obtain sample concentrations of each expected fragment (Figure 3.15).

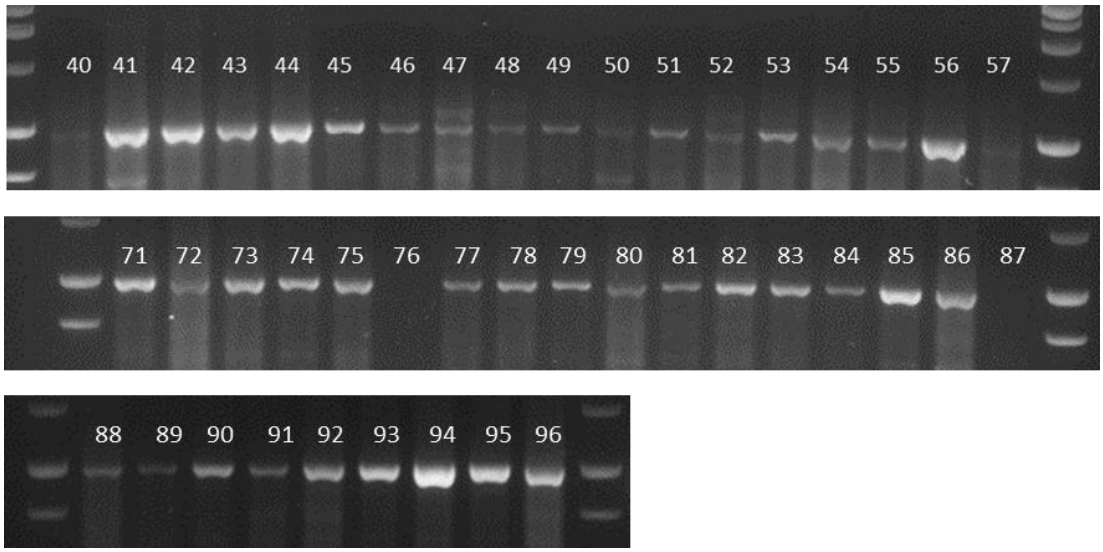


Figure 3.14. Image generated from electrophoresis of SVA D PCR amplification. The PCR product was identified by 40 to 57 and 71 to 96 was run on a 1.5% agarose gel, as detailed in section 2.2, using EtBr and separated by electrophoresis and compared to a 1 Kb plus ladder. A PCR product length of approximately 989 bp was estimated based on in silico assessment of reference genome sequences. Visual inspection suggests the presence of putative size differences between some fragments. Due to the product size, confirming the presence of length polymorphisms in this locus was impossible.

Results with capillary electrophoresis using the Bioanalyzer 2100 showed size variability of ~10% (ranging between 1,020 bp to 1,105 bp), compared to the expected length of 989 bp, which corresponds to a difference between 31 and 116 bp in size measurement.

All in all, both size measurement approaches suggest that the PCR fragments tested were either of similar size/length (in bp), or within a close size range that these methods could not resolve, thereby warranting further investigations. More advanced analytical strategies are required to confirm potential SVA D length polymorphisms in the tested samples.

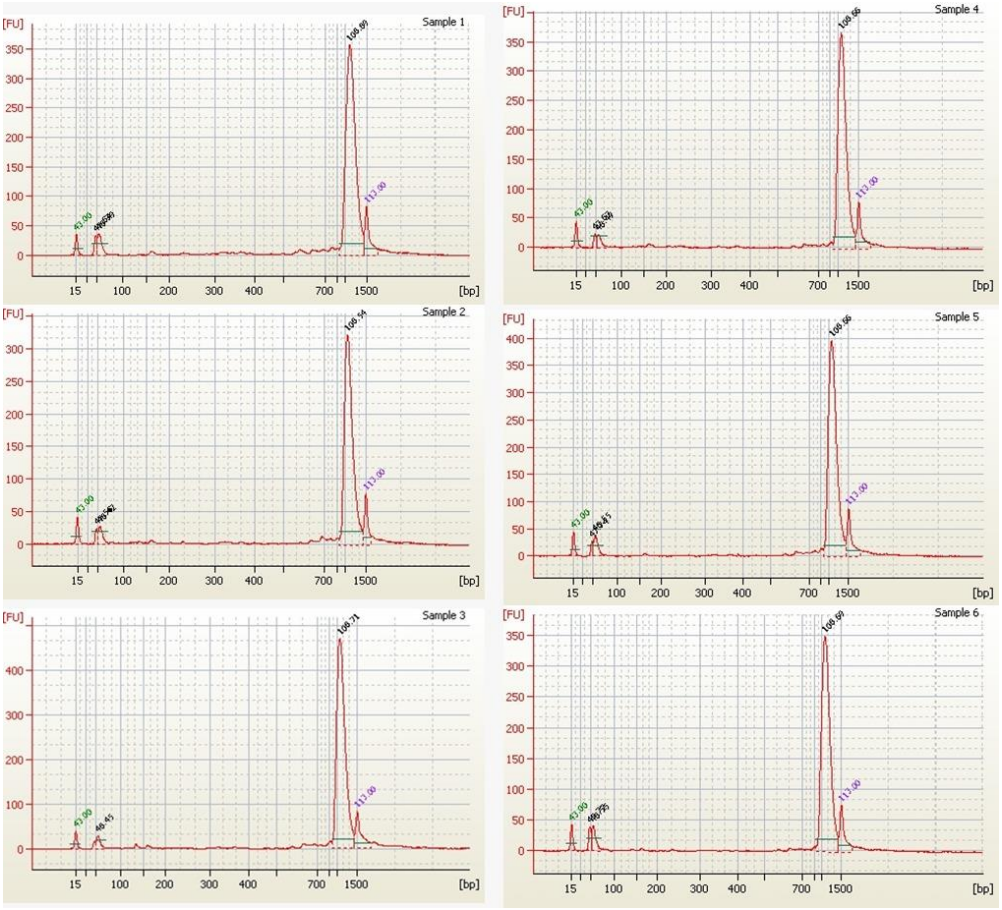


Figure 3.15. Electropherograms generated from a subset of amplified fragments of the SVA D amplification using the 2100 Bioanalyzer capillary electrophoresis system. Runs were analysed using the 'running and evaluating electrophoretic methods" parameters of the Expert software (Agilent). The reference green and purple peaks refer to the low and high size standards, respectively, and were used to measure fragment sizes and concentrations.

The absence of association analysis with the APOC SVA D genotyping was mainly caused by the low numbers of samples able to give a clear s size/length (in bp), due to the large size of the PCR product and the slight variance observed at the samples analysed, as it was difficult to analyse and not sufficiently robust to define the association with this samples numbers further analysis and new approaches would benefit the association analysis.

3.9 discussion chapter 3

3.9.1 The polymorphic poly-T stretch rs523 within TOMM40

We have shown that a primate-specific retrotransposon domain encompassing the genetic variant ('523') within the TOMM40 gene can act as a differential regulator of gene expression based on the genotype of the T repeat and has a significant influence on crystallised intelligence and decline in non-pathological ageing. Specifically, we described that the short poly-T variant was associated with reduced reporter gene expression, decreased vocabulary ability and a slower rate of decline. The association between the TOMM40-APOE locus and cognitive impairment have been together with the most extensively reported findings in the field of complex genetic diseases/traits. The functional relevance of these genes in the aetiology of cognitive impairment adds weight to their involvement. ApoE serves in the transport and breakdown of lipoproteins, influences amyloid-β clearance, and competes with amyloid-β for cellular uptake via ApoE receptors.[217] TOMM40 encodes a translocase of outer mitochondrial membrane 40, a subunit of a larger TOM complex accountable for translocating newly synthesised unfolded proteins into the mitochondrial intermembrane space. [115] Mitochondria are vital for cell survival, and their dysfunction has been linked to pathological and non-pathological cognitive ageing. [218] Until recently, the focus of cognitive/dementia genetic research has significantly relied upon SNP data with little emphasis on other variant types. Previous association studies have demonstrated that the function of non-coding DNA associated with SNPs and advocated for their ability to exert regulatory properties at the locus [67, 207]. The properties of the multiple Alu elements are particularly interesting at this locus, which constitutes a human-specific domain with limited homology to other primates. Retrotransposon domains in non-coding genome regions can establish new splice sites, polyadenylation signals, promoters and regulatory domains that can reorganise gene expression and build new transcription modules [219]. The

location of retrotransposons within introns, promoter, untranslated regions and vicinities further enhances their ability to alter the local architecture and modulate genome regulation [220]. We have previously demonstrated that primate-specific SVA retrotransposons in the promoter of the neurodegenerative candidate genes, PARK7 and FUS, comparably to our present study, are both polymorphic and capable of modulating reporter gene expression [131, 221]. Data from our group [220] indicate that the SINEs block or the FLAM_A domain are plausible candidates to mechanistically modulate vocabulary ability and decline by accounting for differential TOMM40 gene expression, perhaps in response to an additional environmental challenge. However, these regions could also be hotspots for retrotransposition events, affecting the locus' function by multiple mechanisms. With the exception of early embryo development and some malignancies, retrotransposition is suppressed by epigenetic and post-transcriptional modifications in somatic cells, although their mobilisation has also been described in the adult brain. Baillie and colleagues have detected somatic L1 retrotransposition of three active retrotransposon families: LINE-1, Alu and SVA, which resulted in an alteration of the genetic landscape of the human brain [222]. Consistent with CNS mobilisation, sequencing-based studies have shown the existence of significant bias for insertion sites of somatic L1 retrotransposition in neurons of schizophrenics and that these events could have an aetiological role [223]. More recently, dysregulated retrotransposon function has been observed in frontal FTLD, which the authors linked with the activity of the neurodegenerative risk gene TDP-43 [224]. Furthermore, many studies, including our own, have indicated ageing as an important parameter affecting human memory and cognition. This is consistent with the observations in the Drosophila mode, where increased activity of transposable elements in the brain was found to contribute to age-dependent loss of neuronal function [224].

The TOMM40 ‘523’ alleles were previously divided into three groups depending on the length of the T homopolymer: T≤19 (S), Ts=20-29 (L) and T≥30 (VL). The original study by Roses and colleagues reported that longer poly-T length (in particular, T≥27) was associated with earlier age of AD onset compared to individuals with shorter T-homopolymers in people of European descent [103]. Those studies were conducted using three independent cohorts with numbers of cases ranging from 34 to 83 and numbers of controls from 31 to 67. In contrast, a more extensive study of 1,594 AD cases and 1,190 controls (all European descent) found no association between the VL variant in homozygous APOEε3 individuals and age of AD onset but did report that the VL variant was associated with lower AD risk [106]. This finding was supported by an independent group that investigated the ‘523’ variant using 414 AD cases, 173 centenarians and 305 neurologically healthy controls [107]. This study also reported that the L variant was associated with an increased risk of AD and a reduced likelihood of living to 100 years of age. Finally, an even larger study of 11,840 AD patients and 10,931 controls found no association between ‘523’ variants and AD risk [105]. The association between the ‘523’ variant and cognition has been slightly more consistent than the AD reports [211], although the cohort sizes have been smaller. Johnson et al. found that those homozygous for the VL variant (n=35) scored lower on a test of primary retrieval from a verbal list learning task than S homozygous individuals (n=38) (mean age of the cohort, 55 years) [111]. Two further studies by Hayden et al. [110] and Greenbaum et al. [108] (n=127, mean age 80.6; and n=331, ≥65 years, respectively) observed that those homozygous for the S variant performed better on tests of memory and executive function, while in these studies, the results either did not reach significance or were only marginally significant. Finally, a more extensive investigation of 639 cognitively normal individuals aged 21-97 years found that those homozygous for the VL variant had a flattened test-retest improvement that was only observed in those aged under 60 years [109].

Our results showed a faster rate of cognitive decline in volunteers who carried the VL variant, which was marginally significant for vocabulary ability and exhibited a non-significant trend for memory. In contrast to other cognitive studies using neurologically healthy participants, the S variant in our volunteers was associated with reduced vocabulary ability ($p < 0.007$) but not memory ($p = 0.11$), executive function ($p = 0.14$) or processing speed ($p = 0.24$). Our observed differences in neurologically healthy individuals may be related to population stratification. The Greenbaum study investigated TOMM40 in an Israeli Jewish population, and the Johnson study did not mention their population ethnicity. However, the Hayden study was 99% Caucasian, but their results failed to reach significance after correction for multiple testing. Statistical power is unlikely to be an issue in our analysis as our cohort size of 1,613 was larger than the other studies on cognitively healthy individuals combined. When analysing the subgroup of homozygous SS versus homozygous VL individuals, our sample size of 583 volunteers had 93% power to detect a genetic effect size of 2%, assuming a minor allele frequency of 0.45 (frequency of the S allele) and an additive genotypic model of effect.

The cognitive tests used by the different groups also varied, and none used a representative crystallised intelligence test comparable to the one used in this study. Finally, in our study, we used principal components analysis to derive the four factors that represent the four chief domains of cognition and, theoretically, should have been able to detect an association with any other cognitive test [225].

The strong linkage disequilibrium between TOMM40 and APOE genes [226] and their coordinative expression, as well as potential shared regulatory domains, suggests the existence of complex genome interactions with concerted regulatory mechanisms that ultimately determine the level of expression of these clustered genes. Retrotransposons have the ability to modulate gene expression via various processes. The occurrence of a primate-specific mobile DNA element within the

TOMM40-APOE, which may well extend towards neighbouring gene-dense loci, supports its regulatory nature at the same time emphasising the involvement of this locus with higher intellectual traits and, consequently, cognitive dysfunction. Conversely, it is noteworthy that poly-T '523' polymorphism has been challenging to genotype and a simplified categorisation into three groups, has potentially hindered an adequate assessment of an otherwise multi-allelic system. This is further compounded by the fact that the simple analysis of consensus sequences we generated for the locus, clearly indicate the existence of other polymorphisms (SNPs), which were specific for only one of the poly-T variant allele (T40/FLAM41, or simply 41T) and therefore likely to be inherited together.

Nevertheless, this work demonstrates the ability of distinct allele variants within an intronic SINEs block to influence expression within a luciferase reporter gene construct, which in turn were independently associated with cognitive measurements in a non-pathological community-dwelling elderly cohort. Despite these findings, further mechanistic studies are warranted since the precise biological basis for these claims, as well as clarification for the inconsistencies found across studies, have yet to be established.

3.9.2 The SVA D downstream to APOC1

The clustered genes TOMM40-APOE-APOC1 locus is located in a gene-rich region with significant genetic variability. The localisation of the retrotransposons within introns and the local environment is likely to contribute to their ability to alter the local architecture and, consequently, modulate genome regulation. SVA D, located downstream from the APOC1 gene, has been proposed to form secondary structures such as DNA quadruplexes [216], whilst neighbouring inversely arranged Alus (in the opposite strand) can induce the formation of stem-loops and palindromes. These constructed structures are able to influence RNA interference, transcription inititation, DNA replication, and introduce alternative splicings [115].

Moreover, SVAs are GC rich, which can introduce potential methylation sites and lead to the formation of abnormally methylated regions and alterations in the epigenetic profile of the locus [8]. Retrotransposons can also exert an influence on the genome via the action of retrotransposition. Although retrotransposition is suppressed by epigenetic and post-transcriptional modifications in somatic cells, except for early embryo development and some malignancies, the mobilisation has been reported in the brain. Baille et al., have recently detected somatic L1 retrotransposition of three active retrotransposon families: LINE-1, SINE/Alu and SVA, into protein-coding genes in the brain, which resulted in modifications of the genetic landscape of the human brain [218]. SINE-VNTR-Alus (SVAs) are the youngest and one of the minor groups of retrotransposons though they are specific to hominids and can display significant polymorphism, with nearly 2,700 of these elements represented in the human reference genome (hg19) [68]. The SVAs subfamilies are divided from SVA A to F1 with the SVAs subfamilies A, B, and C found across multiple primate species. In contrast, a complete SVA D segment, as well as subfamilies E, F and F1 are considered humans-specific, with many of them likely to carry polymorphic repeat motifs in it. These elements are, therefore, of renewed interest for genetic and functional genomic studies. Certainly, the most recent of the SINEs and quite possibly one the most distinctive evolutionary signatures that make us humans in relation to other primate species.

Human-specific retrotransposon insertions are supposed to be one of the two fundamental driving forces in the evolution of human-specific regulatory arrangements [122], followed by the recombination-based exaptation of highly conserved fragments of ancestral regulatory DNA. This evolution was complex, involving gene regulatory mechanisms in higher primates and humans and quite possibly partly impacted species- and tissue-specific gene regulation and allowed greater diversity comprising epigenetic modulation and response to environmental changes [114]. Bennett et al. found that on average a given individual can have

concealed an estimated 56 polymorphic SVAs within their genome [6], which would be likely to contribute significantly to an individual's genetic variation. Bennett et al. also showed that 79% of human SVAs are missing from their corresponding genomic sites in the chimpanzee genome, indicating that SVAs have played a recent evolutionary task after the split between humans and chimpanzee species [36, 61, 131]. Many transposable element families are established to be distributed non-randomly throughout the human genome. This was emphasised in the early draft of the human genome, in which portions of the X-chromosome were observed to contain exceptionally high concentrations of the LINE-1 retrotransposon sequences, similarly to chromosome 19, for which an increased concentration of Alu elements has been observed [227]. The latter was predicted to be due to the chromosome's high gene and high GC content. It has been shown that SVAs are in fact not distributed at random across the genome, instead their distribution displays a preference for position with high GC content [68][3], which tend to harbour genic regions, with 60% of SVA elements either enclosed by genes or sitting within 10 kb upstream [216]. Wang et al. and Tang et al. emphasize that chromosome 19 has a genome unusually rich in SVAs, compared to other chromosomes [68, 200], with Grim wood et al. explaining that chromosome 19 has a higher proportion of transposable element-derived sequences compared to the remaining genome, 55% versus 44.8% [214]. SVA elements are over-represented in the established Parkinson's disease gene loci (PARK gene cluster) [110], and have been frequent at gene loci that regulate central nervous system (CNS) pathways [211]. Previous studies have verified the capability of SVA elements to control gene expression in human neuroblastoma cell lines (*in vitro*) and chick embryo models (*in vivo*) [103, 210, 213, 217], which would reinforce that SVAs have the potential to regulate gene expression, either through their regulatory characteristics, or by their location within the genome.

Indeed, Kim and Hahn have underlined the ability of SVA insertions to act out as novel promoters at their location of integration, therefore potentially driving the expression of novel human-specific transcripts [61, 68]. Similar observations were given by Kwon et al. who hypothethised that the addition of SVA sequences occurred at numerous human transcripts [228]. Recent work by Tang et al. showed that human-specific retrotransposon insertions had incorporated an addition 14.2 Mbp to the human genome, which by extrapolation corresponded to an increase of 84 kb of expressed human-specific transcripts [200].

Furthermore, it has been observed that at least 12.5% of the human-specific SVAs overlapped with ENCODE ChIP-seq data, in practical terms this would account for the creation of an additional 504 TFBS [200].

A full SVA is encompassed of five main components, beginning with: (i) a simple hexamer repeat of (CCCTCT)n at the 5'end, followed by (ii) an Alu-like region made up of 2 antisenses Alu fragments separated by a region of the intervening sequence, (iii) a variable number tandem repeat (VNTR),(iv) a SINE region resulting from the 3' long terminal repeat (LTR) of the retroviral HERV-K10 element, and lastly (v) a 3' poly(A) signal (Figure 3.4) [68]. A typical full-length SVA element is approximately 0.7–4 kb in length. Nevertheless, due to the recent evolutionary period and also the repetitive pattern in the structure of the SVA, several SVA elements are polymorphic, either within their structure, or in relation to their presence/absence in the human population.

Polymorphisms can arise from several components of its structure; but it is often found, either in its hexamer repeat (CCCTCT)n, within its VNTR core sequence, or in its poly(A) region [213]. SVA insertions are associated as a reported cause of at least 13 diseases [70], for instance, the X-linked dystonia-parkinsonism (XDP) caused by insertion, wheresoever an SVA-F element, located in the intron 32 of the TATA-box binding protein associated factor 1 (TAF1) gene. [225, 229-231].

Intriguingly, this SVA insertion was not only observed to be associated with reduced expression of TAF1, but also with alternative splicing and intron retention, leading to defective gene. The length of its CT hexamer repeat was inversely correlated with the age of onset of the disease [232]. Likewise, polymorphic SVA insertions have been demonstrated to regulate gene expression or isoform expression via intron retention in a population-specific manner [229, 233] This emphasizes their competence to have an impact at several levels of genetic processing and contribute to phenotypic differences within a variety of diseases, including additional neurodegenerative conditions, such as Alzheimer's, Parkinson's and Amyotrophic Lateral Sclerosis.

3.10 CONCLUSION

Retrotransposons have the ability to modulate gene expression via various mechanisms. The presence of these primate-specific, mobile DNA elements within the TOMM40-ApoE-ApoC1 gene locus and their potential regulatory nature supports the association of this gene-dense cluster with higher cognitive traits and, subsequently, cognitive dysfunction.

We demonstrated the ability of different constructs to interfere transcription within a luciferase reporter gene construct, and future work could potentially investigate their ability to modulate reporter gene expression directed by an extended TOMM40/ApoE-specific construct containing regulatory elements, as well as a more detailed investigation of the SVA D, whether or not coupled with the nearby *APOC1*. Indeed haplotype studies indicated the existence of a strong linkage disequilibrium in this intergenic region, which extend towards both, the 5' direction to the TOMM40-APOE gene locus, and the 3' direction to a cluster of Apoliprotein genes, which emphasises the need for additional studies.

Chapter 4

Gene regulation is driven by a variable number of tandem repeats (VNTR)

The work in the following section has been published in the Journal of

Section 4.3 (Part I): Genes, Brain and Behavior: doi.org/10.1111/gbb.12483.

Section 4.4 (Part II): Molecular Neuroscience: MN, doi:10.1007/s12031-018-1044-z.

The work for this chapter, including tissue culture, cell treatments, RNA extraction, cDNA conversion, PCR and data analysis, is all in collaboration and shared with Dr Maurizio Manca.

4.1 INTRODUCTION

Variable number tandem repeats (VNTRs) are a genetic polymorphism defined by the copy number of repeating elements within it. The number of repeat units (R) indicates the number of copies which defines the VNTR genotyping. We focus on VNTRs located in non-coding DNA, which could be regulatory. [7, 234] Monoamine oxidase A (*MAOA*) gene is often considered an excellent example of how gene x environment interaction (GxE) can shape gene expression and phenotype. Specifically concerning central nervous system (CNS) disorders, as it is a key regulator of monoamine neurotransmitters in the brain, including serotonin (SE), dopamine (DA), and norepinephrine (NE) [133, 159, 168]. MAOA has been referred to as a warrior gene, as it is associated with aggressive behaviour [28, 30, 167, 235],

Due to its genomic location at the X chromosome, implying gender could play a role in regulation. [28, 30, 158].

MAOA gene has two variable numbers of tandem repeats near the promoter region of the gene. These polymorphic repetitive DNA domains were termed as the upstream variable number of tandem repeats (uVNTR); it is found 1kb upstream of the transcription start site (TSS) MAOA isoform most cited and with abundant mRNA. The uVNTR comprises a 30-bp motif that can be repeated 2, 3, 3.5, 4, and 5 times. [169, 236] The repeats 2R, 3R, and 5R are commonly defined as low expression variants and, as considered, high expression variants repeats 3.5R and 4R, which have demonstrated an increase of 2- to 10-fold in reporter gene expression. The uVNTR is often recognised as a risk factor for many mental disorders associated with behaviour. [136, 137, 167]

The second VNTR found approximately 500 bp upstream of the promoter region, has been termed dVNTR (distal variable number of tandem repeats). Philibert, et al., has previously described this VNTR, which comprehends a 10 bp repetitive

sequence, arranged by two different decamers, decamer A: CCCCTCCCCG and decamer B: CTCCTCCCCG, and these repeats are arranged in genotypes of 8R to 12R repeats. The alleles 9R and 10R are the most common. [159]

MAOA gene has at least two transcriptional start sites (TSSs), with approximately 1.3 Kb separating both, with subsequent two putative coding isoforms and distinct 5' untranslated regions (5’ UTRs). As the isoforms differ in length, the presence of alternative exons and specific start codons hypothetically can result in the translation of two protein isoforms. In this work, we considered the functional consequences of the coding transcripts of both isoforms and the interaction with the two VNTRs, given that these aspects had not been previously considered in the literature.[20]

The work of this chapter considered the contribution of both VNTRs to the regulation of the MAOA gene. First, we accessed the MAOA regulation to correlate allele-specific gene expression with allele-specific transcription factor binding and epigenetic marks for MAOA, taking advantage of heterozygous DNA for the uVNTR, precisely that of the human neuroblastoma SH-SY5Y cell line (ATCC CRL-2266).

We then addressed the presence or absence of both VNTRs in semi-haploid HAP1 cell lines using genetically engineered to knockout (KO) using the editing tool CRISPR-Cas, of either the uVNTR, dVNTR or both VNTRs. Complementing our functional analysis, we determined the haplotype variation of these VNTRs in distinct cohorts to address the genetic association of both VNTRs in neurodevelopmental disorders.

4.2 AIMS

4.2.1 HYPOTHESIS

Both, uVNTR and dVNTR, have the potential to drive and regulate the expression of MAOA gene, and their respective genetic variation domains could have a role in mental health.

4.2.2. AIMS

- To address if the MAOA variable number of tandem repeats (VNTR) regulates gene expression by
- Relate allele-specific gene expression with allele-specific transcription factor binding and epigenetic marks for the MAOA uVNTR using DNA from the heterozygous for neuroblastoma SH-SY5Y cell line (ATCC CRL-2266).
- Compare endogenous MAOA expression with genetically engineered CRISPR-Cas, deletion of either the uVNTR, dVNTR, or both VNTRs.
- Correlate the MAOA VNTR polymorphism in the Wirral Child Health and Development (WCHADS) cohort to address the contribution of both VNTRs.

4.3 results of allelic-specific regulation of uVNTR and dVNTR MAOA expression.

4.3.1 Variable Number of Tandem Repeats near promote the region of MAOA gene.

MAOA gene is in the short arm (p11.3) of chromosome X, as defined in the UCSC genome browser human genome version 38 (hg 38) (https://genome.ucsc.edu/) and Ensembl (http://www.ensembl.org/).

Here we are evaluating the well-characterised *MAOA* variable number of tandem repeat (uVNTR) polymorphism found in the promoter region of MAOA, and the distal variable number of tandem repeat (dVNTR), located 529 bp upstream of the uVNTR, location defined in table 4.1, has been structured of two decamers. The repeat termed A: CCCCGCCCCT and B: CCCCGCTCCT [159]. The decamer variation in the number of copies repeated, where was found from 8 to 12 copies. The most common copy of repeats is the 9R and 10R, and has been observed transcriptional differences between both repeats; the 9R was increased when compared to 10R, with other genotypes between these 2 copy repeats, utilising reporter gene assays, as described by Philibert et al., 2011. [20]

Table 4.1: The tandem repeats sequences of the MAOA VNTRs.

VNTRs	Start	End	Sequence (bp)	(bp)
uVNTR	43655101	43655205	ACCGGCACCGGCACCAGTACCCGCACCAGT	30
dVNTR	43654464	43654572	CCCCGCCCCT or CCCCGCTCCT	10

The Variable Number of Tandem Repeat (VNTR) loci are chromosomal regions in which a short DNA sequence motif is repeated a variable number of times. For instance, the repeats (R) of dVNTR, containing the 10 bp motif can be tandem repeated from 8 to 12 times.

The dVNTR is a tandem repeat of two different motifs, decamer A: CCCCGCCCCT or B: CCCCGCTCCT, originating five possible alleles from 8 to 12 repetitions, (Figure 4.1).

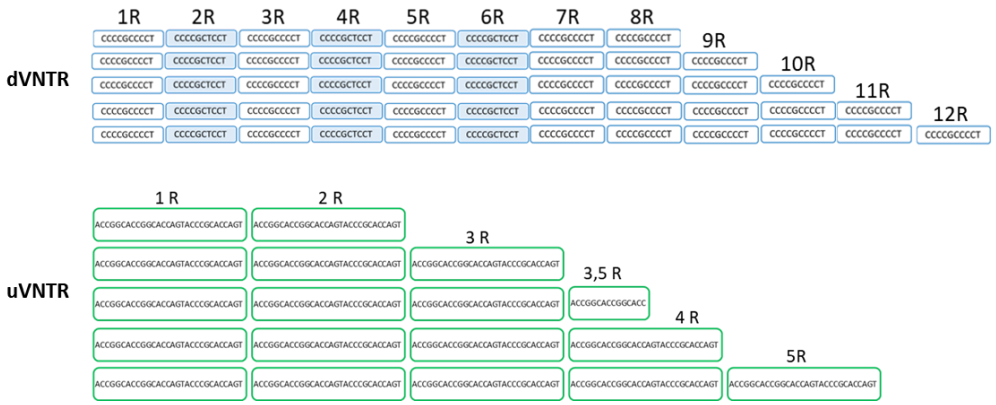


Figure 4.1. Sequences of dVNTR and uVNTR of *MAOA* gene. A) Decamer A and B of dVNTR, the reference sequence formed by 10 bp motif. B) The reference sequence of uVNTR with 3.5R and the motif of 30bp.

4.3.2 uVNTR

A polymorphic repetitive domain termed Upstream Variable Number Tandem Repeat (uVNTR). Located in the promoter of the primary transcript of the MAOA gene, it is recognised as a mediator of MAOA expression. [169]

The uVNTR is a 30 bp repeat element, and the alleles 2R, 3R and 5R are considered as low expression variants (MAOA-L), and alleles 3.5R and 4R are regarded as high expression variants uVNTR (MAOA-H). [154]

Depending on the disorder studied, these low and high-expression alleles can represent either a risk or a protective factor. [28, 30, 135, 152, 154, 165, 166, 237]

It is assumed that the uVNTR acts as a classical transcriptional regulator in the promoter of the MAOA gene and that the distinct variants support differential MAOA expression in a tissue-specific and stimulus-inducible manner.

The uVNTR polymorphism is commonly cited as a case of gene and environment interaction (G × E), meaning that one genotype can have environmental modulation, resulting in consequences to mental health. [165] In other words, the

function of the uVNTR is changed by the balance of active transcription factors in the cell, modified in response to the environment, that are then able to identify and bind to the uVNTR, consequently permitting the domain to play a role of a transcriptional regulator.

4.3.3 Allelic-specific regulation of MAOA expression

To address the allelic-specific regulation expression of *MAOA* uVNTR, was used the human neuroblastoma cell line SH-SY5Y (ATCC/CRL-2266). Cells were cultured as described in section 2.2.8.1. Cells were cultured to 70% to 80% confluence, and 2 μM sodium valproate (Sigma) was added directly to the existing media for 1 hour before harvesting the cells. SH-SY5Y is a well-characterised cell line initially isolated from a 4-year-old female. Genotyping of this cell line showed that it was heterozygous for the uVNTR, with alleles containing either 3R or 4R of the repeat element. This allowed the identification of allelic-specific expression of the MAOA isoform, which initiated from the most 5' TSS because mRNA starting here included the uVNTR in the 5'UTR. The SH-SY5Y cells were submitted to total RNA extraction utilising Trizol reagent (Invitrogen), and the concentration of RNA obtained was measured using a NanoDrop 8000 Spectrophotometer (Thermo Fisher Scientific). Subsequently, the RNA was reverse transcribed to single-stranded cDNA using the GoScript Reverse Transcription System (Promega), utilising Three micrograms of total RNA and random primers according to the manufacturer's protocol. The cDNA was used as a template for amplifying the *MAOA* uVNTR, following the protocol in section 2.2.13.3. Under basal growth conditions, we could detect MAOA mRNA related to expression from both alleles, using the uVNTR length as the distinguishing feature, Figure 4.2 A.

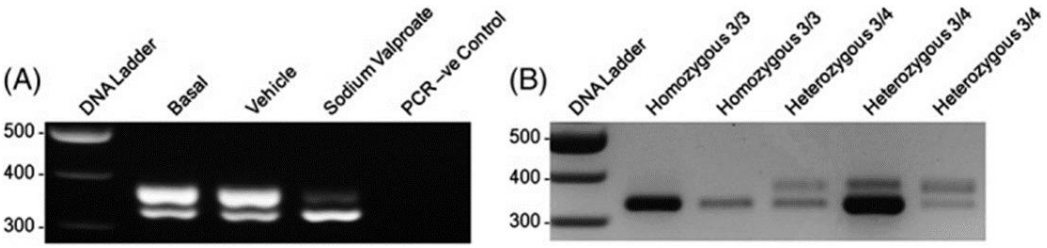


Figure 4.2. Expression of 2 MAOA alleles. Image from electrophoresis in 1.5% agarose gel of MAOA pcr am amplification. Amplicon size 324 bp for 3R uVNTR, 354 bp for 4R uVNTR. A) SH-SY5Y cell line cDNA: lanes from left to right: 100 bp ladder marker; basal condition; vehicle control: H2O 1 hour; sodium valproate treatment 1 hour; PCR negative control. B) Human female cDNA samples: lanes from left to right: 100 bp ladder marker; 2x homozygous 3R/3R uVNTR; 3x heterozygous 3R/4R uVNTR.

4.3.4 Chromatin Immunoprecipitation.

The Human SH-SY5Y cells were also used for chromatin immunoprecipitation. First, they were cultured in a T175 flask to around 80% confluence, followed by 1 hour of treatment under the following conditions, basal anti-Sp1 and anti-CCCTC Binding Factor gene (CTCF) (Millipore). PCR amplification of MAOA uVNTR was performed on the immunoprecipitated samples. The immunoprecipitations were performed using the following antibodies: anti-Histone H3 (Abcam, Cambridge, UK), anti-Nucleolin (Abcam), anti-hnRNP K (Abcam), anti-CCHC-Type Zinc Finger Nucleic Acid Binding Protein gene (CNBP) (Abcam), anti-RNA pol II CTD phosphoSer5 (Active Motif), The supernatant removed was stored for subsequent analysis, as described in section 2.2.16. The beads were washed, and the methylated DNA was eluted. Methylated and unmethylated DNA were also analysed by PCR of the MAOA uVNTR as previously described. In the methylation analysis, the genomic DNA of SH-SY5Y was isolated using QIAmp DNA Mini Kit (Qiagen). The genomic DNA was sonicated to obtain DNA fragments of around 500 bp. The sonication utilised a Bioruptor Plus, where ten cycles of 30-second pulses on high power and then 30-second pauses were performed as detailed in

2.xxx. 300 ng of genomic DNA fragmented was used to perform the methylation analysis using MethylQuest (Millipore), following the manufacturer's protocol.

4.3.5 Expression of 2 MAOA alleles in cDNA obtained from saliva samples.

Likewise, analysis was performed using five heterozygous females' saliva samples from the WCHADS. Previously genotype data allowed us to select only females heterozygous to 3R and 4R. The RNA was extracted using the Trizol protocol and reversed transcribed to cDNA, as described in 2.2.13. The cDNA was used for the expression of MAOA. The results demonstrated that in vivo expression from both alleles could be observed in the same individual, shown in figure 4.2 B. Attempts to use qPCR to measure allele-specific expression over heterozygous VNTRs were problematic, perhaps due to preferential amplification of a specific allele in early rounds of PCR and consistent, reproducible data was difficult to obtain.

However, using the heterozygous cell line SH-SY5Y, we could utilise the size variation of the distinct uVNTRs to compare intensity between and within experiments for the same primer set; this showed that in SH-SY5Y cells in response to exposure to sodium valproate, a mood stabiliser, a switch in the levels of MAOA allelic expression was observed.

Precisely, the ratio of L- to H-allele expression was reproducibly reversed, Figure 4.2 A. This switch in relative levels of expression cannot be explained by PCR allele preference. As the MAOA gene is on the X chromosome, the expression of both alleles in the SH-SY5Y cells might require this region to escape X inactivation, as previously reported for the MAOA locus. [163, 164, 238]

Alternatively, the expression of both alleles could be construed as a mosaic effect of random silencing of one allele in any cell. Thus, the population would appear to express both alleles.

4.3.6 Methylation at the MAOA promoter.

To explore the allelic expression further, we analysed ChIP transcriptional and epigenetic variation at the uVNTR domain in SHSY5Y cells in response to sodium valproate. The methylation status of both alleles was addressed by their ability to bind the Methyl-CpG-binding domain (MBD). By addressing the ratio of each uVNTR allele, this showed an apparent specificity in that the 4R uVNTR promoter was predominantly methylated under basal growth conditions. In contrast, the 3R uVNTR promoter was hypomethylated, Figure 4.3. The methylation pattern was altered when the cells were exposed to sodium valproate.

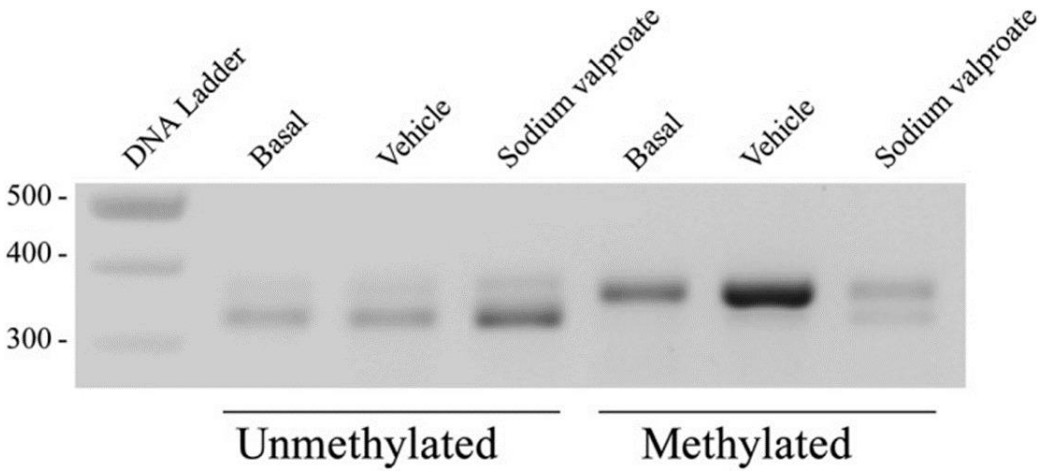


Figure 4.3. Methylation at the MAOA promoter. PCR amplification of sheared gDNA post CpG MethylQuest treatment. Amplicon size 324 bp for 3R uVNTR, 354 bp for 4R uVNTR. Lane 1, 100 bp ladder marker; lanes 2 to 4 PCR of uVNTR under basal condition; vehicle control: H2O; sodium valproate PCR of unmethylated DNA (not bound to methyl domain-binding protein; CpG MethylQuest protein). Lanes 5 to 7, PCR of uVNTR under basal condition; vehicle control: H2O; sodium valproate treatment of methylated DNA (bound to methyl domain-binding protein; CpG MethylQuest protein).

Specifically, variation was most clearly observed in the methylated fractions, in which methylation could now be easily observed over the 3R allele; the ratio of methylated 4R allele to 3R allele was quite distinct from basal conditions (Figure 4.3). Therefore, we explored RNA polymerase II binding and the transcription factors CTCF and SP1 identified from ENCODE data as binding to the uVNTR, to gain insight into potential transcriptional mechanisms operating at this locus.

4.3.7 ChIP analysis of the MAOA VNTRs alleles.

RNA polymerase II was shown to be present on both alleles under basal conditions, consistent with expression from both alleles, Figure 4.4 A, lane 4. This was in stark contrast to Histone 3 (His3) under basal conditions, Figure 4.4 B, lane 4 (His3), which was found predominantly on the shorter, 3R allele, uVNTR allele. Similar to the observations obtained for His3, under basal growth conditions, or exposure to H_20 as the vehicle control, CTCF and SP1 transcription factor binding were predominantly observed on the short, 3R uVNTR allele, Figure 4.4 C, top and middle panel, lanes 7 and 8. Upon exposure to sodium valproate, Figure 4.4 C, bottom panel, lanes 8, there was a significant increase in SP1 binding to the short 3R allele, which is correlated with the switch to increased expression of the longer 5' TSS containing the 3R uVNTR in the 5' UTR observed in Figure 4.2 A.

The distinct promoter architecture suggests a particular/specific mechanism to regulate each MAOA-allelic promoter based in part on the uVNTR repeat copy number. In the literature, CTCF and SP1 bind double-stranded DNA but recognise identical sequences as the single-stranded DNA-binding proteins hnRNPK and CNBP. [239, 240] These single-stranded DNA-binding proteins were also bound predominantly to the short 3R allele under basal conditions, (Figure 4.4 C top and centre panel, lanes 4 and 5), but in response to valproate, they now also bound the 4R allele, (Figure 4.4 C, bottom panel, lanes 4 and 5). This was in stark contrast to SP1 binding, which dramatically increased only on the 3R allele. In contrast, the binding of CTCF was not significantly changed in response to valproate with any change in distribution, and it remained bound predominantly to the 3 R allele.

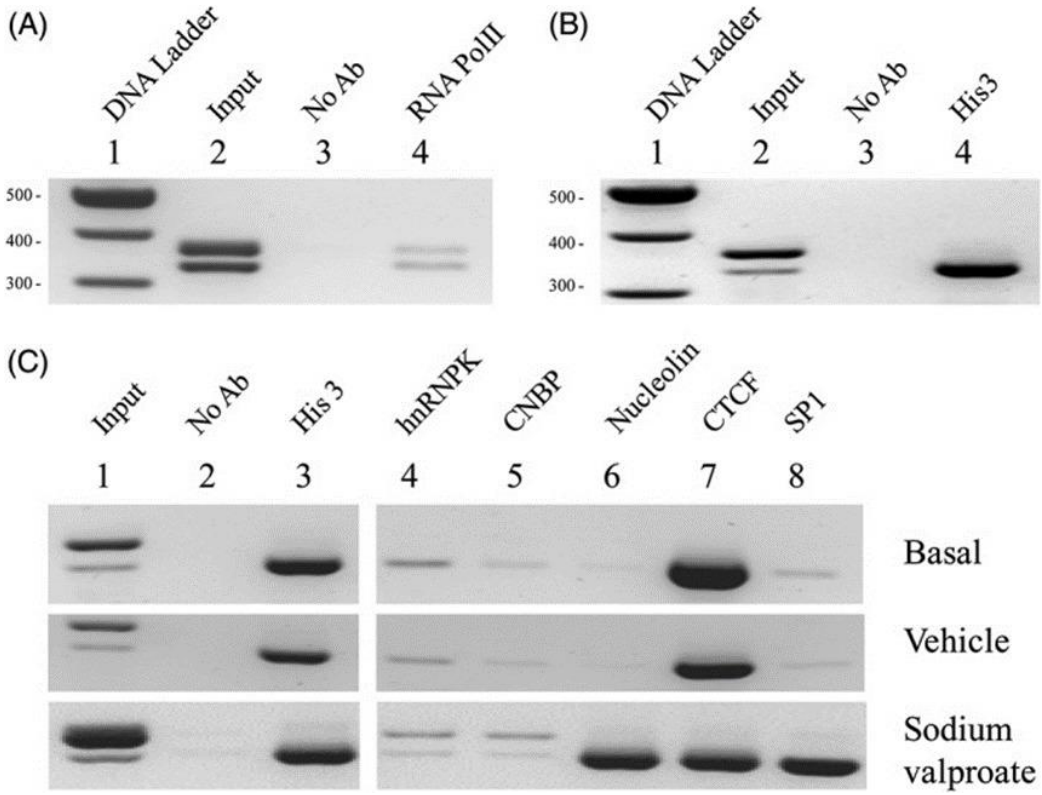


Figure 4.4. ChIP analysis of the MAOA VNTRs alleles. PCR amplification of fragmented gDNA post-enrichment by a specific antibody. Amplicon size 324 bp for 3R VNTR, 354 bp for 4R VNTR. **A)** in vivo interaction of active RNA pol II (CTD phospho Ser5) under basal growth conditions, lanes from left to right: **1.** 100 bp ladder marker; **2.** input DNA (1% sheared chromatin) acts as a positive control for PCR; **3.** no antibody: negative control for non-specific background binding; **4.** active RNA pol II (CTD phospho Ser5) antibody. **B)** in vivo interaction of histone 3 with the MAOA uVNTR alleles under basal growth conditions. Lanes from left to right: **1.** 100 bp ladder marker; **2**. input DNA (1% sheared chromatin); **3.** no antibody; **4.** histone 3 antibody. **C)** ChIP analysis of the in vivo interaction of factors associated with MAOA uVNTR alleles. Image showing results after 1 hour exposure to the following conditions: Basal, vehicle/H2O and 2 uM sodium valproate. Lanes from left to right: **1.** input DNA (1% sheared chromatin); **2.** no antibody; **3.** histone 3 antibody; **4.** single stranded DNA (ssDNA) binding protein hnRNP K antibody; **5.** ssDNA binding protein CNBP antibody; **6.** G4 binding protein nucleolin antibody; **7.** transcription factor CTCF antibody; and **8.** transcription factor Sp1 antibody.

CTCF function on transcription varies, extending from repression to transcriptional pausing, and transactivation, which in part is dependent on interacting partners, one of which is nucleolin which can mediate intra- and interchromosomal interactions. [241]

Analysis of nucleolin by ChIP indicated it binding weakly to the 3R allele under basal conditions with a vast increase in binding after exposure of the cells to valproate, Figure 4.4 C, bottom panel, lane 6. It should be noted that nucleolin can also interact with SP1, and both these factors were dramatically increased in their binding to the short allele after exposure to valproate.

4.4. Results of the role of uVNTR and dVNTR IN the regulation of maoa gene expression.

4.4.1 Bioinformatics of MAOA gene isoforms.

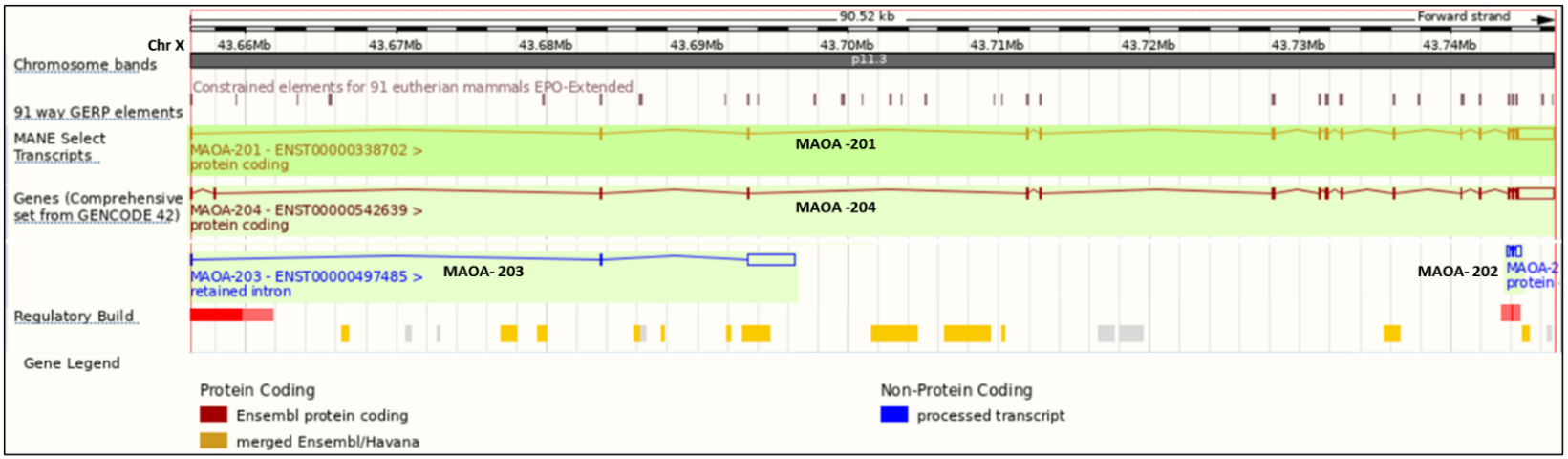
The analysis resulted in two isoforms variants that are transcribed into mRNA (Figure 4.5), The isoform MAOA-204 (ENST00000542639.6), which has 16 exons, genomic, located at hg38 genome position chrX:43,654,907 - 43,746,817 (91,911 bp). Also, the MAOA-201 Transcript (ENST00000338702.4) hg38 chrX:43,656,299 - 43,746,817 (90,519 bp) and has 15 exons, which is the variant 1 or major isoform. Also consulted other public genomes online sources to define the MAOA gene four isoforms, UniProt (http://www.uniprot.org/); Genotype-Tissue Expression (GTEx) (https://www.gtexportal.org/home/) and AceView (http://www.ncbi.nlm.nih.gov/ieb/research/acembly/) identified as isoforms 201 to 204. The data obtained by Ensembl (table 4.2) shows the isoform name and transcript ID, the genomic locations, exons count, transcript length and protein length. The two protein coding isoform variants are displayed on the top, MAOA-201, the primary transcript and 204, the second coding transcript.

Table 4.2: Isoforms of MAOA gene.

Isoform	ID	Genomic Location	Exon count	Transcript (bp)	Protein (aa)	Biotype
201	ENST00000338702.4	43,656,299-43,746,817	15	3931	527	Protein coding
204	ENST00000542639.6	43,654,907-43,746,817	16	5431	394	Protein coding
202	ENST00000490604.1	43,743,704-43,744,658	3	557	no protein	retained intron
203	ENST00000497485.2	43,656,210-43,696,410	3	3420	no protein	retained intron

The genome browsers were interrogated to confirm the mRNA isoforms, where it was found that the MAOA coding transcript isoforms 201 and 204, each predicted distinct protein variants (Figure 4.6).

A



B

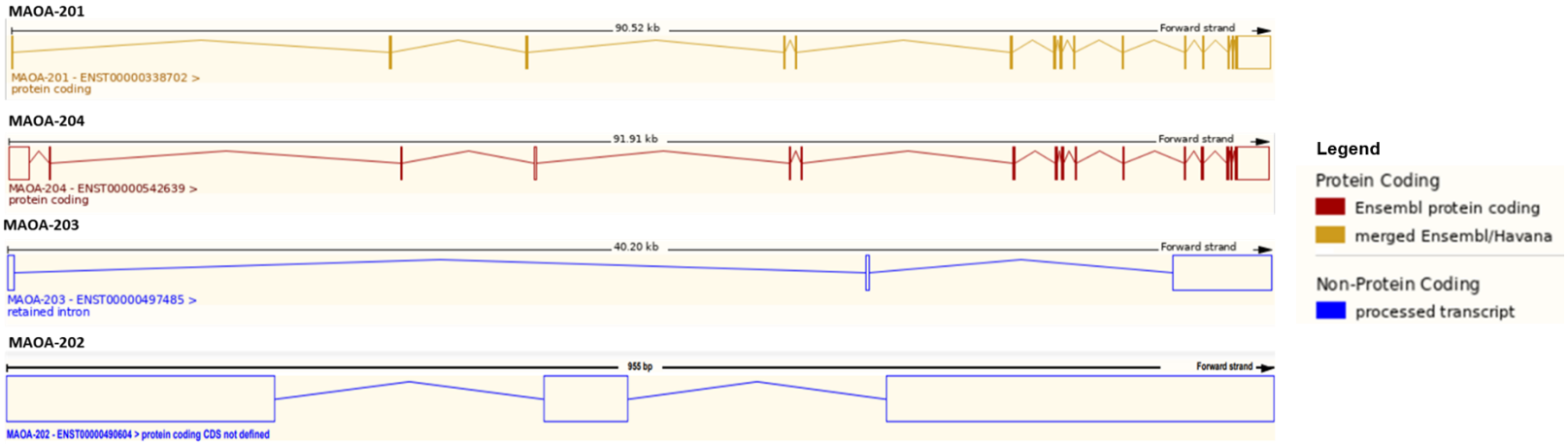


Figure 4.5. Schematics of MAOA gene isoforms on Ensembl (release 108). A) MAOA isoforms 201 major coding isoform, 204 alternative coding isoform, (protein coding), and the non-coding isoforms 202 and 203. B) Representation of each isoform, showing the length in kilo–base-pair (kb), the exons. The isoform 204 has an extra exon, termed exon II, due to the location related to the first exon.

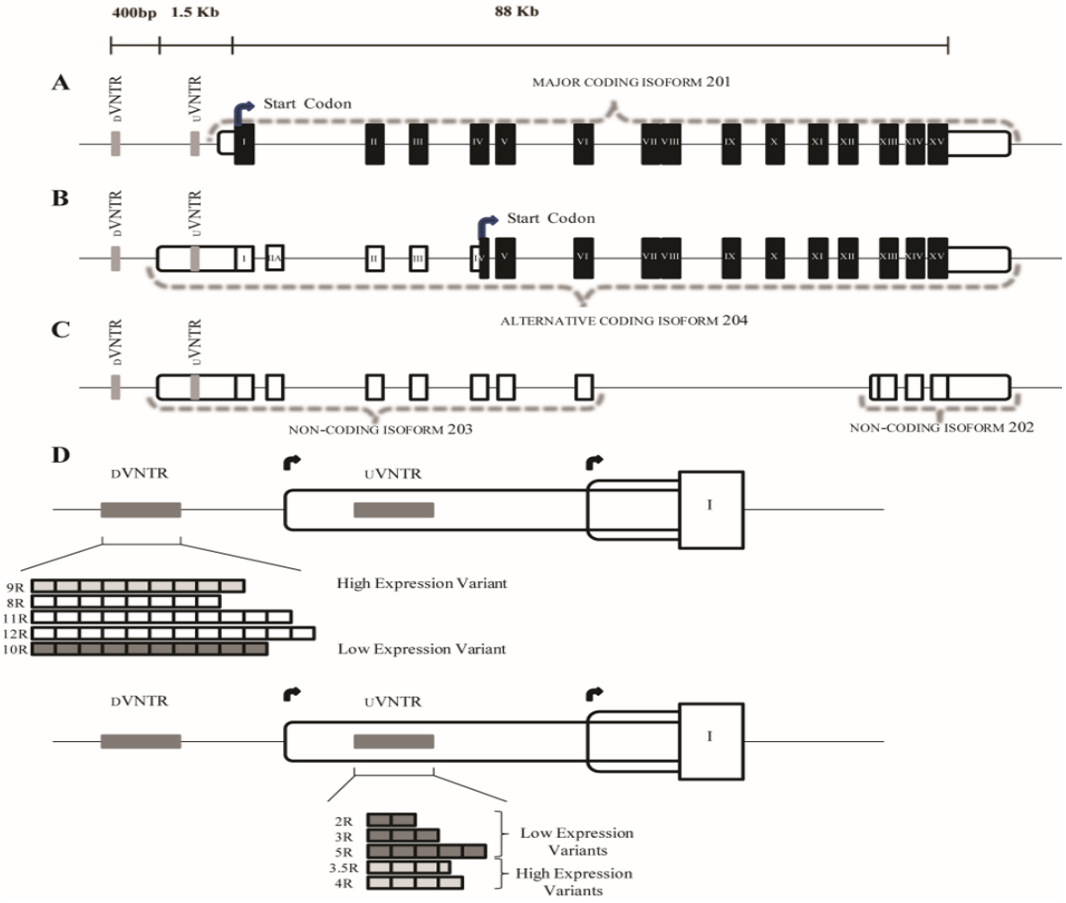


Figure 4.6. Schematics of the isoforms of the monoamine oxidase A (MAOA) gene as described in the UCSC Genome Browser (hg 38).

The following sequences are represented by: 5' and 3' UTRs: by colourless rounded rectangles; Translated exons: by black blocks; Translation start black arrows represent sites, and dVNTR and uVNTR are represented by grey rectangles and located, from left to right, respectively. A) Major MAOA isoform; Ensembl Version 201(NCBI Accession Number: BC008064/Version BC008064.2). B) MAOA secondary isoform; Ensembl Version 204 (NCBI Accession Number: AK293926/Version AK293926.1). C) The untranslated processed transcripts; Ensembl version 203 and 202, respectively. D) Genotyped MAOA dVNTR alleles: 9R represents high expression, 10R low expression, with alleles from 8R to 12R, and each square represents a repeat unit (Philibert et al., 2011). Genotyped MAOA uVNTR alleles: 2R to 5R. The low expression variants have 2R, 3R and 5R, and the high expression variants are 3.5R and 4R, as described for the first time by Sabol et al., 1998.

The established primary isoform of the *MAOA* gene, MAOA-201 (Figure 4.6 A), consists of a 3,931 bp transcript and encodes a full-length protein of 527 amino acids (aa). Here, a VNTR is located 1,094 bp upstream of the transcript start site (TSS) [169]. On this transcript, the 5'UTR extends 124 bp upstream from the first ATG codon.

The MAOA-204, defined as an alternative isoform (Figure 4.6 B)**,** has a 5' UTR 1,392 bp larger than the MAOA-201. Also, this isoform includes the uVNTR, described in section 4.3.3. The MAOA-204 has an alternative exon, exon IIA, located between exon 1 and 3.

The exon IIA includes a premature (TGA) stop codon caused by a shift in the reading frame, resulting in a shorter protein. The next in-frame methionine codon (ATG) situated at exon IV, consequently, may be affected as a result of the shifting in the TSS, which could be the reason this MAOA protein has a truncated amino-terminal, 133 aa shorter than isoform 201, resulting in a protein with 394 aa, as illustrated in Figure 4.6 B.

In this secondary isoform (Figure 4.6 C), the absence of an N-terminal section of the FAD/NAD binding domain was noted, largely overlapping this non-coding isoform 203. However, this MAOA FAD/NAD binding domain is present in the primary isoform. [242, 243]

4.4.2 Horizon Genomics generation of HAP1 VNTR KO cells

The HAP1 cell line is a human near-haploid line, which has been derived from the chronic myelogenous leukaemia (CML) cell line KBM-7. KBM-7 cells were obtained from a male who lacked a Y chromosome. [173, 244] why is this significant.

Many cell lines are diploid. In that way, they can generally have two copies of any allele, except the sex chromosomes, as males have one copy of each sexual chromosome, X and Y and females have two copies of chromosome X. When addressing gene expression of genes located on the X chromosome, it is

necessary to take into consideration that females can be heterozygous, whereas males are hemizygous. One advantage of HAP1 cells in having only one copy of each gene is that we can ensure that the data has no gene dosage interference or are be complicated by additional alleles. The HAP1 cell lines were obtained from Horizon Genomics (https://www.horizondiscovery.com/gene-editing/crispr), and the clones containing specific VNTR deletions were generated using the CRISPR/Cas9 to have clones with the deletion of the markers of interest. The HAP1 clones cell lines generated were confirmed by PCR and Sanger sequencing to prove the presence or absence of the desired mutation at the genomic level. Further details are available at (https://horizondiscovery.com/en/resources/videos/how-do-hap1-knockout-cell-lines-work). (Figure 4.7)

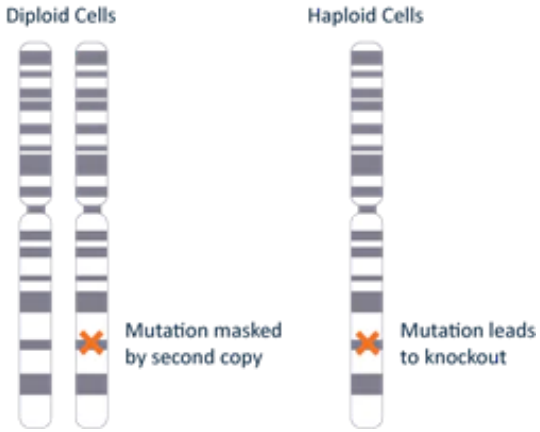


Figure 4.7. Knockout model of HAP1 cells. The difference between haploid and diploid cells and how it can facilitate the complete knockout of a mutation of interest and avoid the occurrence of the results being masked by a second copy, as diploid can have a second mutation at the same loci. (Figure from Horizon - https://horizondiscovery.com/en/engineered-cell-lines/products/human-hap1-knockout-cell-lines?nodeid=entrezgene-9001#description).

4.4.2.1 The HAP1 Clones

To address the function of the uVNTR and dVNTRs in the expression of *MAOA* gene was employed CRISPR/Cas9 and generated clones with the deletion of the

single and both VNTRs of *MAOA* gene and resulting in HAP1 cell line clones with selected distinct VNTR haplotypes.

The HAP1 parental cell line (**P**) was used to generate by CRISPR/Cas9 deletion of the uVNTR and dVNTR (Figure 4.8) in duplicate, resulting in 4 different single knockout cell lines. The two with the uVNTR deletion were clones 9_F4 and 9_E2, identified as **A** and **B,** respectively, and still possess the dVNTR. Quite the opposite, clones 13_B5 and 13_B1, labelled **C** and **D,** in that order, still have the uVNTR and the knockout of the dVNTR.

Two clones generated by each single KO were selected to create a double knockout (DKO) further, applying a second, de-novo, CRISPR/Cas9 deletion to clones **B** and **C**. Clone C generated the DKO du_B5 (E), du_F3 (F) and cloned the other two DKO termed as ud_D8 (G) and ud_F3 (H).

Summarising the Parental lines with a KO of one VNTR originated four clones, A and B with uVNTR KO, and C and D with dVNTR KO. Clone B submitted a second knockout generating two DKO clones, **E** and **F**, as clone C originated the DKO clones **G** and **H**, Figure 4.8.

The deletions have been confirmed by sequencing by Horizon. All clones were also amplified using the PCR with the primer sets for MAOA uVNTR and dVNTR, in that order, Figure 4.9 A-B.

PCR products were run in agarose gel and the virtual gel generated by the QiaXcel, to confirm the deletion for all the cell lines for both uVNTR and dVNTR single KO, also to the four DKO cell lines.

The genotype of MAOA VNTRs observed in the HAP1 cell line was 3R to the uVNTR and 10R to the dVNTR. The results were compared to the SH-SY5Y cell lines to confirm the alleles, Figure 4.9.

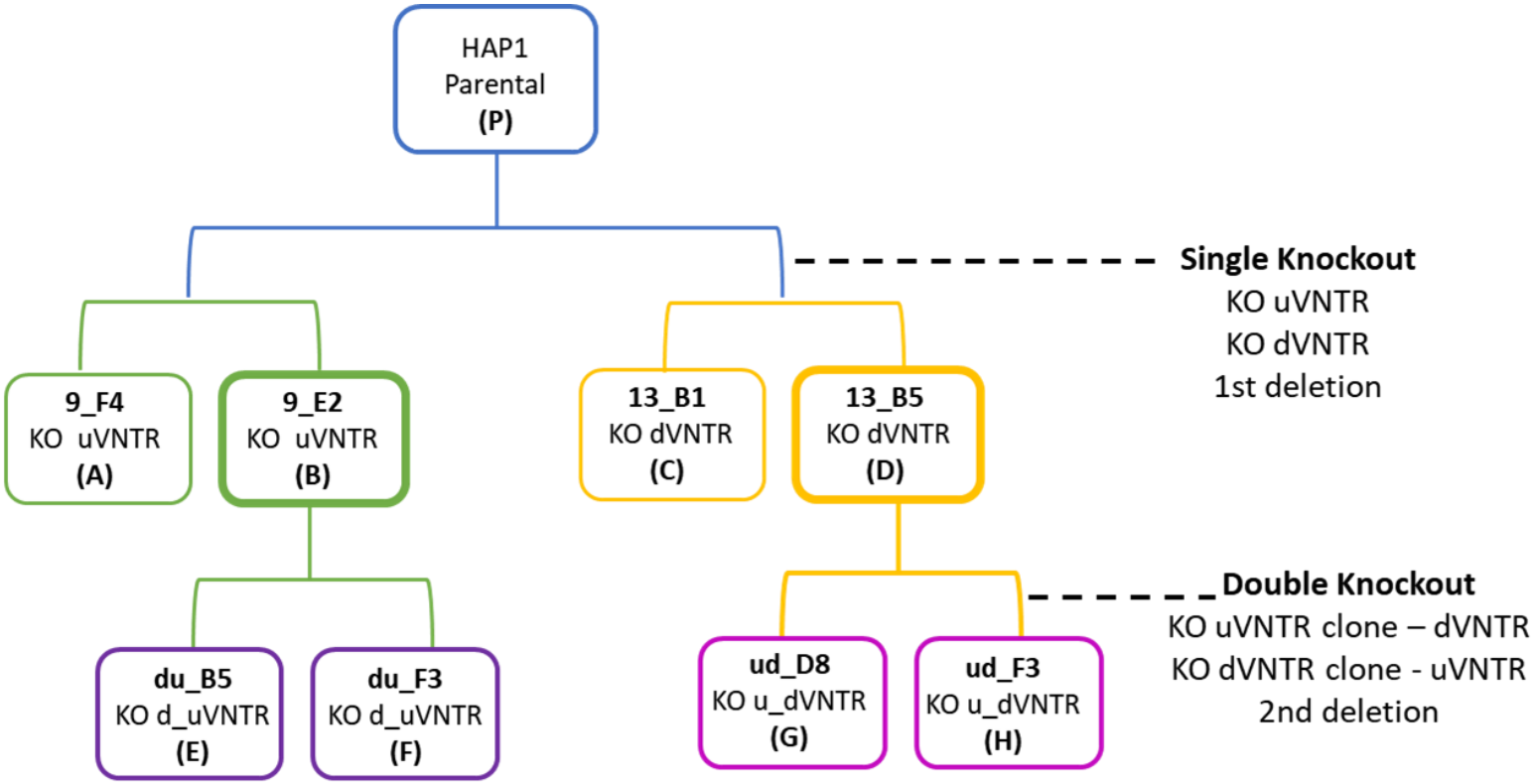


Figure 4.8. Schematics of HAP1 cell clones generation. The schematics of HAP1 cells explaining the clones were generated from the HAP1 parental cell line (P). Fist single VNTR knockout (KO) was generated, by deletion of a single VNTR, resulting in duplicate deletion cells for each VNTR: Clones 9_F4 (A) and 9_E2 (B) uVNTR KO, and clones 13_B1 (C) and 13_b5 (D) dVNTR KO. A second deletion was performed, to obtain double knockout clones, lacking uVNTR and dVNTR, by selecting a single KO clone of each VNTR, uVNTR KO 9_E2 (B), resulting in the double KO clones du_B5 (E) and du_F3 (F), and dVNTR KO 13_b5 (D) originated the DKO clones ud_D8 (G) and ud_F3 (H).

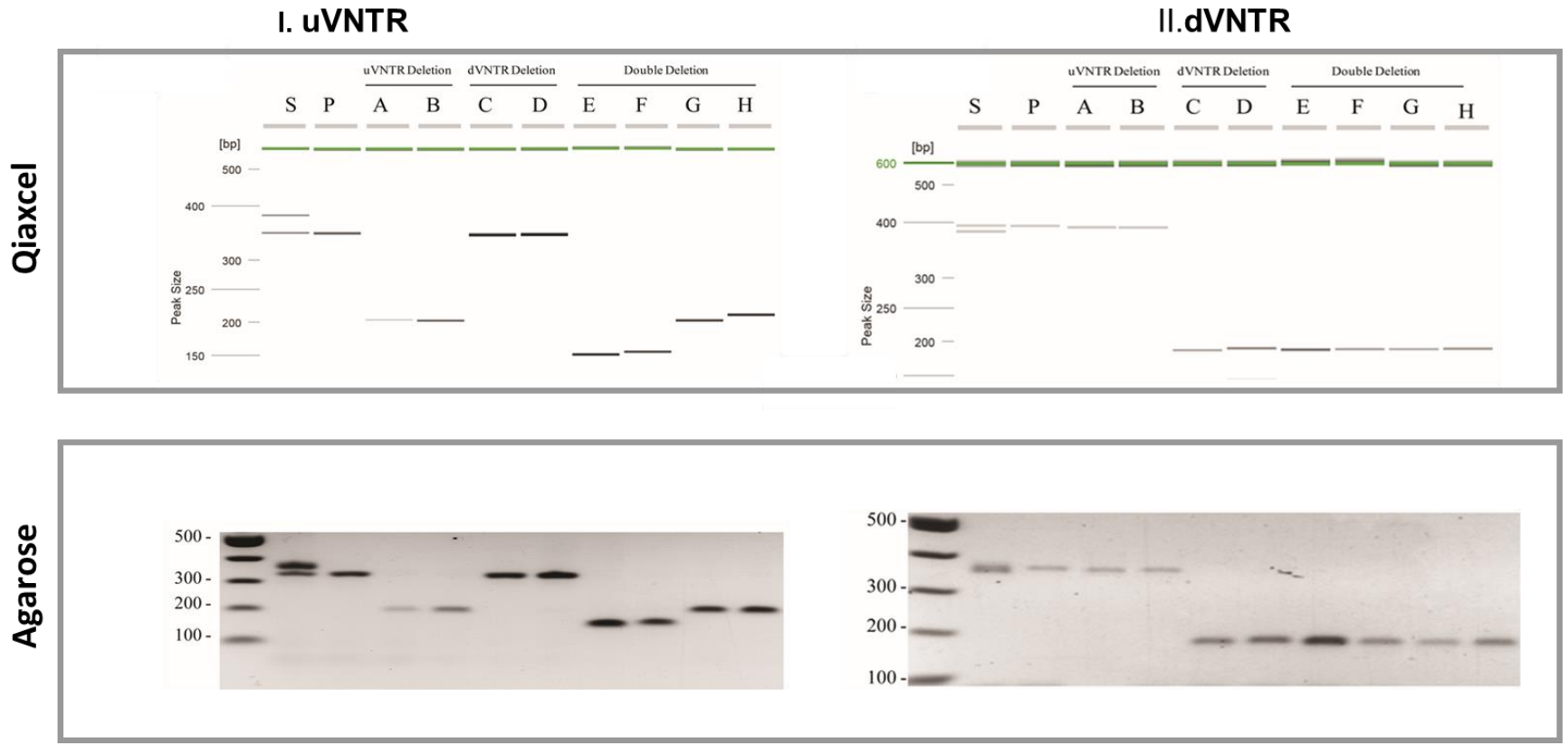


Figure 4.9. HAP1 cell line genotype. HAP1 cell clones generated by single or double knockout (KO) VNTR were submitted to uVNTR and dVNTR MAOA gene amplification. PCR products were run in agarose gels and QiaXcel (Qiagen) to confirm the KO. They are labelled at the top as S: SH-5YSY cell (heterogeneous to both VNTR), P: HAP1 parental cell line (P); Fist single VNTR knockout (KO) was generated, by deletion of a single VNTR, resulting in duplicate deletion cells for each VNTR: A and B: HAP1 with uVNTR KO, Clones 9_F4 (A) and 9_E2 (B); C and D: HAP1 dVNTR KO, clones 13_B1 (C) and 13_b5 (D); Double knockout clones, E and F: DKO clones du_B5 (E) and du_F3 (F); and G and H: DKO clones ud_D8 (G) and ud_F3 (H).

4.4.3 The expression of uVNTR and dVNTR of MAOA gene in HAP1 cells.

To evaluate the impact of uVNTR and dVNTR deletions on the MAOA expression in the HAP1 cell line, RNA was collected and transcribed to cDNA.

The cell lines utilised were cultured at the same conditions as the HAP1 parental, as described in section 2.2.8.1.

Amplifications were performed as explained in section 2.2.13.4, using different primer sets were used to evaluate the expression of the various MAOA mRNA isoforms, the primers for exon III-VI, exon I-IIA, uVNTR and β-actin as housekeeping.

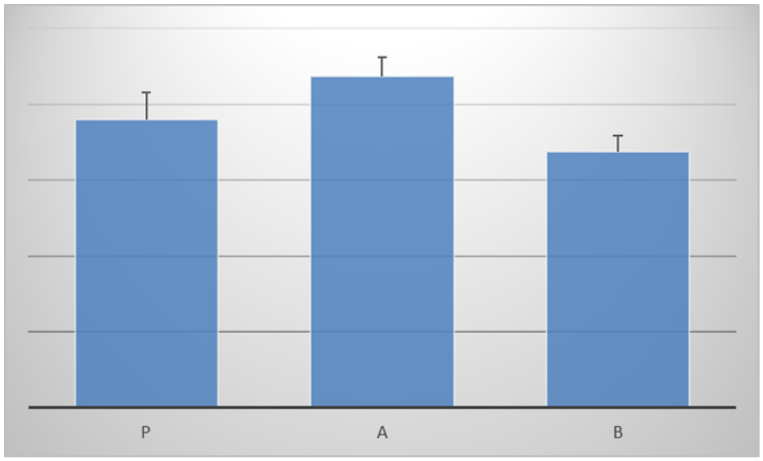
The set of primers that amplify from exon III-VI of the MAOA gene, as it encompasses both the long and short MAOA isoforms.

The comparison between the parental HAP1 cell line and the uVNTR single KO, shown in Figure 4.10, reveals a little variation in the expression supported by the two distinct clones: clone A exhibits a low increase in MAOA expression, whilst clone B displays a low decrease. According to the Student's T-Test, none substantially differs from the expression of the parental cell line. Nevertheless, statistically the difference between the two single KO clones (A vs. B) is significant ($p<0.05$).

Nevertheless, the clones with single deletions of either the clones A and B with KO uVNTR, Figure 4.10, or clones C and D of KO dVNTR, Figure 4.11, had higher levels of expression with the increase of around 3-fold and 2-fold, respectively. Moreover, the data obtained from the DKO derived from clone C (clones G and H), Figure 4.12, supported an additive effect on the expression of this MAOA isoform. Even though the same tendency was observed for KO clones originating from clone B (clones E and F), Figure 4.12, the increase was not statistically significant compared to clone B, used to obtain the DKO.

Unlike the uVNTR KO clones, the dVNTR KO clones C and D seem to support MAOA expression differently. Both cell lines exhibit a substantial decrease in expression relative to the parental cell line (Figure 4.13). This tendency, which is the exact reverse of that suggested by the uVNTR KO clones, clearly shows that the dVNTR plays a more significant role in the regulatory pattern of the MAOA expression than the primary acknowledged marker uVNTR.

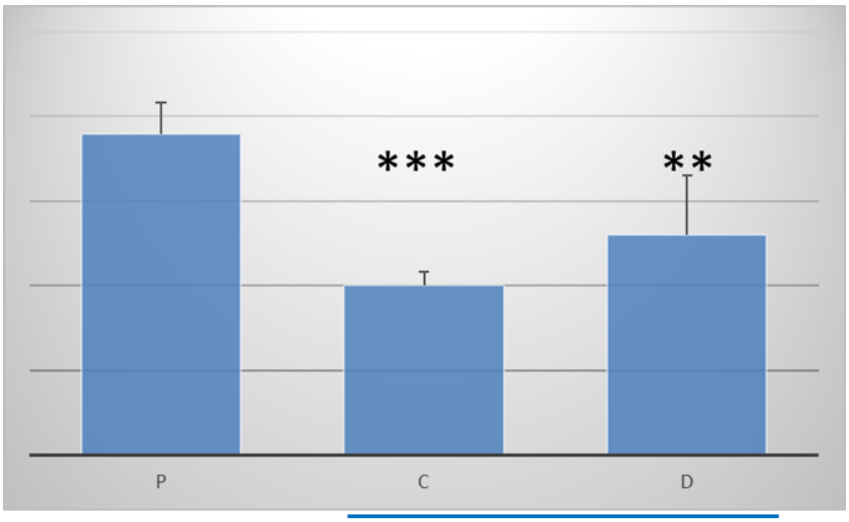
The expression pattern of the four DKO is more constant. As seen in Figure 4.13, the MAOA mRNA expression in the four double KO cell lines is significantly decreased ($p<0.01$) compared to the parental cell line. Furthermore, the expression pattern across the four cell lines is identical.



P vs	A	B
	0.126	0.175
	0.252	0.351

A vs	B
	0.013
	0.026

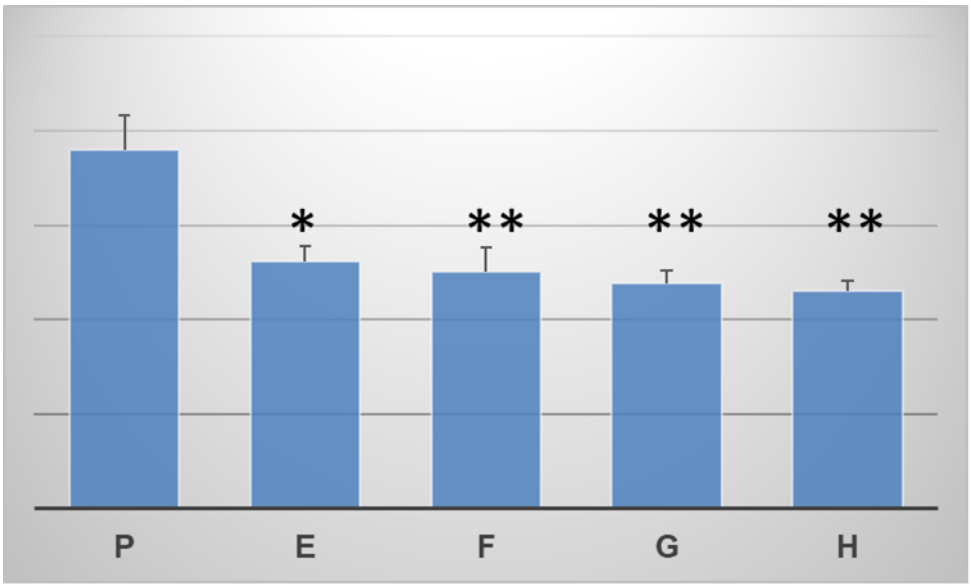
Figure 4.10. MAOA expression in HAP1 cell line with the uVNTR deletion clones. The absolute expression level of MAOA mRNA at basal condition, to region from exon III to exon VI of the MAOA gene. For each clone N=4. P is the parental cell line, A and B are the uVNTR single deletion clones: 9_F4 and 9_E2 in that order. The table at the bottom provides the p-values obtained with the one- and two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001.



P vs	C	D
	0.002	0.008
	0.004	0.015

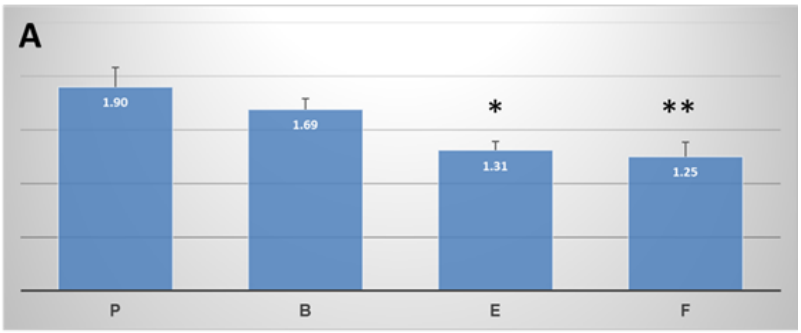
C vs	D
	0.061
	0.121

Figure 4.11. MAOA expression in HAP1 cell line with the dVNTR deletion clones. The absolute expression level of MAOA mRNA at basal condition to region from exon III to exon VI of the MAOA gene. For each clone N of 4 repetitions. P is the parental cell line, and C and D are the dVNTR single deletion clones: 13_B5 and 13_B1, respectively. The table at the bottom provides the p-values obtained with the one- and two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001.



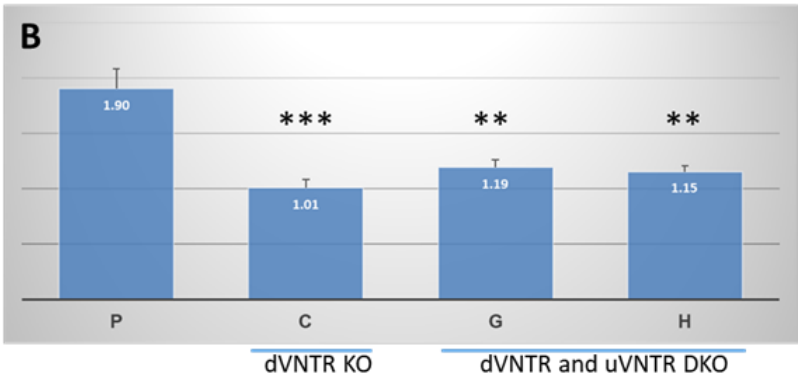
P vs	E	F	G	H
	0.012	0.014	0.005	0.004
	0.024	0.027	0.010	0.007

Figure 4.12. MAOA expression in HAP1 cell lines the double deletion clones. The absolute expression level of MAOA mRNA at basal conditions to region from exon III to exon VI of the MAOA gene. For each clone, N=4. P is the parental cell line, E, F, G and H are the MAOA VNTRs double KO clones: du_B5, du_F3, ud_D8 and ud_F3 in that order. The table at the bottom provides the p-values obtained with the one- and two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001.



P vs	B	E	F
	0.175	**0.012**	**0.014**
	0.351	**0.024**	**0.027**

B vs	E	F
	0.014	**0.021**
	0.027	**0.041**



P vs	C	G	H
	0.002	**0.005**	**0.004**
	0.004	**0.010**	**0.007**

C vs	G	H
	0.064	0.082
	0.128	0.165

Figure 4.13. MAOA expression in HAP1 cell line – single deletion clones and related double KO clones. The absolute expression level of MAOA mRNA at basal condition to region from exon III to exon VI of the MAOA gene. For each clone, N=4. **A.** P is the HAP1 parental cell line, B is a single uVNTR deletion clone, and E and F are the MAOA VNTRs double KO clones generated from Clone B: du_B5, du_F3, respectively. **B.** P is the parental cell line, C is a single dVNTR deletion clone, and G and H are the MAOA VNTRs double KO clones: ud_D8 and ud_F3, respectively. The table at the bottom provides the p-values obtained with the one- and two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001.

The same pattern was observed with the set of primers exon I-IIA when these results were compared to those obtained with uVNTR ($p<0.001$).

The evaluations performed with the DKO clones (E, F, G and H) of isoform 204 and the HAP1 parental cell line (P) resulted in a significant increase in expression (Figure 4.14 A-D).

Our data shows that VNTRs can exhibit isoform-specific modes of regulation, and multiple VNTRs can have an additive effect on gene expression profiles. The dVNTR and uVNTR were found upstream of the monoamine oxidase A (MAOA) gene and confirmed the single knockout (KO) of the dVNTR and the double KO of the dVNTR and uVNTR both led to a significant decrease in total MAOA expression.

However, our measured expression of the minor alternative isoform of MAOA (isoform 204) shows that single KOs of each VNTR can originate a significant increase in the expression of this isoform, and also the DKO induced a further increase in expression, Figure 4.15. All the evidence collected points to the importance of isoform-specific regulatory effects and an additive effect of the two VNTRs in the MAOA gene expression.

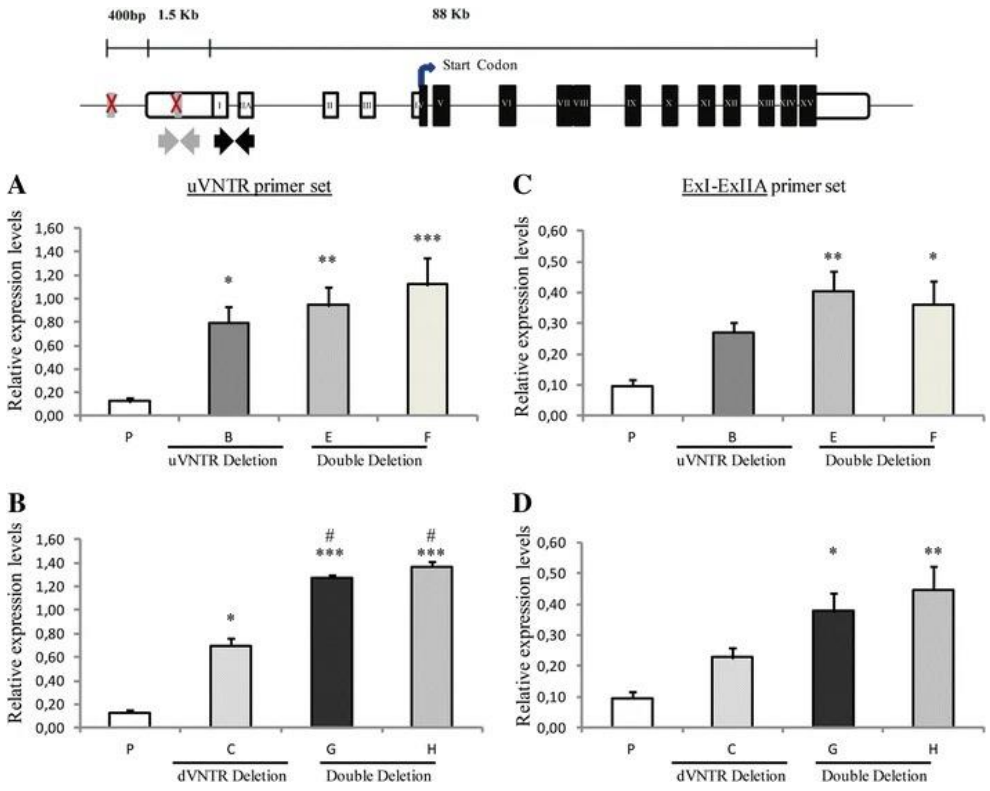


Figure 4.14. MAOA minor isoform expression in HAP1 cell lines under basal condition – KO clones and related DKO clones. The top illustration of MAOA gene is reported in the UCSC genome browser Hg38 and Hg19 version. The 5' and 3' untranslated regions (UTRs) are represented by white blocks, and black boxes are the exons. The curved black arrow indicates the translational start site. Black arrows are the region Exon I to Exon IIA and grey arrows uVNTR, representing the primers forward and reverse binding sites. P is the parental cell line, B is a uVNTR KO clone which has generated the DKO E(du_B5) and F (du_F3), and the C dVNTR KO clone, generated the DKO G (ud_D8) and H(ud_F3). **A and B)** MAOA expression analysis with the uVNTR primer set. **C and D)**. MAOA expression analysis to the region Exon I to Exon IIA. Statistical analysis: #,*p<0.05, ##,**p<0.01, ###,***p<0.001 of univariate analysis followed by a post hoc Bonferroni test for analyses between more than two groups. # represents significance against the single KO cell line; * represents significance against the parental cell line. All values are expressed as mean ± SEM. For each clone N=4. All values were normalised to β-Actin.

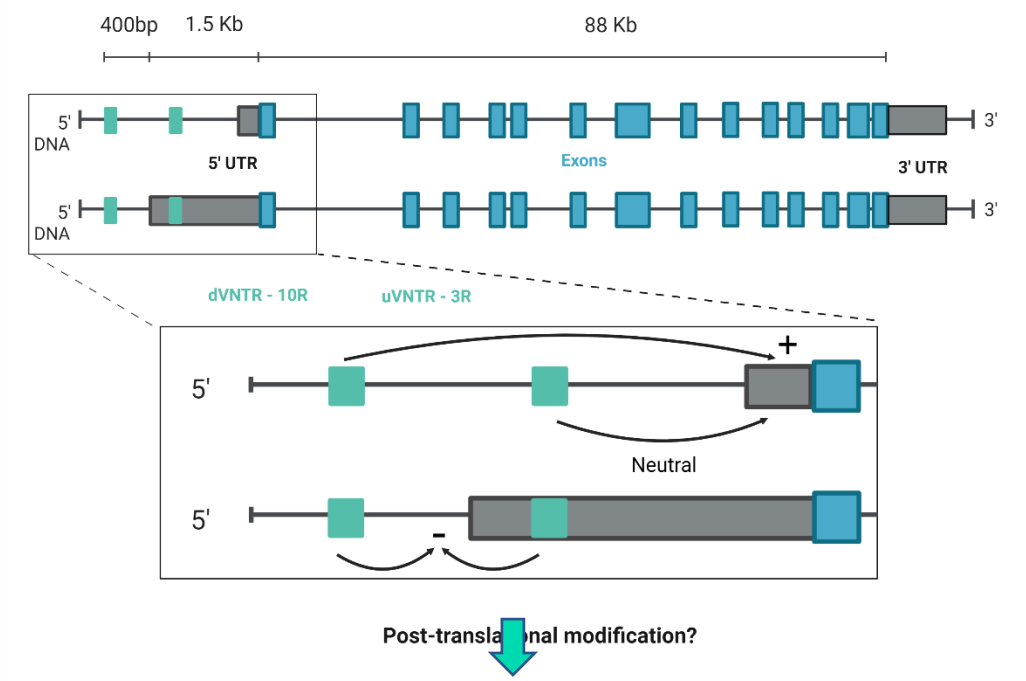


Figure 4.15. The regulation model for monoamine oxidase A (MAOA) expression by the uVNTR and dVNTR. The schematic representation of the MAOA promoter region and the regulatory effect employed by the uVNTR and dVNTR on the two MAOA isoforms. White box exon I; From left to right: Grey bars represent MAOA dVNTR and uVNTR, and overlapped black boxes represent 5'UTRs of the two isoforms. Curved black lines represent the transcriptional start sites (TSSs) for the two MAOA isoforms. Source: Adapted from BioRender tools

4.4.4 MAOA VNTR polymorphisms in the Wirral Child Health and Development (WCHADS) cohort.

4.4.4.1 MAOA VNTRs genotyping

The genomic DNA used to investigate both VNTRs genotypes originated from the collaboration with the Wirral Child Health and Development Study (WCHADS).

Saliva samples were collected, using the established study protocol, and the genomic DNA was extracted as described in section 2.2.2. in these analyses were employed a total of 283 samples, including both females and males. The cohort has been described in section 2.1.4.

The protocol for the genotyping of the MAOA uVNTR has been well established, and it was performed as described in section 2.2.8 [30, 169]. Nevertheless, the genotyping protocol for the MAOA dVNTR, described by Philibert et al., 2011, was quite challenging. First, we tried to replicate the PCR protocol described by Philibert et al., 2001. After several failed attempts, the alternative was to design new primers. Sadly, it was unable to obtain results with conventional PCR protocols, as dVNTR is found in a region rich in CG, yet very repetitive, which increases the difficulty of amplifying this fragment.

The alternative was to improve PCR performance by incorporating three additives to the PCR reagent mix to avoid intramolecular stable stem-loops in GC-rich DNA sequences. We optimised the protocol by combining Betaine, DMSO and 7-deaza-dGTP, to enhance the polymerase chain reaction (PCR) amplification [245]. Betaine lowers the melting temperature, the most common additive used to increase the amplification of sequences rich in GC pairing. Nevertheless, producing a gel with clear amplification was difficult even using betaine. We added dimethyl sulfoxide (DMSO) to disrupt base pairing [246, 247]. The third modification was the use of 7-Deaza-2′-deoxyguanosine 5′-triphosphate (7-deaza-dGTP), which pairs weakly with conventional bases, acts as a dGTP analogue,

reducing the strength of DNA secondary structures, although still utilised by the polymerase

In addition, we modified the cycling conditions. The annealing temperature was decreased by 1°C every cycle, starting from 65 °C e down to 55°C and running a further 35 cycles at 55C [248]. The improvement in the amplification can be observed in figure 4.16.

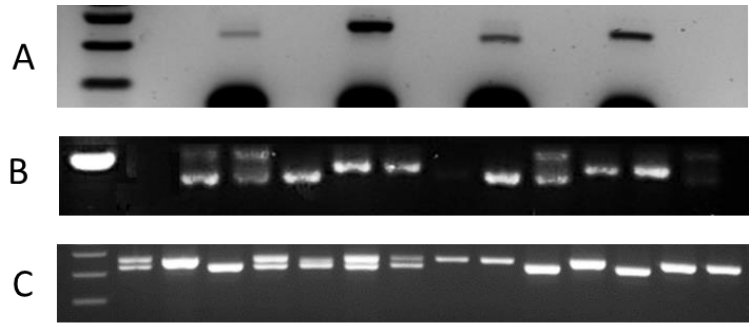


Figure 4.16. Image of agarose gel from electrophoresis of PCR products of *MAOA* gene dVNTR. The electrophoresis was run at a lower voltage (***80 V***), using an agarose gel of 2%. A) The standard PCR reaction and cycling conditions with betaine. B) New reaction with betaine, DMSO, modified dNTPs concentrations, and standard cycling conditions. C) New validated reaction and touchdown cycling conditions.

The MAOA dVNTR PCR reactions for validating were set up using primers described in table 2.1, and using touchdown PCR thermal cycling conditions are reported in section 2.2, with the annealing step from 65°C down to 55 °C over 10 cycles, followed by 35 cycles at 55°C annealing temperature. PCR products were analysed by electrophoresis on a 2% agarose gel, incorporating GelRed™ (Biotium). The same PCR product were analised by capillary electrophoresis (ABI 3130 Life Technologies, Paisley, UK) and using the Genemapper V4.0 (Life Technologies) to the genotypes calls. The results from both methods were analysed blind from each other, and the results are observed in table 4.3.

Table 4.3: Distribution of MAOA uVNTR and dVNTR genotype in the WCHADS cohort. Data of both genders are included, however heterozygous are only present to females. To both genders the most abundant (common) allele is the 4R followed by the 3R. For the dVNTR the most common is the 9R followed by the 10R.

Genotype	Female		Male	
uVNTR	N	%	N	%
3, 3	18	15	46	34
3, 3.5	1	1	*	*
3, 4	47	39	*	*
3, 5	1	1	*	*
3.5, 3.5	1	1	2	1
3.5, 4	1	1		
4, 4	48	40	83	61
4, 5	2	2	*	*
5, 5	0	0	4	3
Total	119		135	

Genotype	Female		Male	
dVNTR	N	%	N	%
9, 9	71	60	101	75
9, 10	34	29	*	*
10, 10	9	8	26	19
10, 11	3	3	*	*
11, 11	2	2	8	6
Total	119		135	

*hemizygous

4.4.4.2 The haplotype of distal and proximal VNTRs in the population.

The haplotype block analysis of dVNTR and uVNTR may constitute other associations or stratifications. The dual VNTR locus was analysed, and allele frequencies for each are shown in Table 4.4.

First, we performed the Hardy-Weinberg equilibrium (HWE) analysis in the female arm. As the males are hemizygous to the locus, they were not included in the analysis. Because of the multi-allelic nature of these two loci, only the common alleles in a bi-allelic system were included.

We analysed the common alleles 3R and 4R of uVNTR and 9R and 10R dVNTR, respectively.

This was further analysed in collaboration , utilising a tool for constructing the haplotype blocks, aiming to verify the inter-allelic linkage disequilibrium that accounted for poly-allelic markers and using the Multiallelic Interallelic Disequilibrium Analysis Software (MIDAS) [249].

Therefore the data were stratified by gender and included the common and rare alleles assigned based on the inferred haplotype blocks. Analysing females into haplotype categories is less straightforward because each woman possesses two X chromosomes. Thus, female participants presented double alleles compared to males. This permitted an evaluation of the expected versus the observed frequency of both common and rare VNTR haplotypes, presuming independent segregation (Table 4.4).

The two common alleles for each locus are included in the haplotype blocks, considering for dVNTR, the 9R and 10R and for uVNTR, the 4R and 3R, significantly diverged from their predicted haplotype frequencies, that observation was emphasised by the significant adjusted chi2 values noted.

Table 4.4: Haplotype distribution in the Wirral child health and development study (WCHADS). Assuming a multiallelic system, adjusted χ2 values were derived from the observed versus expected frequencies built on the allele frequencies obtained in this data. Frequencies and corresponding D′ and correlation r2 values for each haplotype described below were calculated make use of the MIDAS package (Gaunt et al. 2006). haplotype in our cohort. A) Females (2n = 252) and (B) males (n = 136). * *n/a there are no observed frequencies for that.*

A

dVNTR/ uVNTR	Frequency		Yates Chi2	D'	R^2	Haplotype counts
	% Observed	% Expected				
9R:4R	59.13	45.15	**47.8**	0.833	0.402	149
10R:3R	20.07	7.65	**47.1**	0.901	0.399	51
9R:3R	11.77	26.22	**55.5**	-0.846	0.466	29
11R:3R	3.16	1.70	1.36	0.476	0.024	9
9R:5R	1.59	1.17	0.002	1	0.058	4
10R:4R	**1.19**	**13.18**	**41.0**	-0.894	0.348	3
11R:4R	1.19	2.93	0.87	-0.429	0.015	3
9R:3.5R	1.19	0.87	0.04	1	0.044	3
8R:3R	0.40	0.14	0.28	1	0.007	1
10R:3.5R	0.00	0.26	n/a	-0.627	0.001	0
8R:4R	0.00	0.24	0.31	-1	0.006	0
8R:5R	0.00	0.01	n/a	-1	6.00E-05	0
10R:5R	0.00	0.34	0.02	-1	0.044	0
11R:5R	0.00	0.08	n/a	-1	0.008	0
8R:3.5R	0.00	0.00	n/a	-1	5.00E-05	0
11R:3.5R	0.00	0.06	n/a	-1	0.0006	0
						252

B

dVNTR/uVNTR	Frequency		Yates Chi2	D'	R2	Haplotype counts
	% Observed	% Expected				
9R:4R	60.29	46.05	**62.1**	0.952	0.481	82
10R:3R	18.38	6.47	**52.4**	0.942	0.410	25
9R:3R	9.56	25.37	**77.3**	-0.956	0.595	13
11R:3R	5.88	1.99	13.6	1	0.122	8
9R:5R	2.94	2.21	0.34	1	0.010	4
9R:3.5R	2.21	1.38	0.03	1	0.006	3
10R:4R	**0.74**	**11.74**	**42.0**	-0.937	0.330	1
10R:3.5R	0	0.35	0.001	-1	0.004	0
11R:3.5R	0	0.11	n/a	-1	0.001	0
11R:4R	0	3.61	10.9	-1	0.099	0
10R:5R	0	0.56	0.12	-1	0.007	0
11R:5R	0	0.17	n/a	-1	0.002	0
						136

Predicting a multiallelic system for both markers, the haplotype constructs, frequency, and distribution are illustrated in **Figure 4.4** (n=262).

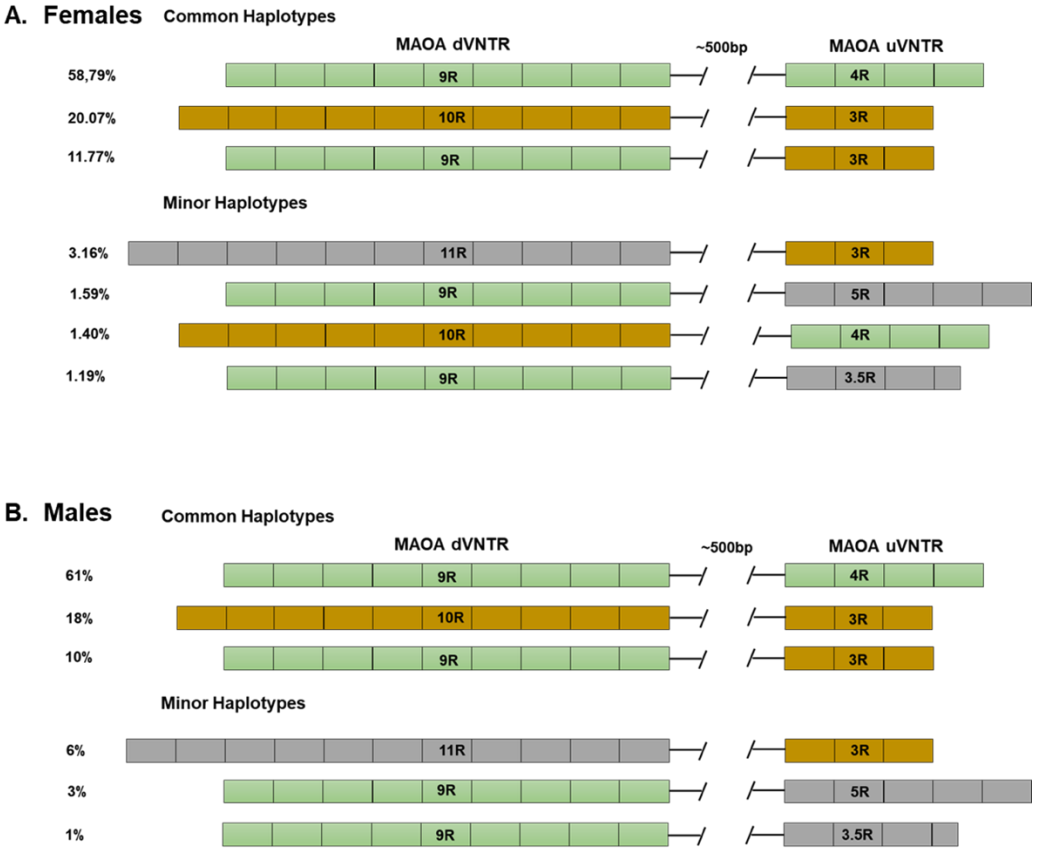
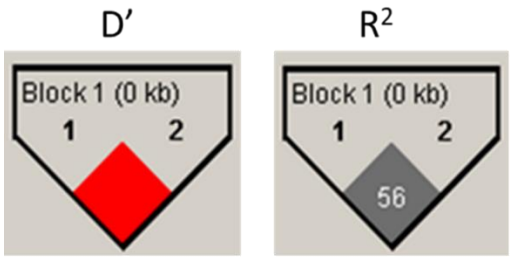


Figure 4.17. Haplotype distribution in the Wirral child health and development study (WCHADS). Assuming a multiallelic system for both markers, the haplotype constructs, frequency, and distribution are illustrated above (n = 262). Haplotypes were initially inferred from males and then the Midas package (Gaunt et al. 2006) was used to work out frequencies in females. Hardy-Weinberg equilibrium (HWE) test for females assumed a biallelic system (p = 0.82 for uVNTR; and p = 0.02 for dVNTR). A) Females; B) Males. NB* HWE test was not conducted for males due to the locus hemizygosity

The haplotype distribution of the alleles in Children of the WCHADS in shown in Table 4.3. the data demonstrates, in these children, the 4R repeat at uVNTR is in linkage disequilibrium with the 9R repeat at MAOA dVNTR. However, the 3R allele at uVNTR

can be further stratified, as the 3R can be associated with the 9R, 10R and 11R dVNTR genotypes. Some alleles that would be categorised as low-activity based on their uVNTR allele might not be categorised as low-activity based on their dVNTR allele, as demonstrated by Philbert et al, 2011, which also discuss the existence of significant linkage disequilibrium (LD) between these two loci, implying a lack of recombination among the two markers, with the alleles expected to separate as part of the same block. Only 3 of the 4 possible combination haplotypes were regularly observed. The incidence of the haplotype formed by the 10R and 4R alleles, of the dVNTR and uVNTR respectively, was rarely observed.



Parameter	uVNTR (2) – dVNTR (1) Relationship	Measure
Linkage Disequilibrium	D prime (D')	1.00
	R squared (R^2)	0.569
	LOD	19.23
Haplotypes	4R-9R (block 4-1)	65.3%
	3R-10R (block 3-2)	23.4%
	3R-9R (block 3-1)	11.4%

Figure 4.18. Linkage disequilibrium between MAOA μVNTR and dVNTR using the Haploview 4.1 and analysing the samples of males and females. The figure above is a depiction of major haplotypes within the u- and dVNTRs of MAOA gene. It was observed that the haplotype blocks are stratified by the 3R, since it can be both inherited with 9R or 10R of dVNTR, whereas the 4R is only inherited with 9R of dVNTR. Frequency of the haplotypes 4R-9R (65.3%), 3R-10R (23.4%) and 3R-9R (11.4%), the haplotype 4R-10R, and no 4R-9R haplotype were observed. The value D' of 1.0 or (100%) R2 was 0.569 (56.9%), the value of 1.0 displays the complete dependence between the VNTRs.

The linkage disequilibrium and haplotype analysis (figure 4.18) permit to place of the 4R and 9R, uVNTR and dVNTR, respectively, and work with the stratification based on the 3R allele of uVNTR, which can form haplotype with the 10R, 9R and 11R of dVNTR.

To summarise, the data demonstrates that the most common haplotype is formed by the major allele of both markers, 9R and 4R, this haplotype makes up about 60% of our cohort data, for both genders, 59.13% in females and 60.29% in males.

The second most common haplotype is the 10R-3R, with a frequency of 20.07% in females and 18.38% in males, followed by the 9R-3R haplotype, with 11.77% in females and 9.56% in males. The rare haplotype formed by the 10R-4R, was observed with a frequency of 1%. although the expected frequency was above 10%.

To the minor alleles found in the analysis of haplotype distribution, the most frequent haplotype combination was the 3R variant of uVNTR and the 11R dVNTR

followed by the 3.5R and 5R of uVNTR that were only linked with the 9R variant of dVNTR. No significant difference in gender differences in the distribution of haplotypes between males and females (t-test $p>0.05$) was observed.

4.5 Discussion Chapter 4

4.5.1 The role of uVNTR and dVNTR in the regulation of MAOA gene expression and how it relates to mental health.

The MAOA uVNTR is well discussed in the literature regarding the G × E interactions in the. The recognised theory is that the different repeat copy numbers of the uVNTR charge the different expressions of uVNTR in response to the environment. This is a reasonably simple justification for the male gender, which possesses only one allele for MAOA. Nevertheless, the situation is more complex in females for various reasons, including the gene dosage, the potential for escape X-inactivation, and the heterozygosity of the uVNTR, whose genotype may direct distinct MAOA gene expression patterns.

Here investigated the molecular processes underlying MAOA expression in the SH-SY5Y, a human neuroblastoma cell line heterozygous for the uVNTR; the heterozygosity of the uVNTR is required to permits us to relate mRNA expression from each allele of uVNTR MAOA. Followed by the epigenetic marks over the uVNTR and transcription factor binding to the locus.

The finds revealed activity in both alleles of uVNTR using the SH-SY5Y cells, although by significantly distinct molecular mechanisms established on the methylation marks and binding of proteins over this region. Most outstandingly, even though both alleles had very diverse chromatin structures (methylation and transcription factor interactions), when ChIP was executed for an active RNA polymerase, binding was detected at both alleles, which was consistent with the cDNA studies showing that mutual alleles were active, in spite of the distinct chromatin architecture, Figure 4.4A.

The methylation of a single allele which apparently did not repress MAOA expression, Figure 4.3, would be consistent with current work on X chromosome gene expression profiles which showed that methylation did not essentially correlate with gene repression

but somewhat operated as a parameter disturbing a region's capacity to escape X inactivation [163]

To further explore the regulation of MAOA gene expression by the uVNTR, we examined the binding of five transcription factors: CTCF, SP1, hnRNP K, CNBP and nucleolin. Our data delivered indication that these proteins recognised the proximal MAOA promoter surrounding the uVNTR, and more significantly, a distinct pattern of binding was noticed over each allele which was modulated both by the genotype of the uVNTR, e confirmed with the exposure of the cells to sodium valproate, Figure 4.4C.

The importance of gender in the regulation of MAOA expression through epigenetic mechanisms. Wong et al. demonstrated the gender-specific alterations of methylation in a longitudinal study of twins. [250] Domschke et al. similarly showed hypomethylation in the pathogenesis of panic disorder, particularly in female. [251] Both studies were consistent with our finds.

Our data could enlighten the modifications in uVNTR genotype association for definite mental health issues when comparing males and females. Such examples include MAOA uVNTR moderating the association between childhood mistreatment and several disorders, as an example the dysthymia only present in females, [168] gender alterations comprising noticeable effects on anxiety and Attention-deficit hyperactivity disorder (ADHD) in females, however, to bipolar disorder, autism and aggressive behaviour the males are perceptible more affected.[139, 154] Indeed, many MAOA G × E studies report only on males due to compounding effects, including female heterozygosity for the uVNTR. [30, 135, 165] MAOA gene is one of the roughly 15% of X chromosome genes predictable to escape X inactivation, and that alone can further complicate the regulation.[163, 164, 238]

We still have a considerable discussion regarding X inactivation in vivo, as the X inactivation of the second allele in heterozygotes may not depend on G x E interactions,

as well as the gene called a master regulator of X chromosome inactivation, the gene X Inactive Specific Transcript (XIST), is overexpressed in women with major affective disorders, which is consistent with this hypothesis. [162] Thus, gender- and disorder-based differences, especially in mental health, are likely, and this study highlights a potential mechanism to explain these differences.

4.5.2 Allelic-specific regulation of uVNTR MAOA expression

The study of this section aimed to investigate the role of two VNTRs in the MAOA promoter in directing expression from each of the TSS.

To overcome the difficult task of analysing the gene expression of a gene located at the X chromosome, due to gender bias, the approach utilised a near-haploid cell line HAP1, which was then submitted to CRISPR/Cas9 editing, to generate cell lines with a single knockout of uVNTR and dVNTR, and double knockout by deleting both VNTRs and compare to the parental HAP1 cell line. This permitted us to observe the expression patterns of each VNTRs and the distinction between MAOA isoforms in the presence and when lacking one or both VNTRs.

Bioinformatics searches using data from hg38, provided ENCODE data and transcriptomic mapping of the MAOA gene, revealed multiple isoforms to this gene, and the data generated in this study was able to confirm that, as two main isoforms were observed. The namely primary isoform, was the established MAOA isoform, comprising 15 exons, and a shorter 5' UTR (Figure 4.2, isoform 201), and an isoform, with an extra exon termed IIA, compromising 16 exons, and an elongated 5' UTR, contained in this sequence the uVNTR.

In our experiment, no significant modulation in the expression of total MAOA mRNA was observed when using the in vitro models of single uVNTR knockout, under basal growth conditions. Nevertheless, it was observed a significantly reduced expression of total

MAOA mRNA (Figure 4.10 B) within the dVNTR deletion, under the same conditions. The result indicates it is a positive regulator of the primary MAOA isoform, this observation has not been previously described. After that impact of the double knockout was addressed, and the finds were in an agreement with the single VNTRs KO.

Interestingly, the main mediator to the expression of the primary MAOA isoform in the DKO cell lines was the dVNTR (Figure 4), been even significantly lower than the HAP1 parental cell line. However, we could not find significant changes comparing the expression of the double KO clones and the single dVNTR deletion (Figure 4.10 B).

This data provides evidence of the major role of dVNTR in driving the gene expression of the primary MAOA isoform, further explained by the important reduction of this isoform found in the double KOs originated from the uVNTR KO (Figure 4.10 A).

The isoform 204, which contained the uVNTR within its 5' UTR (Figure 4.11 A - D), showed that deletion of either the dVNTR or uVNTR resulted in an increased expression. Consequently, both VNTRs individually appeared to act as negative regulators of this isoform.

The VNTR DKO equally significantly increased the expression of this transcript relating to the parental cell line (P) (Figures 4.11 C - D). When performed a PCR analysis, targeting the uVNTR (Figure 4.11 A - B)., and the result was replicated.

Our finds are coherent with previous data from Philibert et al, 2011, demonstrating that distinct combinations of the dVNTR and uVNTR (9R:3R and 9R:4R) encouraged distinct reporter gene expression patterns [159] as it implied that the haplotype of the d- and u-VNTRs could account for significant differential regulation of MAOA expression.

The DNA sequences of the dVNTR and uVNTR are quite different and are consequently expected to bind a different set of transcription factors and intermediate a differential response to the same challenge, our group and others have demonstrated for the well-

established regulatory VNTR in the SLC6A4 and SLC6A3 genes (Michelhaugh et al., 2001, Vasiliou et al., 2012).

These data indicate that the uVNTR and dVNTR are part of the same haplotype block, and that alone can contribute to the analysis performed, where only the uVNTR is taken into consideration, as it is inevitably associated with the dVNTR, this could enlighten the investigation of diverse disorders associated to the central neuro system. Still, this needed to the investigated in larger, well prefilled and diverse cohorts to address the frequencies of the dVNTR and uVNTR, to be able to understand how this association can contribute to mental health.

4.6 Conclusion Chapter 4

The focus of MAOA regulation has always been attributed exclusively to the uVNTR. Several studies and publications are considering the only to account for the variations to the expression of MAOA and the only also to address the possibility of finding answers in neuropsychiatric conditions, behaviour, aggressiveness, and others. Here we discuss the contribution of dVNTR to MAOA gene expressions and the import synergism between the two markers. However, our data from the double knockout VNTRs did not prevent the expression of the gene. Thus, it was able to provide significantly lower levels of mRNA transcripts. The primary transcript isoform seemed to have the dVNTR as a major regulator, other hand and the minor transcript isoform 204, Figure 4.13, had a negatively regulate expression by both VNTRs.

Our data was generated using a cell line able to express both isoforms of MAOA, HAP1 cells, and it is a haploid karyotype, which enabled successful CRISPR deletion of regulatory elements. This model supports combining both VNTRs in MAOA regulation, considering that the expression of MAOA can be differentiated depending on the inducible stimulus manner, tissue-specific or regulation in a neuronal context.

The conservations of both MAOA VNTRs, show that they are primate-specific, which proposes different key regulatory domains for MAOA expression in mammals. Additionally, work from Warburton et al., 2016 [252] demonstrated a dual function attribute to a VNTR in the mir137 gene, depending on the expressed mRNA isoform. Similarly, our data also indicate that the uVNTR could have a dual function in both transcription and post-transcriptional regulation of the MAOA gene as it is contained within the 5'UTR of one of the putative coding isoforms.

In conclusion, these data provide important considerations for understanding the regulation of MAOA expression and how it can act in the modulation by the VNTRS.

Our work has also shown that VNTRs can exhibit isoform-specific modes of regulation, and multiple VNTRs can have an additive effect on gene expression profiles. They assessed two VNTRs (dVNTR and uVNTR) found upstream of the monoamine oxidase A (MAOA) gene and discovered that a single knockout (KO) of the dVNTR and a double KO of the dVNTR and uVNTR both led to a significant decrease in total MAOA expression. They also measured the expression of a minor alternative isoform of MAOA. They found that single KOs of each VNTR led to a significant increase in the expression of this isoform. A double KO induced a further increase in expression, highlighting isoform-specific regulatory effects and an additive effect of the two VNTRs at this locus [20].

Chapter 5

The characterisation of MAOA VNTRs in Schizophrenia.

5.1 INTRODUCTION

Accordingly to the WHO, about 1 in 300 people worldwide has been affected with schizophrenia, a severe psychiatric disorder with significant impairments in perception and changes in behaviour, including persistent delusions, hallucinations, disorganised thinking, highly disorganised behaviour, or extreme agitation, those symptoms are divided into three groups, positive, negative and cognitive, and have significant public health implications. People with schizophrenia may experience persistent difficulties with their cognitive functioning and have a life expectancy of ten to twenty years below that of the general population.

In addition, the pharmacological treatments available are effective for about half of the patients and aim to improve mainly positive symptoms, such as hallucinations and thought disorders, which are the core of the disease. In contrast, the negative symptoms such as flat affect and social withdrawal or cognitive symptoms resulting in learning and attention disorders remain unaffected. Asides from this, various effective treatment options include psychoeducation, family interventions, and psychosocial rehabilitation. Nonetheless, the cause of SZ remains to be understood entirely, and it is known that it is high heritability and is linked with complex polygenic factors. [253, 254]

Pharmacological research has shown that dysfunction of dopaminergic neurons could be involved in developing schizophrenia.[253] In the brain, dopamine degradation is catalysed by monoamine oxidase (MAO) and catechol-o-methyltransferase (COMT).[255, 256] Many studies have investigated the association between MAOA and Schizophrenia [256-259] as MAOA plays an essential role in mental illnesses such as schizophrenia.[260, 261]

Several studies associate uVNTR and SZ with an upstream (u) variable number of tandem repeats (VNTR) in the MAOA gene promoter. The majority evaluated the 3.5R

and 4R, which are the high-expression alleles, however, many of these results are inconsistent. [169, 262-264] The novel VNTR, namely distal (d)VNTR, also located in the MAOA gene promoter region, are associated with uVNTR. Koks et al., found that nicotine dependence is associated with the haplotype of dVNTR and uVNTR in association with neuropsychiatric disorders. [265] Also, haplotype was associated with schizophrenia in females, which may increase the risk of development and help reveal the molecular mechanism of this disease. Tanifuji et al., observed that the haplotype of dVNTR and uVNTR was associated with an increased risk of developing schizophrenia in females. [266]

Here we explored the association between the two VNTRs located in the promoter region of MAOA using samples from a schizophrenia cohort and health control.

Our study has several limitations, as our sample size was relatively small, and our cohort was comprised of participants of European descent. Further studies with large sample sizes and different populations are required to validate our findings. Also, we did not have access to data to analyse the longitudinal effects or detailed history of the symptoms, such as whether negative or positive symptoms were dominant, also, information regards previous and current use of antipsychotic medication would benefit our research.

5.2 AIM

Our aim was to investigate the role of MAOA uVNTR and dVNTR in schizophrenia patients and controls and explore associations between MAOA gene VNTRs in SCZ.

To achieve this aim, specific goals were also defined:

I. Analyse the genotypes of VNTRs located in the monoaminergic genes MAOA in SCZ samples and controls.
II. Use the genotype and haplotype data to perform Clump analysis.
III. Perform association analysis comparing genotype and haplotype data to risk factors linked to SCZ.

5.3 RESULTS

5.3.11. Characterisation of MAOA genotype and haplotype in schizophrenia.

For the characterisation of MAOA genotype and haplotype, we used 250 control samples (male n=124 and female n=126), with sub-samples mean age 46.27 and age range from 19 to 72, and 262 samples from schizophrenia patients (male n=183 and female n=79), with mean age 37.66 years and range age from 18 to 71. All DNA samples were from the Schizophrenia cohort described in section 2.1.4.3.

Genotyping of uVNTR and dVNTR in the MAOA gene was performed using 10 ng of DNA from each sample, as described in sections 2.2.14.2 and 2.2.14.3, and using primers described in Table 2.1.

Functional analysis of the MAOA gene supported differential gene regulation dependent on both copy numbers of the VNTRs. We therefore hypothesised that the VNTRs copy number and the haplotype of u- and dVNTR might be an essential

influencing factors in developing neurological dysfunction, such as schizophrenia.

That might be especially true in combination with an environmental factor to modulate the function of the VNTRs. To test this hypothesis,we analysed the the genotype of both VNTRs in 287 schizophrenic patients and 332 healthy controls, as described in section 2.1.4. We found alleles from 2 to 4 copies of the 30 bp promoter uVNTR were identified in our samples, and as for the dVNTR, we found alleles from 8 to 12 copies of the 10 bp repeat. The distribution of genotype frequencies of males and females is outlined in Table 5.1 and Table 5.2.

5.3.2 Clump analysis

The data were analysed using Clump analysis [204]. Differences between the groups were analysed using $\chi 2$ to determine significant differences between the Controls and Cases, and the threshold for statistical significance was defined as $p < 0.05$.

Since the X chromosome contains the MAOA gene, we analysed each sex separately. The genotype and allelic frequency for dVNTR, and uVNTR, are shown in Tables 5.1, 5.2 and 5.3. Neither the genotype nor allelic frequency of the VNTRs significantly differed between the schizophrenia and control groups.

The results obtained from this statistical model showed no significant difference under the conventional chi-squared (X2) approach to the samples stratified by gender. Next, we analysed each marker seperately, uVNTR and dVNTR, and the allelic distribution of females to both of VNTRs. The data is shown in Tables 5.1, 5.2 and 5.3, respectively.

Table 5.1: Genotype distribution and Clump analysis in males. A) Genotype data for uVNTR and dVNTR of the MAOA gene in males of the schizophrenia cohort. B) Significance-testing genotype frequency data between schizophrenia cases and healthy controls using Clump analysis. Values represent adjusted P-values for permutation testing of the differences between cases and controls concerning genotype frequencies.

A

Genotype	Cases		Controls	
uVNTR	**N**	**Frequency (%)**	**N**	**Frequency (%)**
3	65	**35.71**	59	**41.26**
3.5	1	0.55	2	1.40
4	116	**63.74**	82	**57.34**
dVNTR				
9	105	**57.69**	81	**56.64**
10	56	**30.77**	46	**32.17**
11	18	9.89	14	9.79
12	3	1.65	2	1.40

B

VNTR	T1	T2	T3	T4
uVNTR	0.4050	0.2410	0.2410	0.4642
dVNTR	0.9920	0.9637	0.7874	0.9893

Table 5.2: Genotype distribution and Clump analysis in females. A) Genotype data for uVNTR and dVNTR of the MAOA gene in females of the schizophrenia cohort. B) Significance-testing genotype frequency data between schizophrenia cases and healthy controls using Clump analysis. Values represent adjusted P-values for permutation testing of the differences between cases and controls concerning genotype frequencies.

A

Genotype	Cases		Controls	
uVNTR	**N**	**Frequency (%)**	**N**	**Frequency (%)**
2,4	1	0.95	2	1.06
3,3	12	**11.43**	25	**13.23**
3,4	50	**47.62**	93	**49.21**
3,5	0	0.00	2	1.06
4, 4	42	**40.00**	67	**35.45**
dVNTR	**N**	**Frequency (%)**	**N**	**Frequency (%)**
8,8	0	0	3	1.59
9,9	52	**49.52**	79	**41.80**
9,10	23	**21.90**	40	**21.16**
9,11	5	4.76	13	6.88
9,12	0	0	1	0.53
10, 10	14	**13.33**	26	**13.76**
10,11	8	7.62	21	11.11
10, 12	0	0	2	1.06
11, 11	3	2.86	2	1.06
11,12	0	0	2	1.06
N SZ/Control	105		189	

B

MAOA	T1	T2	T3	T4
uVNTR	0.7900	0.6614	0.4390	0.9217
dVNTR	0.5355	0.5429	0.2017	0.5423

Table 5.3: Allelic distribution of uVNTR and dVNTR of the MAOA gene in females of the schizophrenia cohort. A) uVNTR data**; B)** dVNTR data**; B)** Significance-testing allelic distribution data between schizophrenia cases and healthy controls using Clump analysis. Values represent adjusted P-values for permutation testing of the differences between cases and controls concerning genotype frequencies.

A *uVNTR*

	Cases		Controls	
Allelic distribuition	**N**	**Frequency (%)**	**N**	**Frequency (%)**
2	1	0.48	2	0.53
3	74	**35.24**	145	**38.36**
4	135	**64.29**	229	**60.58**
5	0	0	2	0.53
	210		378	

B *dVNTR*

	Cases		Controls	
Allelic distribuition	**N**	**Frequency (%)**	**N**	**Frequency (%)**
8	0	0	6	1.59
9	132	**62.86**	212	**56.08**
10	59	28.10	115	30.42
11	19	9.05	40	10.58
12	0	0	5	1.32
	210		378	

C

MAOA	T1	T2	T3	T4
uVNTR	0.6213	0.3756	0.3756	0.8721
dVNTR	0.10202	0.16996	0.11032	0.44436

5.3.3 Association analysis

The Assessment of cognitive endophenotypes to correlate our findings was performed by Dr Bettina Konte, University of Halle-Wittenberg, Halle, Germany. No significant results were observed for uVNTR and dVNTR genotypes and haplotypes for both genders versus affection status, Table 5.4. In association analysis, Affection status refers to the affection status line that gives each individual an affection status (i.e., 1 = unaffected, 2 = affected). After imputation, the batches were combined, and each VNTR genotype and the haplotype of the *MAOA* gene, using clump corrected for affection status for each gender. The association analysis of the dVNTR and uVNTR haplotypes is given in Tables 5.4 and 5.5. The haplotype distribution comprising two VNTRs showed no significant association to both markers of each gender.

Table 5.4: Clumping on genotypes uVNTR, dVNTR and haplotype. Including data from Tables 5.9, 5.10 and 5.11 versus affection status (AS) comparing schizophrenia and matched controls.

Affection status		T1	T2	T3	T4
dVNTR	males	0.995100	0.948705	0.979402	0.997400
	females	0.700930	0.857814	0.727327	0.808019
uVNTR	males	0.286371	0.112789	0.133287	0.245775
	females	0.759124	0.702630	0.707829	0.845915
Haplotype	males	0.691631	0.880112	0.833417	0.731927
	females	0.744526	0.876412	0.547545	0.645435

In Table 5.5, the suicide attempts trait was compared to both VNTRs' genotypes and the haplotypes, including both genders. Still, a slight trend was observed in the uVNTR genotype in males, yet not enough to define it as a risk biomarker.

Table 5.5: Clumping on genotypes uVNTR and dVNTR and haplotype versus suicide attempt comparing schizophrenia and matched controls. The analysis included data from Tables 5.1, 5.2 and 5.6.

Suicide Attempt		T1	T2	T3	T4
dVNTR	males	0.113689	0.288171	0.284872	0.468953
	females	0.921908	0.550545	0.743426	0.815718
uVNTR	males	**0.067493**	**0.066393**	**0.054695**	**0.054695**
	females	0.985001	0.898310	0.935306	0.926307
Haplotype	males	0.164284	0.214779	0.152185	0.326767
	females	0.962004	0.896310	0.866313	0.900410

Fisher's exact test was performed, Table 5.6, due to be more accurate than the chi-square and can be employed when the sample size is small. The samples were grouped by the allele activity to the uMAOA, following as parameter activity as 3R and 5R low activity and 3.5R and 4R high activity [28]. The data show that for males in both activity groups, defined as low_low or high_high, the variant gene expression presented a minor trend towards the risk of this trait accordingly. Further analysis applying logistic regression for males of activity group low-low with the uVNTR (3R) has been statistically significantly associated (p 0.036) with the suicide attempt risk.

Association analysis involving risk factors of a suicide attempt versus allele activity of MAOA uVNTR in Table 5.6, categorised as low activity for alleles 3R and 5R and high activity for alleles 3.5R and 4R, including females and males, was not statistically significant. Fisher's exact test for allelic activity based on the MAOA uVNTR was grouped based on the low and high activity allele, in female homozygous p = 0.4723 and heterozygous p = 0.6717; and males p = 0.1362, as shown in Table 5.6.

However, the logistic regression for males with the activity group low for the uVNTR (3R) has been statistically significantly associated (p 0.036) with the suicide attempt risk.

Table 5.6: Fisher's exact test to allele activity of MAOA uVNTR. Data was grouped with the following parameters: low activity for alleles 3R and 5R and high activity for alleles 3.5R and 4R. Where the analysis of low_high activity was only performed on females

		high high		low_low		low_high		Fisher's exact
		N	%	N	%	N	%	p-value
Males	SZ	95	0.56	74	0.44			0.1362
	Control	131	0.64	73	0.36			
females homozygous	SZ	79	0.73	29	0.27			0.4723
	Control	51	0.78	14	0.22			
females heterozygous	SZ	79	0.38	29	0.14	100	0.48	0.6717
	Control	51	0.43	14	0.12	54	0.45	

5.4 Discussion Chapter 5

Here we characterised the genetic variation with MAOA VNTRs in SCZ patients and compared them to findings in controls.

SCZ is a chronic and severe mental disorder affecting more than 21 million people worldwide and affecting the process of thinking, feeling, and behaving [206]. Schizophrenia is associated with several physical comorbidities, which can be responsible for 60% of premature deaths. Because of this, it is known as a "life-shortening disease" [267], which can abbreviate life expectancy between one to two decades, as those comorbidities more than double the chances of death compared to the general population [268, 269]. The early onset and the chronic aspect of schizophrenia make the disorder of relatively high prevalence and contribute to being amongst the leading causes of disabilities. Schizophrenia is among the top 25 global social and economic burdens, representing around 2% of health costs [270]. Another social impairment for schizophrenia patients is the persistence of symptoms, which can cause them to be unsuitable to work [271]. Also, early onset can cause educational disruption, difficulty in starting a career and lack of employment; all those together can cause a poor quality of life for schizophrenia patients [272].

The characteristic age of onset to the first episode of schizophrenia is during the late years of adolescence or early years of adulthood. It is infrequent in childhood (<13 years); some studies suggest that it might be an essential factor in the severity of the disease. The peak age at onset is 20 to 29 years for both genders, whereas females have a second peak at 30-39 years [2, 273] [274]. Several studies support the neurodevelopment hypothesis of schizophrenia, where critical changes during brain development caused by early life stress events, such as adversities in childhood, violence, and a history of childhood trauma, indicate that neurodevelopment is a target phase for the risk of schizophrenia [275-277]. The importance of the environment to epigenetic changes implicates the interaction between biological vulnerability and interaction versus environmental risks, also known as gene × environment. This

complex interaction may induce epigenetic changes, which can be an added risk to schizophrenia [278].

Several studies have concluded no association or consensus between MAOA uVNTR polymorphism and suicide attempts [236, 279-282]. Studies have also reported that the high-activity allele 4R is associated with a suicide attempt. The 4R of MAOA could be associated with traits orienting suicidal behaviour towards a violent act or suicide attempt by violent means. [283] Lung et al. concluded that the 4R allele of MAOA is associated with enhanced vulnerability to suicide in depressed males, as this can result from the 4R allele affecting vulnerability to suicide through the mediating factor of depressive symptoms [236]. They also concluded that there is an association between MAOA promoter VNTR allele 4R and suicidality in bipolar patients, particularly in females. Huang et al. observed that separating the sample into those with and without childhood abuse revealed an exploratory finding that females with the low activity allele 3R and a history of early abuse had an association with suicide attempt behaviour as adults [281]. Alves et al. found in a meta-analysis review that the L allele of 5HTTLPR increased the probability of suicide behaviour to several mental illnesses, with the strongest correlation to suicide being bipolar disorder, major depressive disorder, and schizophrenia. Also, an association has been reported between genetic variation in MAOA and anger behaviours in suicidal males [282].

However, our results indicated that in our population, the 3R had been associated with attempting suicide in males.

It is essential to observe that heterozygous females were included in the analysis. also, each gender was also investigated separately.

Our regression analysis did not find any positive association between the variants of the uVNTR to the risk of a suicide attempt; regardless of the results demonstrating a trend in the males, the literature is quite contradictory in defining which allele provides the increase to the risk of suicidal behaviour, as it has been linked to the high allele

in males with depression, and also in females with bipolar disorder [236]. Huang et al., in a study investigating the history of early abuse, found that the 3R was the risk allele to females [281].

The importance of these variants in disease predisposition may only indicate the participation of MAOA. It has a prominent role in GxE; also, the protein of MAOA operates in a fundamental neurologic process of the balance of neurotransmitters, which requires a constant response as it is dynamically challenged.

5.5 Conclusion chapter 5

It was previously reported that uVNTR (low-expression alleles) is associated with schizophrenia. In this work, we observed a significant association between the low-expression allele and schizophrenia. Tanifuji et al. have also found that the haplotypes containing dVNTR(9R) may enhance the role of MAOA in the pathophysiology of schizophrenia, which is a critical observation of the combined effects of several variants on the risk of schizophrenia; this approach can elucidate the contribution of MAOA VNTRs, also help to raise our current knowledge of the molecular mechanism underlying schizophrenia.

Therefore, our findings warrant reconsideration of previous studies while also considering the effects of VNTRs, haplotypes, and sex differences.

Our study has several limitations. First, our sample size was relatively small, and our cohort comprised participants of European descent. Further studies with large sample sizes and different populations must validate and clarify the association between MAOA VNTRs and mental health outcomes. Second, it would be necessary to consider the longitudinal effects, with a comprehensive history of the symptoms, such as whether negative or positive symptoms were dominant, age of onset, family history and environmental data.

Chapter 6

Thesis Summary

6. Thesis Summary

6.1.1 - Background

Despite the advances in human genomics research in the past decades, genetic risk factors within non-coding regions remain unknown. Much of this is due to our disproportionate knowledge of the well-characterised coding versus non-coding genome, formerly known as junk DNA.

Mutations of coding sequences have the potential to cause both qualitative and quantitative alterations to their encoded proteins, which automatically put them as the primary investigative source of heritable diseases, meaning that the functional role of the ncDNA is often overlooked. The ncDNA contains sequences that act as regulatory domains, predicted to function as transcription factor binding sites (TFBS) to assemble the complex DNA-RNA machinery, in which transcription factors (TFs) can account for either upregulation-downregulation, tissue-specific expression, environmental-induced activation-repression, or other complex forms of modulation.

In addition to quantitatively regulating gene expression, polymorphisms in ncDNA can also impact gene structure, conformational changes in preRNA maturation, retention of introns during RNA splicing (exonification) and translation of truncated proteins. In this context, repetitive elements and mobile retrotransposon elements displaying length variabilities, such as VNTRs and SINEs, are of particular interest due to their abundance, location and pleiotropic nature. Add to the fact that some of these elements are located within coding regions or UTRs of several genes.

Despite their functional role, there is an apparent difficulty in pinpointing variable ncDNA regions that can play a differential role in health and disease due to their poor characterisation and lack of standardised procedures, coupled with a high abundance of redundant elements. Some established regions resulted from previous investigative studies with syndromes and monogenic diseases or chance characterisations from traditional genetic techniques, not the integrative result of concerted mechanistic

approaches. This work builds on continuous functional work carried out in the field by our neurobiology group and represents a research effort to generate evidence and bridge knowledge gaps towards a clear understanding of polymorphic repetitive elements, particularly within ncDNA.

6.1.2 SINEs as regulatory drivers of the TOMM40-APOE-APOC1 gene cluster

We have shown that a primate-specific retrotransposon domain encompassing the genetic variant ('523') within the TOMM40 gene can act as a differential regulator of gene expression based on the genotype of the T repeat and has a significant influence on crystallised intelligence and decline in non-pathological ageing. Specifically, we described that the short poly-T variant was associated with reduced reporter gene expression, decreased vocabulary ability and a slower rate of decline. The association between the TOMM40-APOE locus and cognitive impairment have been together with the most extensively reported findings in complex genetic diseases/traits. The functional relevance of these genes in the aetiology of cognitive impairment adds weight to their involvement. TOMM40 encodes a translocase of outer mitochondrial membrane 40, a subunit of a larger TOM complex accountable for translocating newly synthesised unfolded proteins into the mitochondrial intermembrane space. [115] Mitochondria are vital for cell survival, and their dysfunction has been linked to pathological and non-pathological cognitive ageing. [218] Until recently, the focus of cognitive/dementia genetic research has significantly relied upon SNP data with little emphasis on other variant types. Previous association studies have demonstrated the function of non-coding DNA associated with SNPs and advocated for their ability to exert regulatory properties at the locus [67, 207]. The properties of the multiple Alu elements are awakening at this locus, which constitutes a human-specific domain with limited homology to other primates. Retrotransposon domains in non-coding genome regions can establish new splice sites, polyadenylation signals, promoters and

regulatory domains that can reorganise gene expression and build new transcription modules [219]. The location of retrotransposons within introns, promoters, untranslated regions and vicinities further enhances their ability to alter the local architecture and modulate genome regulation [220]. We have previously demonstrated that primate-specific SVA retrotransposons in the promoter of the neurodegenerative candidate genes, PARK7 and FUS, comparably to our present study, are both polymorphic and capable of modulating reporter gene expression [131, 221]. Data from our group [220] indicate that the SINEs block or the FLAM_A domain are plausible candidates to mechanistically modulate vocabulary ability and decline by accounting for differential TOMM40 gene expression, perhaps in response to an additional environmental challenge. However, these regions could also be hotspots for retrotransposition events, affecting the locus' function by multiple mechanisms.

Except for early embryo development and some malignancies, retrotransposition is suppressed by epigenetic and post-transcriptional modifications in somatic cells. However, their mobilisation has also been described in the adult brain. Baillie and colleagues have detected somatic L1 retrotransposition of three active retrotransposon families: LINE-1, Alu and SVA, which resulted in an alteration of the genetic landscape of the human brain [222]. Consistent with CNS mobilisation, sequencing-based studies have shown significant bias for insertion sites of somatic L1 retrotransposition in neurons of people with schizophrenia and that these events could have an aetiological role [223]. More recently, dysregulated retrotransposon function has been observed in frontal FTLD, which the authors linked with the activity of the neurodegenerative risk gene TDP-43 [224]. Furthermore, many studies, including ours, have indicated ageing as an important parameter affecting human memory and cognition. This data is consistent with the observations in the Drosophila mode, where the increased activity of transposable elements in the brain contributed to an age-dependent loss of neuronal function [224].

The association between the TOMM40 poly-T ‘523’ variant and cognition has been reported by several groups [211]. Our results showed a faster rate of cognitive decline in volunteers who carried the VL variant, which was marginally significant for vocabulary ability and exhibited a non-significant trend for memory. In contrast to other cognitive studies using neurologically healthy participants, the S variant in our volunteers was associated with reduced vocabulary ability ($p < 0.007$) but not memory ($p = 0.11$), executive function ($p = 0.14$) or processing speed ($p = 0.24$). Our observed differences in neurologically healthy individuals may be related to population stratification. The Greenbaum study investigated TOMM40 in an Israeli Jewish population, and the Johnson study did not mention their population ethnicity. However, the Hayden study was 99% Caucasian, but their results failed to reach significance after correction for multiple testing. Statistical power is unlikely to be an issue in our analysis as our cohort size of 1,613 was more significant than the other studies on cognitively healthy individuals combined. When analysing the subgroup of homozygous SS versus homozygous VL individuals, our sample size of 583 volunteers had 93% power to detect a genetic effect size of 2%, assuming a minor allele frequency of 0.45 (frequency of the S allele) and an additive genotypic model of effect.

The cognitive tests used by the different groups also varied, and none used a representative crystallised intelligence test comparable to the one used in this study. Finally, in our study, we used principal components analysis to derive the four factors that represent the four chief domains of cognition and, theoretically, should have been able to detect an association with any other cognitive test [130].

The strong linkage disequilibrium between TOMM40 and APOE genes [226] and their coordinative expression, as well as potential shared regulatory domains, suggests the existence of complex genome interactions with concerted regulatory mechanisms that ultimately determine the level of expression of these clustered genes. Retrotransposons can modulate gene expression via various processes. The

occurrence of a primate-specific mobile DNA element within the TOMM40-APOE, which may well extend towards neighbouring gene-dense loci, supports its regulatory nature while emphasising the involvement of this locus with higher intellectual traits and, consequently, cognitive dysfunction. Conversely, it is noteworthy that poly-T '523' polymorphism has been challenging to genotype and simplified categorisation into three groups, potentially hindering an adequate assessment of an otherwise multi-allelic system. It is further compounded by the fact that the simple analysis of consensus sequences we generated for the locus clearly indicates the existence of other polymorphisms (SNPs), which were specific for only one of the poly-T variant allele (T40/FLAM41, or simply 41T) and therefore likely to be inherited together.

Nevertheless, this work demonstrates the ability of distinct allele variants within an intronic SINEs block to influence expression within a luciferase reporter gene construct independently associated with cognitive measurements in a non-pathological community-dwelling elderly cohort. Further mechanistic studies are warranted despite these findings since the precise biological basis for these claims and clarification for the inconsistencies found across studies have yet to be established.

6.1.3 – The potential functional role of the SVA D downstream to APOC1 in regulating gene expression

The clustered genes TOMM40-APOE-APOC locus is located in a gene-rich region with significant genetic variability. The localisation of the retrotransposons within introns and the local environment is likely to contribute to their ability to alter the local architecture and, consequently, modulate genome regulation. SVA D, located downstream from the APOC1 gene, has been proposed to form secondary structures such as DNA quadruplexes [8], whilst neighbouring inversely arranged Alus (in the opposite strand) can induce the formation of stem-loops and palindromes. These constructed structures can influence RNA interference, transcription initiation, and

DNA replication and introduce alternative splicings [19]. Moreover, SVAs are GC-rich, which can introduce potential methylation sites and lead to the formation of abnormally methylated regions and alterations in the epigenetic profile of the locus [8]. Retrotransposons can also exert an influence on the genome via the action of retrotransposition. Although retrotransposition is suppressed by epigenetic and post-transcriptional modifications in somatic cells, except for early embryo development and some malignancies, the mobilisation has been reported in the brain. Baille et al. has recently detected somatic L1 retrotransposition of three active retrotransposon families: LINE-1, SINE/Alu and SVA, into protein-coding genes in the brain, which resulted in modifications of the genetic landscape of the human brain [20]. SINE-VNTR-Alus (SVAs) are the youngest and one of the minor groups of retrotransposons though they are specific to hominids and can display significant polymorphism, with nearly 2,700 of these elements represented in the human reference genome (hg19) [3]. The SVAs subfamilies are divided from SVA A to F1, with the SVAs subfamilies A, B, and C found across multiple primate species.

In contrast, a complete SVA D segment and subfamilies E, F and F1 are considered humans-specific, with many likely to carry polymorphic repeat motifs. These elements are, therefore, of renewed interest for genetic and functional genomic studies. Indeed, the most recent SINEs could be underlying the most distinctive evolutionary signatures that make us humans and differentiate us from other primates.

Human-specific retrotransposon insertions are supposed to be one of the two fundamental driving forces in the evolution of human-specific regulatory arrangements [4], followed by the recombination-based exaptation of highly conserved fragments of ancestral regulatory DNA. This evolution was complex, involving gene regulatory mechanisms in higher primates and humans and quite possibly partly impacted species- and tissue-specific gene regulation and allowed greater diversity comprising epigenetic modulation and response to environmental changes [5]. Bennett et al. found that, on average, a given individual can have

concealed an estimated 56 polymorphic SVAs within their genome [6], which would likely contribute significantly to an individual's genetic variation. Bennett et al. also showed that 79% of human SVAs are missing from their corresponding genomic sites in the chimpanzee genome. It indicates that SVAs have played a recent evolutionary task after the split between humans and chimpanzee species [6–9]. Many transposable element families are established to be distributed non-randomly throughout the human genome. That was emphasised in the early draft of the human genome, in which portions of the X-chromosome were observed to contain exceptionally high concentrations of the LINE-1 retrotransposon sequences, similar to chromosome 19, for which an increased concentration of Alu elements has been observed [10]. The latter was predicted due to the chromosome's high gene and GC content. It has been shown that SVAs are, in fact, not distributed at random across the genome; instead, their distribution displays a preference for a position with high GC content [3], which tend to harbour genic regions, with 60% of SVA elements either enclosed by genes or sitting within 10 kb upstream [11]. Wang et al. and Tang et al. emphasize that chromosome 19 has a genome unusually rich in SVAs, compared to other chromosomes [3,12], with Grim Wood et al. explaining that chromosome 19 has a higher proportion of transposable element-derived sequences compared to the remaining genome, 55% versus 44.8% [13]. SVA elements are over-represented in the established Parkinson's disease gene loci (PARK gene cluster) [14] and have been frequent at gene loci that regulate central nervous system (CNS) pathways [15]. Previous studies have verified the capability of SVA elements to control gene expression in human neuroblastoma cell lines (*in vitro*) and chick embryo models (*in vivo*) [11,16–18], which would reinforce that SVAs have the potential to regulate gene expression, either through their regulatory characteristics or by their location within the genome.

Indeed, Kim and Hahn have underlined the ability of SVA insertions to act out as novel promoters at their location of integration, potentially driving the expression of novel

human-specific transcripts [7,8]. Similar observations were given by Kwon et al., who hypothesised that the addition of SVA sequences occurred in numerous human transcripts [9]. Recent work by Tang et al. showed that human-specific retrotransposon insertions had incorporated an additional 14.2 Mbp to the human genome, which by extrapolation corresponded to an increase of 84 kb of expressed human-specific transcripts [12].

Furthermore, it has been observed that at least 12.5% of the human-specific SVAs overlapped with ENCODE ChIP-seq data. In practical terms, this would account for the creation of an additional 504 TFBS [12].

Polymorphisms can arise from several components of its structure, but it is often found, either in its hexamer repeat (CCCTCT)n, within its VNTR core sequence, or in its poly(A) region [11]. SVA insertions are associated as a reported cause of at least 13 diseases [70], for instance, the X-linked dystonia-parkinsonism (XDP) caused by insertion, wheresoever an SVA-F element, located in the intron 32 of the TATA-box binding protein associated factor 1 (TAF1) gene. [225, 229-231]. Intriguingly, this SVA insertion was not only observed to be associated with reduced expression of TAF1, but also with alternative splicing and intron retention, leading to a defective gene. The length of its CT hexamer repeat was inversely correlated with the age of onset of the disease [232]. Likewise, polymorphic SVA insertions have been demonstrated to regulate gene expression or isoform expression via intron retention in a population-specific manner [229, 233]. This emphasizes their competence to have an impact at several levels of genetic processing and contribute to phenotypic differences within various diseases, including additional neurodegenerative conditions, such as Alzheimer's, Parkinson's and Amyotrophic Lateral Sclerosis.

Retrotransposons can modulate gene expression via various mechanisms. The presence of these primate-specific, mobile DNA elements within the TOMM40-ApoE-ApoC1 gene locus and their potential regulatory nature supports the association of

this gene-dense cluster with higher cognitive traits and, subsequently, cognitive dysfunction.

We demonstrated the ability of different constructs to interfere with transcription within a luciferase reporter gene construct. Future work could potentially investigate their ability to modulate reporter gene expression directed by an extended TOMM40/ApoE-specific construct containing regulatory elements, as well as a more detailed investigation of the SVA D, whether or not coupled with the nearby *APOC1*. Indeed, haplotype studies indicated the existence of strong linkage disequilibrium in this intergenic region, which extends towards both, the 5' direction to the TOMM40-APOE gene locus and the 3' direction to a cluster of Apolipoprotein genes, which emphasises the need for additional studies.

6.1.4 Allelic-specific regulation of uVNTR and dVNTR MAOA expression

The MAOA uVNTR is well discussed in the literature regarding the G × E interactions. The recognised theory is that the different repeat copy numbers of the uVNTR charge the different expressions of uVNTR in response to the environment. It is a reasonably simple justification for the male gender, which possesses only one allele for MAOA. Nevertheless, the situation is more complex in females for various reasons, including the gene dosage, the potential for escape X-inactivation, and the heterozygosity of the uVNTR, whose genotype may direct distinct MAOA gene expression patterns.

Here investigated the molecular processes underlying MAOA expression in the SH-SY5Y, a human neuroblastoma cell line heterozygous for the uVNTR; the heterozygosity of the uVNTR is required to permit us to relate mRNA expression from each allele of uVNTR MAOA. Followed by the epigenetic marks over the uVNTR and transcription factor binding to the locus.

The finds revealed activity in both alleles of uVNTR using the SH-SY5Y cells, although by significantly distinct molecular mechanisms established on the methylation marks and binding of proteins over this region. Most outstandingly, even though both alleles

had very diverse chromatin structures (methylation and transcription factor interactions) when ChIP was executed for an active RNA polymerase, binding was detected at both alleles, which was consistent with the cDNA studies showing that mutual alleles were active, despite the distinct chromatin architecture, Figure 4.4A.

The methylation of a single allele which did not repress MAOA expression, Figure 4.3, would be consistent with current work on X chromosome gene expression profiles, which showed that methylation did not essentially correlate with gene repression but somewhat operated as a parameter disturbing a region's capacity to escape X inactivation [163].

To further explore the regulation of MAOA gene expression by the uVNTR, we examined the binding of five transcription factors: CTCF, SP1, hnRNP K, CNBP and nucleolin. Our data delivered indication that these proteins recognised the proximal MAOA promoter surrounding the uVNTR, and more significantly, a distinct pattern of binding was noticed over each allele which was modulated both by the genotype of the uVNTR, e confirmed with the exposure of the cells to sodium valproate, Figure 4.4C.

The importance of gender in the regulation of MAOA expression through epigenetic mechanisms has been demonstrated by Wong et al., where the gender-specific alterations of methylation in a longitudinal study of twins. [250] Domschke et al. similarly showed hypomethylation in the pathogenesis of the panic disorder, particularly in females. [251] Both studies were consistent with our finds.

Our data could enlighten the modifications in uVNTR genotype association for definite mental health issues when comparing males and females. Such examples include MAOA uVNTR moderating the association between childhood mistreatment and several disorders, as an example the dysthymia only present in females, [168] gender alterations comprising noticeable effects on anxiety and Attention-deficit hyperactivity disorder (ADHD) in females, however, to bipolar disorder, autism and aggressive behaviour the males are perceptible more affected.[139, 154] Indeed, many MAOA G

× E studies report only on males due to compounding effects, including female heterozygosity for the uVNTR. [30, 135, 165] MAOA gene is one of the roughly 15% of X chromosome genes predictable to escape X inactivation, and that alone can further complicate the regulation.[163, 164, 238]

We still have a considerable discussion regarding X inactivation in vivo, as the X inactivation of the second allele in heterozygotes may not depend on G x E interactions, as well as the gene called a master regulator of X chromosome inactivation, the gene X Inactive Specific Transcript (XIST), is overexpressed in women with major affective disorders, which is consistent with this hypothesis. [162] Thus, gender- and disorder-based differences, especially in mental health, are likely, and this study highlights a potential mechanism to explain these differences.

6.1.5 The role of uVNTR and dVNTR in the regulation of MAOA gene expression and how it relates to mental health

The focus of MAOA regulation has always been attributed exclusively to the uVNTR. Several studies and publications consider the only to account for the variations in the expression of MAOA and the only to address the possibility of finding answers in neuropsychiatric conditions, behaviour, aggressiveness, and others. Here we discuss the contribution of dVNTR to MAOA gene expressions and the import synergism between the two markers. However, our data from the double knockout VNTRs did not prevent gene expression. Thus, it was able to provide significantly lower levels of mRNA transcripts. The primary transcript isoform seemed to have the dVNTR as a major regulator. On the other hand, the minor transcript isoform 204, Figure 4.13, had a negatively regulated expression by both VNTRs.

Our data was generated using a cell line able to express both isoforms of MAOA, HAP1 cells, and it is a haploid karyotype, which enabled successful CRISPR deletion of regulatory elements. This model supports combining both VNTRs in MAOA

regulation, considering that the expression of MAOA can be differentiated depending on the inducible stimulus manner, tissue-specific or regulation in a neuronal context. The conservations of both MAOA VNTRs, show that they are primate-specific, which proposes that there are further critical regulatory domains for MAOA expression in mammals. Additionally, work from Warburton et al., 2016 [252], demonstrated a dual function attribute to a VNTR in the mir137 gene, depending on the expressed mRNA isoform. Similarly, our data also indicate that the uVNTR could have a dual function in transcription and post-transcriptional regulation of the MAOA gene as it is contained within the 5'UTR of one of the putative coding isoforms.

In conclusion, these data provide important considerations for understanding the regulation of MAOA expression and how it can act in the modulation by the VNTRS. Our work has also shown that VNTRs can exhibit isoform-specific modes of regulation, and multiple VNTRs can have an additive effect on gene expression profiles. They assessed two VNTRs (dVNTR and uVNTR) found upstream of the monoamine oxidase A (MAOA) gene and discovered that a single knockout (KO) of the dVNTR and a double KO of the dVNTR and uVNTR both led to a significant decrease in total MAOA expression. They also measured the expression of a minor alternative isoform of MAOA. They found that single KOs of each VNTR led to a significant increase in the expression of this isoform. A double KO induced a further increase in expression, highlighting isoform-specific regulatory effects and an additive effect of the two VNTRs at this locus [20].

6.1.6 Characterization of haplotypes of monoaminergic VNTRs and inference of composite 'risk' alleles associated with behaviour conditions

The interplay between the environment and genetic variations of VNTRs related to genes within the monoaminergic pathway is compatible with molecular biology discoveries that gene expression activation depends on transcriptional signals originating from internal and exterior environments. [284]

Nevertheless, often we found inconsistency in the results associating these mental health conditions to a specific genotype due to the complexity of these behavioural or mental disorders based on the outcome of a sole polymorphism analysis, resulting in the unreliable G x E results may be in part supported by analyses that include only a single genomic variant. Therefore, the solution could lie in the use of a combined polymorphism analysis. Here, we consider approaches that hold several polymorphisms accountable can contribute to a more reliable analysis in light of the complexity involved in psychiatric and behavioural disorders [256, 285]. There is a substantial amount of evidence that the uVNTR and dVNTRS of MAOA, 5-HTTLPR, STin2, and DRD4 can aid in elucidating the predisposition to many neurological disorders, as well as the G x E that make these biomarkers an excellent candidate to be considered either a risk or a protective influence on neurological disturbances [134, 256, 261]. Compared to an isolated analysis, the combined force of a combined genetic profile may have a more significant effect. This method could also tell if any crucial changes happen at different neurodevelopment stages [84, 87, 286]. Our genotype results found in our population present similar distributions to the literature. Only the DRD4 exon 3 VNTRs, was evaluated by Holmboe et al., merging the alleles 2R and 7R, as it was coded as 4R vs 2R/7R. However, we performed separate analyses to verify the inclusion and exclusion of allele 7R, less frequent but without significative changes to the results, and we opted to use the merged alleles [287, 288]. We then tried to apply the linear regression, as suggested by Zhang et al. [134]. However, we cannot perform the analysis due to the lack of clinical data, as no association could be proven (Data shown in Appendix). The analysis conformed to the Hardy-Weinberg equilibrium regarding the heterozygous distribution in the genotypes of all five VNTRs with the WCHADS samples.

To validate and use the combined VNTR profile in an attempt to investigate if the haplotypes and genotypes combined could support any risk or protective factor for

mental health [180, 289], will be necessary further analyses, including cohort with both, cases versus health matched controls, and robust numbers.

As the VNTRs in question are well recognised linked to many neurodevelopmental disorders and conditions, such as schizophrenia and aggressive behaviour [144, 176, 188, 197, 288, 290-292], and we need to take into consideration the G x E, which are well established with these genes linked to essential neurotransmitters, which are responsible for keeping the balance of the CNS. Schizophrenia considered a neurodevelopmental disorder, is an example of a disorder that significantly impacts the patient, and an early diagnostic or early approach to identify the risks could reduce the negative consequences. In the second part of chapter 5, we characterised the genetic variation with MAOA VNTRs in SCZ patients and compared it to finds in controls.

Schizophrenia is a chronic and severe mental disorder affecting more than 21 million people worldwide and affecting thinking, feeling, and behaving [206]. Schizophrenia is associated with several physical comorbidities, which can be responsible for 60% of premature deaths. Because of this, it is known as a "life-shortening disease" [267], which can abbreviate life expectancy between one to two decades, as those comorbidities more than double the chances of death, compared to the general population [268, 269]. The early onset and the chronic aspect of schizophrenia make the disorder relatively high prevalence and contribute to being among the leading causes of disabilities. It is among the top 25 global social and economic burdens, representing around 2% of health costs [270]. Another social impairment for schizophrenia patients is the persistence of symptoms, which can cause them to be unsuitable to work [271]. Also, early onset can cause educational disruption, difficulty in starting a career and lack of employment; all those together can cause a poor quality of life for schizophrenia patients [272].

The characteristic age of onset to the first episode of schizophrenia is during the late years of adolescence or early years of adulthood. It is infrequent in childhood (<13

years); some studies suggest that it might be an essential factor in the severity of the disease. The peak age at onset is 20 to 29 years for both genders, whereas females have a second peak at 30-39 years [2, 273] [274]. Several studies support the neurodevelopment hypothesis of schizophrenia, where critical changes during brain development caused by early life stress events, such as adversities in childhood, violence, and a history of childhood trauma, indicate that neurodevelopment is a target phase for the risk of schizophrenia [275-277]. The importance of the environment to epigenetic changes implicates the interaction between biological vulnerability and interaction versus environmental risks, also known as gene × environment. This complex interaction may induce epigenetic changes, which can be an added risk to schizophrenia [278].

Several studies have concluded no association or consensus between MAOA uVNTR polymorphism and suicide attempts [236, 279-282]. Few studies have also reported that the high-activity allele 4R is associated with a suicide attempt. The 4R of MAOA could be associated with traits orienting suicidal behaviour towards a violent act or suicide attempt by violent means [283]. Lung et al. concluded that the 4R allele of MAOA is associated with enhanced vulnerability to suicide in depressed males. That can be a result of the 4R allele affecting vulnerability to suicide through the mediating factor of depressive symptoms [236], also concluded that there is an association between MAOA promoter VNTR allele 4R and suicidality in bipolar patients, particularly in females. Huang et al. observed that separating the sample into those with and without childhood abuse revealed an exploratory finding that females with the low activity allele 3R and a history of early abuse had an association with suicide attempt behaviour as adults [281]. Alves et al., found in a meta-analysis review that the L allele of 5HTTLPR increased the probability of suicide behaviour to several mental illnesses, with the strongest correlation to suicide being bipolar disorder, major depressive disorder, and schizophrenia. Also has been reported an association between genetic variation in the MAOA and anger behaviours in suicidal males [282].

However, our results indicated that in our population, the 3R had been associated with attempting suicide in males; the evaluation of genotype to uVNTR and dVNTR was not statistically significant in the data analysed.

Association analysis involving risk factors of a suicide attempt versus allele activity of MAOA uVNTR, categorised as low (3R;5R) and high (3.5R; 4R) alleles, here, our data was unable to show any statistically significant association.

Our regression analysis did find an association between males' carrier of the low variant to the uVNTR to the risk of a suicide attempt; regardless of the results being statistically significant in males in our studies, the literature is quite contradictory in defining which allele provides the increase to the risk of suicidal behaviour, as it has been linked to the high allele in males with depression, and also females with of and bipolar disorder [236]. Huang et al., in a study investigating the history of early abuse, found that the 3R was the risk allele to females [281].

The importance of these variants in disease predisposition may only indicate the participation of MAOA. It has a prominent role in GxE; also, the protein of MAOA operates in a fundamental neurologic process of the balance of neurotransmitters, which requires a constant response as it is dynamically challenged.

Chapter 7

Appendix

7.1 Introduction

The human central nervous system (CNS) begins to develop at three weeks of gestation, and this will carry on throughout infancy until adulthood; the changes will inevitably continue during a lifetime as it adapts to environmental and other physiological and chemical processes.

This work aimed to investigate the VNTRs of MAOA gene and other monoaminergic genes that can contribute to a better understanding of these interactions. Monoaminergic neurotransmitters play a crucial role in maintaining this equilibrium. Also, we included the VNTRs of the Serotonin Transporter-Linked Polymorphic Region (5-HTTLPR) and Serotonin Transport Intro 2 (STin2), which are both situated on the Serotonin Transporter gene (SLC6A4); the VNTR of exon 3 from the Dopamine receptor D4 (DRD4); and MAOA gene polymorphism at uVNTR and dVNTR, were genotyped, and all theses markers have been associated in several studies related to behaviour and mental health.[266]

The aim was to evaluate the relevance of the VNTRs of MAOA, SLC6A4, DRD4 in clinical samples from the Wirral Child Health and Development Study (WCHADS).

A combined genotyping and haplotype analysis of these VNTRs was conducted to study the interaction.

According to the data, haplotypes are co-inherited with the MAOA VNTRS and have a certain tendency in the SLC6A4 markers. The findings from the combination of genotypes reveal how they interact. Compared with clinical outcomes, these variations may be better understood with further investigation into neuropsychiatric disorders.

7.2 Genotyping and Allele Distribution of the Monoaminergic VNTRs

The aim was to investigate the genetic profile of MAOA, SLC6A4 and DRD4 VNTRs, and potential multi-loci combinations of their 'risk' alleles since these genes have been associated with behaviour and other neurodevelopmental disorders. First, we Analyse the genotypes of VNTRs located in the monoaminergic genes MAOA, SLC6A4 and DRD4, followed by the analysis of the genotype results, including Hardy-Weinberg equilibrium and Linkage disequilibrium of haplotype to MAOA, SLC6A4 VNTRs. To understand the correlation between the monoaminergic VNTRs to mental health.

Using DNA extracted from saliva samples from the Wirral child health and development study (WCHADS) cohort, the DNA extraction and cohort methods are mentioned in sections 2.1.4.1 and 2.2.2, respectively. The repeats of the polymorphisms of MAOA uVNTR and dNVTR, the 5HTTLPR and STin2 of the SLC6A4 gene and the VNTR in the exon3 of DRD4 gene, detailed in Table 5.1, were investigated using primers described in table 2.1 and of the polymerase chain reaction (PCR) method described in section 2.2.15. The genotype results were double-blind checked, and results and contradictory results were excluded from the analysis. The differences in the number of samples results to each marker also resulted from unsuccessful amplification or poor-quality bands in the gels examined. The genotype results can be seen in Tables 5.2 to 5.6.

Table 7.1: VNTRs from the monoamine genes

VNTRs	Gene	Chr	Start	End	Repeat (bp)	Common allele
uVNTR	MAOA	X	43514243	43514566	30	**4R** vs. 3R
dVNTR	MAOA	X	43654464	43654572	10	**9R** vs. 10R
5HTTLPR	SLC6A6	17	30237061	30237479	44	**16R** (**L**) vs. 14R (S)
STin2	SLC6A6	17	30221314	30221701	16 or 17	**12R** vs.10R
DRD4ex3	DRD4	11	639989	640194	48	**4R** and 2R+7R

For the MAOA gene VNTRs, the genotype data found in Table 7.2, to the uVNTR given as a major allele 4R to both genders, with males and females having 56.7% and 42.1%. Respectively. In the sample of females, the number of 4R enhanced when considering the heterozygous genotypes. Opposite to the 3R variant, that is increased to males (39.4%). However, when the heterozygous females are included in the count, there is no difference between genders, resulting in 39.5%, as before, homozygous was only 13.2%. Our found corroborate other studies [169](sabol, Caspi uvntr).

Table 7.2: Genotype and allele frequencies of MAOA uVNTR in WCHADS cohort.

uVNTR	females/males		males		females	
	frequency	%	frequency	%	frequency	%
Genotypes						
3_3	56	25.7	41	39.4	15	13.2
3_3.5	1	0.5	0	0	1	0.9
3.5_3.5	1	0.5	1	1	0	0
3_4	45	20.6	0	0	45	39.5
3.5_4	1	0.5	0	0	1	0.9
3.5_5	1	0.5	0	0	1	0.9
4_4	107	49.1	59	56.7	48	42.1
4_5	3	1.4	0	0	3	2.6
5_5	3	1.4	3	2.9	0	0
N	**218**		**104**		**114**	
Alleles						
3R	158	36.2	41	39.4	76	33.3
3.5R	5	1.1	1	1	3	1.3
4R	263	60.3	59	56.7	145	63.6
5R	10	2.3	3	2.9	4	1.8
N	436		104		228	

In that order, the major allele of dVNTR, 9R compromises 71.6% and 57.3% of the genotypes, to males and females. In females, 27.1% of heterozygous have one 9R allele, and only 8.3% are 10R/10R, as hemizygous males presented 24.7% of the 10R allele. In the samples analysed, there is a preference for at least one allele 9R, the major allele, as observed in Table 7.3.

Table 7.3: Genotype and allele frequencies of MAOA dVNTR in WCHADS cohort.

dVNTR	males + females		males		females	
	Frequency	%	frequency	%	frequency	%
Genotypes						
9_9	113	63.8	58	71.6	**55**	57.3
9_10	26	14.7	0	0	26	27.1
9_11	5	2.8	0	0	5	5.2
10_10	28	15.8	20	24.7	8	8.3
10_11	2	1.1	0	0	2	2.1
11_11	3	1.7	3	3.7	0	0
N	**177**		**81**		**96**	
Alleles						
9R	257	72.6	58	71.6	141	73.4
10R	84	23.7	20	24.7	44	22.9
11R	13	3.7	3	3.7	7	3.6
N	354		81		**192**	

The SLC6A4 gene, which encodes for the serotonin transporter, contains a 43 bp variable-number tandem repeat polymorphism in the promoter region (5-HTTLPR), with alleles often known as Long (L) with 16 copies of the repeat and short (S) with 14R. The L/L genotype is increased in males (9.5%) and appears at least 6.1% higher in the alleles frequency than in females, Table 7.4. A second VNTR located in intro 2 of the SLC6A4 gen, with 9 to 12 repeats alleles, has also been analysed, showing that the heterozygous 10R-12R has been the most frequent, followed by the 12R and 10R, respectively. The 12R is the allele most frequent, with >60% of alleles frequency, to both gender.

Table 7.4: Genotype and allele frequencies of 5HTTLPR genotype in WCHADS cohort.

5HTTLPR	males + females		males		females	
	Frequency	%	Frequency	%	Frequency	%
Genotypes						
LL	61	28.6	36	33.3	**25**	23.8
SL	102	47.9	48	44.4	54	51.4
SS	50	23.5	24	22.2	26	24.8
N	**213**		**108**		**105**	
Alleles						
L	224	52.6	120	55.6	104	49.5
S	202	47.4	96	44.4	106	50.5
N	426		216		210	

Table 7.5: Genotype and alleles frequencies of STin2 in WCHADS cohort.

STin2	males + females		males		females	
	Frequency	%	Frequency	%	Frequency	%
Genotypes						
9_10	3	1.6	1	1.1	2	2
9_12	5	2.6	3	3.2	2	2
10_10	27	14	12	12.8	15	15.2
10_12	84	43.5	41	43.6	43	43.4
12_12	74	38.3	37	39.4	37	37.4
N	**193**		**94**		**99**	
Alleles						
9R	8	2.1	4	2.1	4	2
10R	141	36.5	66	35.1	75	37.9
12R	237	61.4	118	62.8	119	60.1
N	386		188		198	

In our study population, six different DRD4ex3 alleles with 2 to 8 tandem repeats, Table 7.6, and forming fifteen genotypes were present. The 4R and 7 repeat alleles were the most common, with 69.25% and 15% of the allele's frequency. We agreement controls found in other studies. Nevertheless, there is a lack of homozygous 7/7 and an increase in the allele 2R.

Table 7.6: DRD4 Exon3 genotype and alleles frequencies in WCHADS cohort.

DRD4	males + females		males		females	
Ex3	Frequency	%	Frequency	%	Frequency	%
Genotypes						
2_2	5	2.5	3	3.1	2	2
2_3	2	1	0	0	2	2
2_4	21	10.5	11	11.2	10	9.8
2_5	1	0.5	0	0	1	1
2_7	5	2.5	2	2	3	2.9
3_3	2	1	2	2	0	0
3_4	6	3	3	3.1	3	2.9
3_5	1	0.5	0	0	1	1
3_7	4	2	1	1	3	2.9
4_4	101	50.5	50	51	51	50
4_5	2	1	0	0	2	2
4_7	45	22.5	22	22.4	23	22.5
4_8	1	0.5	0	0	1	1
7_7	3	1.5	3	3.1	0	0
8_8	1	0.5	1	1	0	0
N	**200**		**98**		**102**	
Alleles						
2R	39	9.75	19	9.7	20	9.8
3R	17	4.25	8	4.1	9	4.4
4R	277	69.25	136	69.4	141	69.1
5R	4	1	0	0	4	2
7R	60	15	31	15.8	29	14.2
8R	3	0.75	2	1	1	0.5
N	400		196		204	

7.3.2 Statistical analysis

Following the genotypes results, to allow biallelic analysis, the data were categorised based on the common alleles, as required, Table 5.1. The MAOA uVNTRs were coded as 4R vs.3R and dVNTR as 9R vs. 10R. The 5-HTTLPR was coded 16R/L vs 14R/S and STin2 10R vs. 12R. [169]

Due to the multiple alleles and genotypes of DRD4, to examined was necessary to categorise the alleles based on the frequency of the alleles. For that, the 2R and 7R were merged into the same allele, and DRD4 was coded as 4R vs. 2R/7R, as proposed by Holmboe et al. [287, 288]

Haplotype frequencies for DRD4, STin2, MAOA (uVNTR; dVNTR) and 5HTTLPR, we were unable to perform the Linkage analysis (LD), as it only applies to polymorphisms within the same chromosome.

7.3.3 Distribution of VNTRs conformed to the Hardy-Weinberg equilibrium.

For the heterozygous distribution of the Hardy-Weinberg equilibrium, Table 7.7 is observed p-values of 0.29 to DRD4ex3, 0.37 to STin2, 0.36 and 0.22 to uVNTR and dVNTR, respectively and to the 5HTTLPR 0.52. In the number of heterozygous observed and predicted in our samples, we observed a variance of 0.01 to 0.03, and the dVNTR, this variance was twice compared to the range noted 0.06. It is probably a result of the high homozygous outcome, related to the allele 9R representing 57.3% of female samples, as here we are analysing only heterozygous only females included in the dVNTR and uVNTR. The distribution of heterozygous in VNTRs from the samples analysed conformed to the Hardy-Weinberg equilibrium.

Table 7.7: **Hardy-Weinberg equilibrium for the observed and predicted heterozygous frequency.**

VNTR	HET observed	HET predicted	HW p-value	MAF	X^2
DRD4ex3	0.38	0.41	0.2998	0.29	2.1403
STin2	0.46	0.47	0.8502	0.373	0.035
uVNTR	0.43	0.46	0.6708	0.358	0.0626
dVNTR	0.29	0.35	0.1746	0.224	0.8803
5HTTLPR	0.47	0.5	0.5153	0.467	0.3356

7.4 Discussion

The interplay between the environment and genetic variations of VNTRs related to genes within the monoaminergic pathway is compatible with molecular biology

discoveries that gene expression activation depends on transcriptional signals originating from both the internal and exterior environments. [284]

Nevertheless, often we found inconsistency in the results associating these mental health conditions to a specific genotype due to the complexity of these behavioural or mental disorders based on the outcome of a sole polymorphism analysis, resulting in the unreliable G x E results may be in part supported by analyses that include only a single genomic variant. Therefore, the solution could lie in the use of a combined polymorphism analysis. Here, we consider approaches that hold several polymorphisms accountable can contribute to a more reliable analysis in light of the complexity involved in psychiatric and behavioural disorders [256, 285]. There is a substantial amount of evidence that the uVNTR and dVNTRS of MAOA, 5-HTTLPR, STin2, and DRD4 can aid in elucidating the predisposition to many neurological disorders, as well as the G x E that make these biomarkers an excellent candidate to be considered either a risk or a protective influence on neurological disturbances [134, 256, 261]. Compared to an isolated analysis, the combined force of a combined genetic profile may have a more significant effect. This method could also evaluate changes at different neurodevelopment stages [84, 87, 286]. Our genotype results found in our population present similar distributions to the literature. Only the DRD4 exon 3 VNTRs, was evaluated as Holmboe et al., merging the alleles 2R and 7R, as it was coded as 4R vs. 2R/7R. However, we performed separate analyses to verify the inclusion and exclusion of allele 7R, less frequent but without significative changes to the results, and we opted to use the merged alleles [287, 288]. We then tried to apply the linear regression, as Zhang et al. [134] suggested. However, due to the lack of clinical data, we cannot perform the analysis, as no hypothesis could be generated. The analysis conformed to the Hardy-Weinberg equilibrium regarding the heterozygous distribution in the genotypes of all five VNTRs with the WCHADS samples. The LD analysis presented a positive LD to uVNTR and dVNTR, but the 5HTTLPR and STin2 have not been observed LD, just a trend, as there is a slight

preference between the two markers. However, the missing correlation with clinical data was prohibitive to carry out an extensive analysis investigating the biological context of the analysed samples, as it would be to determine if any trait could be potentially identified. Our data had several limitations, first is necessary to consider the sample size since the number of individuals for some genotypes was insufficient to provide a firm conclusion, as it requires supporting statistical analysis, avoiding inferring the results, and having clarity in the conclusion. Second was the access to the clinical profile, as it was required to determine if any findings were associated with risk or protective factors in a range of mental disorders, e.g., behaviour and other neurodevelopmental disorders intended to be investigated.

To validate and use the combined VNTR profile in an attempt to investigate if the haplotypes and genotypes combined could support any risk or protective factor for mental health [180, 289], will be necessary further analyses, including cohort with both, cases versus health matched controls, and robust numbers.

As the VNTRs in question are well recognised linked to many neurodevelopmental disorders and conditions, such as schizophrenia and aggressive behaviour [144, 176, 188, 197, 288, 290-292] and we need to take into consideration the G x E, which are well established with these genes linked to essential neurotransmitters, which are responsible for keeping the balance of the CNS. Schizophrenia considered a neurodevelopmental disorder, is an example of a disorder that significantly impacts the patient, and an early diagnostic or early approach to identify the risks could reduce the negative consequences.

7.5 Conclusion

Conclude, our data on the genotypes and haplotype analysis of the VNTRS of MAOA, SLC6A4 and DRD4 indicates a complete LD between the MAOA VNTRs, and the results of 9R and 4R haplotypes has been proven to be accountable, and In evidence

in the 4R preference to the 9R, this has been observed in the samples from the WCHADS and Schizophrenia cohort. The SLC6A4 markers only provided a partial LD. Genotypes agreed to HWE, and no differences in the genotype distribution were observed compared to the literature. Indeed our samples analysed were from a populational-based study, which includes both sets of individuals, with stratification to the risk of behavioural problems, based on the assessments since the 20 weeks of pregnancy, when the recruitment was realised. They have been assessed periodically since as part of a longitudinal study. Additional analysis of these VNTRs would include a cohort of cases and controls to be able to analyse if the combined analysis could provide any potential risk or protective factor in mental health.

Chapter 8

References

8. References

1. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* 2001.
2. Zhou, X., et al., *Non-coding variability at the APOE locus contributes to the Alzheimer's risk.* Nature Communications, 2019. **10**(1): p. 3310.
3. Perenthaler, E., et al., *Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development.* Frontiers in Cellular Neuroscience, 2019. **13**(352).
4. Mendizabal, I., et al., *Cell type-specific epigenetic links to schizophrenia risk in the brain.* Genome Biol, 2019. **20**(1): p. 135.
5. Anwar, S.L., W. Wulaningsih, and U. Lehmann, *Transposable Elements in Human Cancer: Causes and Consequences of Deregulation.* International Journal of Molecular Sciences, 2017. **18**(5): p. 974.
6. Billingsley, K.J., et al., *Regulatory characterisation of the schizophrenia-associated CACNA1C proximal promoter and the potential role for the transcription factor EZH2 in schizophrenia aetiology.* Schizophrenia Research, 2018. **199**: p. 168-175.
7. Marshall, J.N., et al., *Variable number tandem repeats - Their emerging role in sickness and health.* Exp Biol Med (Maywood), 2021. **246**(12): p. 1368-1376.
8. Treangen, T.J. and S.L. Salzberg, *Repetitive DNA and next-generation sequencing: computational challenges and solutions.* Nat Rev Genet, 2011. **13**(1): p. 36-46.
9. de Koning, A.P.J., et al., *Repetitive Elements May Comprise Over Two-Thirds of the Human Genome.* PLOS Genetics, 2011. **7**(12): p. e1002384.
10. Auton, A., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
11. Van der Auwera, S., et al., *Genome-wide gene-environment interaction in depression: A systematic evaluation of candidate genes: The childhood trauma working-group of PGC-MDD.* American journal of medical genetics. Part B, Neuropsychiatric genetics, 2018. **177**(1): p. 40-49.
12. Sullivan, P.F., et al., *Psychiatric Genomics: An Update and an Agenda.* American Journal of Psychiatry, 2018. **175**(1): p. 15-27.
13. Lee, H., et al., *Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs.* 2013.
14. Power, R.A., et al., *Genome-wide Association for Major Depression Through Age at Onset Stratification:Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium.* 2017.
15. Davidson, S., et al., *Analysis of the effects of depression associated polymorphisms on the activity of the BICC1 promoter in amygdala neurones.* The Pharmacogenomics Journal, 2016. **16**(4): p. 366-374.

16. Hing, B., et al., *A Polymorphism Associated with Depressive Disorders Differentially Regulates Brain Derived Neurotrophic Factor Promoter IV Activity.* Biological Psychiatry, 2012. **71**(7): p. 618-626.
17. Gianfrancesco, O., V.J. Bubb, and J.P. Quinn, *SVA retrotransposons as potential modulators of neuropeptide gene expression.* Neuropeptides, 2017. **64**: p. 3-7.
18. Gianfrancesco, O., et al., *Identification and Potential Regulatory Properties of Evolutionary Conserved Regions (ECRs) at the Schizophrenia-Associated MIR137 Locus.* Journal of Molecular Neuroscience, 2016. **60**(2): p. 239-247.
19. Kundaje, A., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**.
20. Manca, M., et al., *The Regulation of Monoamine Oxidase A Gene Expression by Distinct Variable Number Tandem Repeats.* Journal of Molecular Neuroscience, 2018. **64**(3): p. 459-470.
21. Vasiliou, S.A., et al., *The SLC6A4 VNTR genotype determines transcription factor binding and epigenetic variation of this gene in response to cocaine in vitro: Epigenetics of the SLC6A4 gene.* Addiction biology, 2012. **17**(1): p. 156-170.
22. Nagy, C., et al., *Astrocytic abnormalities and global DNA methylation patterns in depression and suicide.* Molecular psychiatry, 2015. **20**(3): p. 320-328.
23. Murgatroyd, C., et al., *Effects of prenatal and postnatal depression, and maternal stroking, at the glucocorticoid receptor gene.* Transl Psychiatry, 2015. **5**: p. e560.
24. Zhao, J., et al., *An integrative functional genomics framework for effective identification of novel regulatory variants in genome–phenome studies.* Genome Medicine, 2018. **10**(1): p. 7-7.
25. Bakhtiari, M., et al., *Targeted genotyping of variable number tandem repeats with adVNTR.* Genome research, 2018. **28**(11): p. 1709-1719.
26. Breen, G., et al., *Variable number tandem repeats as agents of functional regulation in the genome.* IEEE engineering in medicine and biology magazine, 2008. **27**(2): p. 103-108.
27. Castillo-Lizardo, M., G. Henneke, and E. Viguera, *Replication slippage of the thermophilic DNA polymerases B and D from the Euryarchaeota Pyrococcus abyssi.* Frontiers in Microbiology, 2014. **5**.
28. Hill, J., et al., *Evidence for interplay between genes and maternal stress in utero: monoamine oxidase A polymorphism moderates effects of life events during pregnancy on infant negative emotionality at 5 weeks.* Genes, Brain and Behavior, 2013. **12**(4): p. 388-396.
29. Haddley, K., et al., *Behavioural Genetics of the Serotonin Transporter.* Behavioral Neurogenetics, 2012. **12**: p. 503-535.
30. Pickles, A., et al., *Evidence for interplay between genes and parenting on infant temperament in the first year of life: monoamine oxidase A polymorphism moderates*

effects of maternal sensitivity on infant anger proneness. Journal of child psychology and psychiatry, 2013. **54**(12): p. 1308-1317.

31. Warburton, A., et al., *Characterization of a REST-Regulated Internal Promoter in the Schizophrenia Genome-Wide Associated Gene MIR137.* Schizophrenia Bulletin, 2015. **41**(3): p. 698-707.
32. Bonvicini, C., et al., *DRD4 48 bp multiallelic variants as age-population-specific biomarkers in attention-deficit/hyperactivity disorder.* Transl Psychiatry, 2020. **10**(1): p. 70.
33. De Roeck, A., et al., *An intronic VNTR affects splicing of ABCA7 and increases risk of Alzheimer's disease.* Acta neuropathologica, 2018. **135**(6): p. 827-837.
34. Bakhtiari, M., et al., *Variable number tandem repeats mediate the expression of proximal genes.* Nat Commun, 2021. **12**(1): p. 2075.
35. Lu, T.-Y. and M.J.P. Chaisson, *The motif composition of variable-number tandem repeats impacts gene expression.* bioRxiv, 2022: p. 2022.03.17.484784.
36. Jönsson, M.E., et al., *Transposable Elements: A Common Feature of Neurodevelopmental and Neurodegenerative Disorders.* Trends Genet, 2020. **36**(8): p. 610-623.
37. Bourque, G., et al., *Ten things you should know about transposable elements.* 2018.
38. Ravindran, S., *Barbara McClintock and the discovery of jumping genes.* Proceedings of the National Academy of Sciences, 2012. **109**(50): p. 20198-20199.
39. Elbarbary, R.A., B.A. Lucas, and L.E. Maquat, *Retrotransposons as regulators of gene expression.* Science (American Association for the Advancement of Science), 2016. **351**(6274): p. 679-679.
40. Kazazian, H.H., Jr. and J.V. Moran, *Mobile DNA in Health and Disease.* N Engl J Med, 2017. **377**(4): p. 361-370.
41. Platt, R.N., 2nd, M.W. Vandewege, and D.A. Ray, *Mammalian transposable elements and their impacts on genome evolution.* Chromosome Res, 2018. **26**(1-2): p. 25-43.
42. Savage, A.L., et al., *Retrotransposons in the development and progression of amyotrophic lateral sclerosis.* Journal of Neurology, Neurosurgery & Psychiatry, 2019. **90**(3): p. 284.
43. Ayarpadikannan, S. and H.-S. Kim, *The Impact of Transposable Elements in Genome Evolution and Genetic Instability and Their Implications in Various Diseases.* Genomics & Informatics, 2014. **12**(3): p. 98-104.
44. Grandi, N. and E. Tramontano, *Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses.* Front Immunol, 2018. **9**: p. 2039.
45. Douville, R.N. and A. Nath, *Human endogenous retroviruses and the nervous system.* Handb Clin Neurol, 2014. **123**: p. 465-85.
46. Dolei, A., et al., *Expression of HERV Genes as Possible Biomarker and Target in Neurodegenerative Diseases.* Int J Mol Sci, 2019. **20**(15).

47. Thompson, P.J., T.S. Macfarlan, and M.C. Lorincz, *Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire.* Mol Cell, 2016. **62**(5): p. 766-76.
48. Küry, P., et al., *Human Endogenous Retroviruses in Neurological Diseases.* Trends Mol Med, 2018. **24**(4): p. 379-394.
49. Brudek, T., et al., *B cells and monocytes from patients with active multiple sclerosis exhibit increased surface expression of both HERV-H Env and HERV-W Env, accompanied by increased seroreactivity.* Retrovirology, 2009. **6**(1): p. 104.
50. Perron, H., et al., *Molecular characteristics of Human Endogenous Retrovirus type-W in schizophrenia and bipolar disorder.* Transl Psychiatry, 2012. **2**(12): p. e201.
51. Douville, R., et al., *Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis.* Ann Neurol, 2011. **69**(1): p. 141-51.
52. Li, W., et al., *Human endogenous retrovirus-K contributes to motor neuron disease.* Sci Transl Med, 2015. **7**(307): p. 307ra153.
53. Grow, E.J., et al., *Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells.* Nature, 2015. **522**(7555): p. 221-225.
54. Baar, T., et al., *RNA transcription and degradation of Alu retrotransposons depends on sequence features and evolutionary history.* G3 Genes|Genomes|Genetics, 2022. **12**(5).
55. Gianfrancesco, O., et al. *The Role of SINE-VNTR-Alu (SVA) Retrotransposons in Shaping the Human Genome*. International journal of molecular sciences, 2019. **20**, E5977 DOI: 10.3390/ijms20235977.
56. Beck, C.R., et al., *LINE-1 Retrotransposition Activity in Human Genomes.* Cell, 2010. **141**(7): p. 1159-1170.
57. Payer, L.M. and K.H. Burns, *Transposable elements in human genetic disease.* 2019. **20**(12): p. 760-772.
58. Deininger, P., et al., *A comprehensive approach to expression of L1 loci.* Nucleic Acids Res, 2017. **45**(5): p. e31.
59. Brouha, B., et al., *Hot L1s account for the bulk of retrotransposition in the human population.* Proceedings of the National Academy of Sciences, 2003. **100**(9): p. 5280-5285.
60. Denli, Ahmet M., et al., *Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity.* Cell, 2015. **163**(3): p. 583-593.
61. Raiz, J., et al., *The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery.* Nucleic Acids Research, 2012. **40**(4): p. 1666-1683.
62. Deininger, P., *Alu elements: know the SINEs.* Genome Biology, 2011. **12**(12): p. 236.
63. Konkel, M.K., et al., *Sequence Analysis and Characterization of Active Human Alu Subfamilies Based on the 1000 Genomes Pilot Project.* Genome Biology and Evolution, 2015. **7**(9): p. 2608-2622.

64. Batzer, M.A., et al., *Standardized nomenclature for Alu repeats.* J Mol Evol, 1996. **42**(1): p. 3-6.
65. Batzer, M.A. and P.L. Deininger, *Alu repeats and human genomic diversity.* Nature Reviews Genetics, 2002. **3**(5): p. 370-379.
66. Ade, C., A.M. Roy-Engel, and P.L. Deininger, *Alu elements: an intrinsic source of human genome instability.* Curr Opin Virol, 2013. **3**(6): p. 639-45.
67. Larsen, P.A., et al., *The Alu neurodegeneration hypothesis: A primate-specific mechanism for neuronal transcription noise, mitochondrial dysfunction, and manifestation of neurodegenerative disease.* Alzheimer's & Dementia, 2017. **13**(7): p. 828-838.
68. Savage, A.L., et al., *Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns.* BMC Evolutionary Biology, 2013. **13**(1): p. 101-101.
69. Zabolotneva, A.A., et al., *Transcriptional regulation of human-specific SVAF1 retrotransposons by cis-regulatory MAST2 sequences.* Gene, 2012. **505**(1): p. 128-136.
70. Hancks, D.C. and H.H. Kazazian Jr, *Roles for retrotransposon insertions in human disease.* Mobile DNA, 2016. **7**: p. 1-28.
71. Hancks, D.C., et al., *Exon-trapping mediated by the human retrotransposon SVA.* Genome Res, 2009. **19**(11): p. 1983-91.
72. Hall, A., et al., *A SINE-VNTR-Alu in the LRIG2 Promoter Is Associated with Gene Expression at the Locus.* International journal of molecular sciences, 2020. **21**(22): p. 8486.
73. Ali, F.R., et al., *Combinatorial interaction between two human serotonin transporter gene variable number tandem repeats and their regulation by CTCF.* Journal of Neurochemistry, 2010. **112**(1): p. 296-306.
74. Roberts, J., et al., *Differential Regulation of the Serotonin Transporter Gene by Lithium Is Mediated by Transcription Factors, CCCTC Binding Protein and Y-Box Binding Protein 1, through the Polymorphic Intron 2 Variable Number Tandem Repeat.* The Journal of neuroscience, 2007. **27**(11): p. 2793-2801.
75. Guindalini, C., et al., *A Dopamine Transporter Gene Functional Variant Associated with Cocaine Abuse in a Brazilian Sample.* Proceedings of the National Academy of Sciences - PNAS, 2006. **103**(12): p. 4552-4557.
76. MacKenzie, A. and J. Quinn, *A Serotonin Transporter Gene Intron 2 Polymorphic Region, Correlated with Affective Disorders, Has Allele-Dependent Differential Enhancer-Like Properties in the Mouse Embryo.* Proceedings of the National Academy of Sciences - PNAS, 1999. **96**(26): p. 15251-15255.
77. Warburton, A., et al., *A GWAS SNP for Schizophrenia Is Linked to the Internal MIR137 Promoter and Supports Differential Allele-Specific Expression.* Schizophrenia Bulletin, 2016. **42**(4): p. 1003-1008.

78. Bilgin Sonay, T., et al., *Tandem repeat variation in human and great ape populations and its impact on gene expression divergence.* Genome research, 2015. **25**(11): p. 1591-1599.
79. Breen, G., et al., *Engineering in Genomics variable number tandem repeats as agents of functional regulation in the genome.* IEEE engineering in medicine and biology magazine, 2008. **27**(2).
80. Quilez, J., et al., *Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans.* Nucleic acids research, 2016. **44**(8): p. 3750-3762.
81. Sawaya, S., et al., *Microsatellite Tandem Repeats Are Abundant in Human Promoters and Are Associated with Regulatory Elements.* PLoS ONE, 2013. **8**(2): p. e54710-e54710.
82. Dolzhenko, E., et al., *Detection of long repeat expansions from PCR-free whole-genome sequence data.* 2017.
83. Ziats, M.N., L.P. Grosvenor, and O.M. Rennert, *Functional genomics of human brain development and implications for autism spectrum disorders.* Translational Psychiatry, 2015. **5**(10): p. e665-e665.
84. Owen, M.J. and M.C. O'Donovan, *Schizophrenia and the neurodevelopmental continuum:evidence from genomics.* World psychiatry, 2017. **16**(3): p. 227-235.
85. MacKenzie, A. and J.P. Quinn, *Post-genomic approaches to exploring neuropeptide gene mis-expression in disease.* Neuropeptides, 2004. **38**(1): p. 1-15.
86. Gaspar, P., O. Cases, and L. Maroteaux, *The developmental role of serotonin: news from mouse molecular genetics.* Nature reviews. Neuroscience, 2003. **4**(12): p. 1002-1012.
87. Schork, A.J., et al., *A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment.* Nature Neuroscience, 2019. **22**(3): p. 353-361.
88. Soldner, F., et al., *Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression.* Nature, 2016. **533**(7601): p. 95-9.
89. Feigin, M.E., et al., *Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma.* Nature Genetics, 2017. **49**(6): p. 825-833.
90. Huang, F.W., et al., *Highly recurrent TERT promoter mutations in human melanoma.* Science, 2013. **339**(6122): p. 957-9.
91. Horn, S., et al., *TERT promoter mutations in familial and sporadic melanoma.* Science, 2013. **339**(6122): p. 959-61.
92. Bell, R.J., et al., *Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer.* Science, 2015. **348**(6238): p. 1036-9.
93. Killela, P.J., et al., *TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal.* Proc Natl Acad Sci U S A, 2013. **110**(15): p. 6021-6.

94. Rachakonda, P.S., et al., *TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism.* Proc Natl Acad Sci U S A, 2013. **110**(43): p. 17426-31.
95. Mansour, M.R., et al., *Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element.* Science, 2014. **346**(6215): p. 1373-7.
96. Hoyt, S.J., et al., *From telomere to telomere: The transcriptional and epigenetic state of human repeat elements.* Science, 2022. **376**(6588): p. eabk3112.
97. Bakulski, K.M., et al., *Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex.* J Alzheimers Dis, 2012. **29**(3): p. 571-88.
98. Davies, G., et al., *A genome-wide association study implicates the APOE locus in nonpathological cognitive ageing.* Mol Psychiatry, 2014. **19**(1): p. 76-87.
99. Sachs-Ericsson, N.J., et al., *APOE epsilon4 allele carriers: Biological, psychological, and social variables associated with cognitive impairment.* Aging Ment Health, 2010. **14**(6): p. 679-91.
100. Hamilton, G., et al., *Alzheimer's disease genes are associated with measures of cognitive ageing in the lothian birth cohorts of 1921 and 1936.* Int J Alzheimers Dis, 2011. **2011**: p. 505984.
101. Bekris, L.M., F. Lutz, and C.-E. Yu, *Functional analysis of APOE locus genetic variation implicates regional enhancers in the regulation of both TOMM40 and APOE.* Journal of Human Genetics, 2012. **57**(1): p. 18-25.
102. Lyall, D.M., et al., *Alzheimer's Disease Susceptibility Genes APOE and TOMM40, and Hippocampal Volumes in the Lothian Birth Cohort 1936.* PloS one, 2013. **8**(11): p. e80513-e80513.
103. Roses, A.D., et al., *A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease.* The Pharmacogenomics Journal, 2010. **10**(5): p. 375-384.
104. Lutz, M.W., et al., *Genetic variation at a single locus and age of onset for Alzheimer's disease.* Alzheimers Dement, 2010. **6**(2): p. 125-31.
105. Jun, G., et al., *Comprehensive Search for Alzheimer Disease Susceptibility Loci in the APOE Region.* Archives of neurology (Chicago), 2012. **69**(10): p. 1-10.
106. Cruchaga, C., et al., *Association and Expression analyses with SNPs in TOMM40 in Alzheimer's Disease.* Archives of neurology (Chicago), 2011. **68**(8): p. 1013-1019.
107. Maruszak, A., et al., *TOMM40 rs10524523 Polymorphism's Role in Late-Onset Alzheimer's Disease and in Longevity.* Journal of Alzheimer's disease, 2012. **28**(2): p. 309-322.
108. Greenbaum, L., et al., *The TOMM40 poly-T rs10524523 variant is associated with cognitive performance among non-demented elderly with type 2 diabetes.* European neuropsychopharmacology, 2014. **24**(9): p. 1492-1499.

109. Caselli, R.J., et al., *Longitudinal modeling of cognitive aging and the TOMM40 effect.* Alzheimer's & dementia, 2012. **8**(6): p. 490-495.
110. Hayden, K.M., et al., *A homopolymer polymorphism in the TOMM40 gene contributes to cognitive performance in aging.* Alzheimers Dement, 2012. **8**(5): p. 381-8.
111. Johnson, S.C., et al., *The effect of TOMM40 poly-T length on gray matter volume and cognition in middle-aged persons with APOE ε3/ε3 genotype.* Alzheimer's & dementia, 2011. **7**(4): p. 456-465.
112. Bernardi, L., et al., *Role of TOMM40 rs10524523 polymorphism in onset of alzheimer's disease caused by the PSEN1 M146L mutation.* J Alzheimers Dis, 2013. **37**(2): p. 285-9.
113. Deters, K.D., et al., *TOMM40-APOE haplotypes are associated with cognitive decline in non-demented Blacks.* Alzheimers Dement, 2021. **17**(8): p. 1287-1296.
114. Payton, A., et al., *A TOMM40 poly-T variant modulates gene expression and is associated with vocabulary ability and decline in nonpathologic aging.* Neurobiology of aging, 2015. **39**: p. 217.e1-217.e7.
115. Humphries, A.D., et al., *Dissection of the Mitochondrial Import and Assembly Pathway for Human Tom40.* The Journal of biological chemistry, 2005. **280**(12): p. 11535-11543.
116. Chacinska, A., et al., *Importing mitochondrial proteins: machineries and mechanisms.* Cell, 2009. **138**(4): p. 628-44.
117. Fortney, K., et al., *Genome-Wide Scan Informed by Age-Related Disease Identifies Loci for Exceptional Human Longevity.* PLoS Genet, 2015. **11**(12): p. e1005728.
118. Swerdlow, R.H. and S.M. Khan, *The Alzheimer's disease mitochondrial cascade hypothesis: An update.* Experimental neurology, 2009. **218**(2): p. 308-315.
119. Valant, V., et al., *TOMM40 in Cerebral Amyloid Angiopathy Related Intracerebral Hemorrhage: Comparative Genetic Analysis with Alzheimer's Disease.* Translational stroke research, 2012. **3**(Suppl 1): p. 102-112.
120. Davies, G., et al., *A genome-wide association study implicates the APOE locus in nonpathological cognitive ageing.* Molecular psychiatry, 2014. **19**(1): p. 76-87.
121. Chen, L.-L. and G.G. Carmichael, *Gene regulation by SINES and inosines: biological consequences of A-to-I editing of Alu element inverted repeats.* Cell Cycle, 2008. **7**(21): p. 3294-3301.
122. Larsen, P.A., et al., *The Alu neurodegeneration hypothesis: A primate-specific mechanism for neuronal transcription noise, mitochondrial dysfunction, and manifestation of neurodegenerative disease.* Alzheimers Dement, 2017. **13**(7): p. 828-838.
123. Nishikura, K., *A-to-I editing of coding and non-coding RNAs by ADARs.* Nat Rev Mol Cell Biol, 2016. **17**(2): p. 83-96.
124. Larsen, P.A., et al., *Warning SINEs: Alu elements, evolution of the human brain, and the spectrum of neurological disease.* Chromosome Res, 2018. **26**(1-2): p. 93-111.

125. Linnertz, C., et al., *The cis-regulatory effect of an Alzheimer's disease-associated poly-T locus on expression of TOMM40 and apolipoprotein E genes.* Alzheimers Dement, 2014. **10**(5): p. 541-51.
126. Babenko, V.A., et al., *Miro1 Enhances Mitochondria Transfer from Multipotent Mesenchymal Stem Cells (MMSC) to Neural Cells and Improves the Efficacy of Cell Recovery.* Molecules, 2018. **23**(3).
127. Lutz, M.W., et al., *New Genetic Approaches to AD: Lessons from APOE-TOMM40 Phylogenetics.* Current Neurology and Neuroscience Reports, 2016. **16**(5): p. 48.
128. Kulminski, A.M., et al., *Associations of the APOE ε2 and ε4 alleles and polygenic profiles comprising APOE-TOMM40-APOC1 variants with Alzheimer's disease biomarkers.* Aging, 2022.
129. Kulminski, A.M., et al., *Definitive roles of TOMM40-APOE-APOC1 variants in the Alzheimer's risk.* Neurobiology of Aging, 2022. **110**: p. 122-131.
130. Payton, A., et al., *A TOMM40 poly-T variant modulates gene expression and is associated with vocabulary ability and decline in non-pathological ageing.* Neurobiology of Aging.
131. Savage, A.L., et al., *Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns.* BMC Evolutionary Biology, 2013. **13**(1): p. 101.
132. Hancks, D.C. and H. Kazazian, *SVA retrotransposons: Evolution and genetic instability.* Seminars in cancer biology, 2010. **20**(4): p. 234-245.
133. Andreou, D., et al., *Polymorphisms in genes implicated in dopamine, serotonin and noradrenalin metabolism suggest association with cerebrospinal fluid monoamine metabolite concentrations in psychosis.* Behavioral and Brain Functions, 2014. **10**(1): p. 1-10.
134. Zhang, Y., et al., *Gene-Gene-Environment Interactions of Serotonin Transporter, Monoamine Oxidase A and Childhood Maltreatment Predict Aggressive Behavior in Chinese Adolescents.* Front Behav Neurosci, 2017. **11**: p. 17.
135. Fergusson, D.M., et al., *Moderating role of the MAOA genotype in antisocial behaviour.* British journal of psychiatry, 2012. **200**(2): p. 116-123.
136. Haberstick, B.C., et al., *MAOA Genotype, Childhood Maltreatment, and Their Interaction in the Etiology of Adult Antisocial Behaviors.* Biological Psychiatry, 2014. **75**(1): p. 25-30.
137. Byrd, A.L. and S.B. Manuck, *MAOA, Childhood Maltreatment, and Antisocial Behavior: Meta-analysis of a Gene-Environment Interaction.* Biological Psychiatry, 2014. **75**(1): p. 9-17.
138. Mentis, A.-F.A., et al., *From warrior genes to translational solutions: novel insights into monoamine oxidases (MAOs) and aggression.* Translational Psychiatry, 2021. **11**(1): p. 130.

139. Reif, A., et al., *Meta-analysis argues for a female-specific role of MAOA-uVNTR in panic disorder in four European populations.* American journal of medical genetics. Part B, Neuropsychiatric genetics, 2012. **159B**(7): p. 786-793.
140. Takahashi, A., et al., *Behavioral and Pharmacogenetics of Aggressive Behavior*, in *Behavioral Neurogenetics*, J.F. Cryan and A. Reif, Editors. 2012, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 73-138.
141. Brunner, H.G., et al., *Abnormal Behavior Associated with a Point Mutation in the Structural Gene for Monoamine Oxidase A.* Science, 1993. **262**(5133): p. 578-580.
142. Godar, S.C., et al., *The role of monoamine oxidase A in aggression: Current translational developments and future challenges.* Progress in Neuro-Psychopharmacology and Biological Psychiatry, 2016. **69**: p. 90-100.
143. Popova, N.K., *From genes to aggressive behavior: the role of serotonergic system.* BioEssays, 2006. **28**(5): p. 495-503.
144. Ficks, C.A. and I.D. Waldman, *Candidate genes for aggression and antisocial behavior: a meta-analysis of association studies of the 5HTTLPR and MAOA-uVNTR.* Behav Genet, 2014. **44**(5): p. 427-44.
145. Comings, D.E., et al., *A multivariate analysis of 59 candidate genes in personality traits: the temperament and character inventory.* Clin Genet, 2000. **58**(5): p. 375-85.
146. Comings, D.E., et al., *The DRD4 gene and the spiritual transcendence scale of the character temperament index.* Psychiatr Genet, 2000. **10**(4): p. 185-9.
147. Adkins, D.E., et al., *The Influence of Five Monoamine Genes on Trajectories of Depressive Symptoms across Adolescence and Young Adulthood.* Development and psychopathology, 2012. **24**(1): p. 267-285.
148. Wermter, A.-K., et al., *From nature versus nurture, via nature and nurture, to gene × environment interaction in mental disorders.* European Child & Adolescent Psychiatry, 2010. **19**(3): p. 199-210.
149. Zhuravleva, E., et al., *Acyl Coenzyme A Thioesterase Them5/Acot15 Is Involved in Cardiolipin Remodeling and Fatty Liver Development.* Molecular and Cellular Biology, 2012. **32**(14): p. 2685-2697.
150. Guffanti, G., et al., *LINE1 insertions as a genomic risk factor for schizophrenia: Preliminary evidence from an affected family.* American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2016. **171**(4): p. 534-545.
151. Aberg, K.A., et al., *Methylome-wide association study of schizophrenia: Identifying blood biomarker signatures of environmental insults.* JAMA Psychiatry, 2014. **71**(3): p. 255-264.
152. Beach, S.R.H., et al., *Child Maltreatment Moderates the Association of MAOA With Symptoms of Depression and Antisocial Personality Disorder.* Journal of family psychology, 2010. **24**(1): p. 12-20.
153. Sinha, S. and B.M. Jones, *Behavior-related gene regulatory networks: A new level of organization in the brain.* 2020. **117**(38): p. 23270-23279.

154. Melas, P.A., et al., *Genetic and epigenetic associations of MAOA and NR3C1 with depression and childhood adversities.* The international journal of neuropsychopharmacology, 2013. **16**(7): p. 1513-1528.
155. Magwai, T., et al., *DNA Methylation and Schizophrenia: Current Literature and Future Perspective.* Cells, 2021. **10**(11).
156. Hill, J., et al., *Evidence for interplay between genes and maternal stress in utero: monoamine oxidase A polymorphism moderates effects of life events during pregnancy on infant negative emotionality at 5weeks.* Genes, brain and behavior, 2013. **12**(4): p. 388.
157. Briere, J. and C.E. Jordan, *Childhood maltreatment, intervening variables, and adult psychological difficulties in women: An overview.* Trauma, Violence, & Abuse, 2009. **10**: p. 375-388.
158. Holz, N., et al., *Evidence for a Sex-Dependent MAOA× Childhood Stress Interaction in the Neural Circuitry of Aggression.* Cerebral Cortex, 2014.
159. Philibert, R.A., et al., *Gene Environment Interactions with a Novel Variable Monoamine Oxidase A Transcriptional Enhancer are Associated with Antisocial Personality Disorder.* Biological psychology, 2011. **87**(3): p. 366-371.
160. Manca, M., et al., *The regulation of monoamine oxidase A gene expression by distinct variable number tandem repeats.* 2018.
161. Zhang, M., et al., *Interactions between Intestinal Microbiota and Host Immune Response in Inflammatory Bowel Disease.* Front Immunol, 2017. **8**: p. 942.
162. Ji, B., et al., *Over-expression of XIST, the Master Gene for X Chromosome Inactivation, in Females With Major Affective Disorders.* EBioMedicine, 2015. **2**(8): p. 909-918.
163. Joo, J.E., et al., *Human active X-specific DNA methylation events showing stability across time and tissues.* European journal of human genetics : EJHG, 2014. **22**(12): p. 1376-1381.
164. Mugford, J.W., et al., *Evidence for Local Regulatory Control of Escape from Imprinted X Chromosome Inactivation.* Genetics (Austin), 2014. **197**(2): p. 715-723.
165. Caspi, A., et al., *Role of Genotype in the Cycle of Violence in Maltreated Children.* Science, 2002. **297**(5582): p. 851-854.
166. Kim-Cohen, J., et al., *MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis.* Mol Psychiatry, 2006. **11**(10): p. 903-913.
167. Åslund, C., et al., *Maltreatment, MAOA, and Delinquency: Sex Differences in Gene–Environment Interaction in a Large Population-Based Cohort of Adolescents.* Behavior Genetics, 2011. **41**(2): p. 262-272.
168. Nikulina, V., C.S. Widom, and L.M. Brzustowicz, *Child Abuse and Neglect, MAOA, and Mental Health Outcomes: A Prospective Examination.* Biological Psychiatry, 2012. **71**(4): p. 350-357.

169. Sabol, S.Z., S. Hu, and D. Hamer, *A functional polymorphism in the monoamine oxidase A gene promoter.* Human Genetics, 1998. **103**(3): p. 273-279.
170. Fan, M., et al., *Meta-analysis of the association between the monoamine oxidase-A gene and mood disorders.* Psychiatric genetics, 2010. **20**(1): p. 1-7.
171. Philibert, R.A., et al., *Gene environment interactions with a novel variable Monoamine Oxidase A transcriptional enhancer are associated with antisocial personality disorder.* 2011.
172. Reif, A., et al., *MAOA and mechanisms of panic disorder revisited: from bench to molecular psychotherapy.* Molecular psychiatry, 2014. **19**(1): p. 122-128.
173. Carette, J.E., et al., *Haploid Genetic Screens in Human Cells Identify Host Factors Used by Pathogens.* Science, 2009. **326**(5957): p. 1231-1235.
174. Daw, J., et al., *Genetic sensitivity to peer behaviors: 5HTTLPR, smoking, and alcohol consumption.* Journal of health and social behavior, 2013. **54**(1): p. 92-108.
175. Kravić, N., et al., *Association analysis of maoa and SLC6A4 gene variation in South East European War related posttraumatic stress disorder.* Psychiatria Danubina, 2019. **31**(2): p. 211-218.
176. Haberstick, B.C., A. Smolen, and J.K. Hewitt, *Family-based association test of the 5HTTLPR and aggressive behavior in a general population sample of children.* Biological Psychiatry, 2006. **59**(9): p. 836-843.
177. Liu, Z., et al., *MAOA Variants and Genetic Susceptibility to Major Psychiatric Disorders.* Molecular Neurobiology, 2015: p. 1-9.
178. Kilpatrick, D.G., et al., *National estimates of exposure to traumatic events and PTSD prevalence using DSM-IV and DSM-5 criteria.* Journal of traumatic stress, 2013. **26**(5): p. 537-547.
179. Maul, S., et al., *Genetics of resilience: Implications from genome-wide association studies and candidate genes of the stress response system in posttraumatic stress disorder and depression.* American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2020. **183**(2): p. 77-94.
180. Yin, L., et al., *TPH, SLC6A2, SLC6A3, DRD2 and DRD4 Polymorphisms and Neuroendocrine Factors Predict SSRIs Treatment Outcome in the Chinese Population with Major Depression.* Pharmacopsychiatry, 2015. **48**(3): p. 95-103.
181. Gokturk, C., et al., *Serotonin transporter (5-HTTLPR) and monoamine oxidase (MAOA) promoter polymorphisms in women with severe alcoholism.* Archives of Women's Mental Health, 2008. **11**(5-6): p. 347-355.
182. Greenberg, B.D., et al., *Genetic variation in the serotonin transporter promoter region affects serotonin uptake in human blood platelets.* American journal of medical genetics, 1999. **88**(1): p. 83-87.
183. Wang, S.-W., et al., *Jun dimerization protein 2 in oxygen restriction; control of senescence.* Current Pharmaceutical Design, 2011. **17**(22): p. 2278-2289.

184. Praschak-Rieder, N., et al., *Novel 5-HTTLPR allele associates with higher serotonin transporter binding in putamen: a [11C] DASB positron emission tomography study.* Biological psychiatry, 2007. **62**(4): p. 327-331.
185. Haase, C.M., et al., *Short alleles, bigger smiles? The effect of 5-HTTLPR on positive emotional expressions.* Emotion, 2015. **15**(4): p. 438.
186. Xie, P., et al., *Serotonin transporter 5-HTTLPR genotype moderates the effects of childhood adversity on posttraumatic stress disorder risk: A replication study.* American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2012. **159**(6): p. 644-652.
187. Pietrzak, R.H., et al., *Examining the relation between the serotonin transporter 5-HTTPLR genotype x trauma exposure interaction on a contemporary phenotypic model of posttraumatic stress symptomatology: a pilot study.* Journal of affective disorders, 2013. **148**(1): p. 123-128.
188. Sarosi, A., et al., *Association of the STin2 polymorphism of the serotonin transporter gene with a neurocognitive endophenotype in major depressive disorder.* Progress in Neuro-Psychopharmacology and biological psychiatry, 2008. **32**(7): p. 1667-1672.
189. Heils, A., et al., *Allelic variation of human serotonin transporter gene expression.* Journal of neurochemistry, 1996. **66**(6): p. 2621-2624.
190. Hemmings, S.M.J., et al., *Appetitive and reactive aggression are differentially associated with the STin2 genetic variant in the serotonin transporter gene.* Scientific reports, 2018. **8**(1): p. 1-9.
191. Saab, B.J., et al., *NCS-1 in the dentate gyrus promotes exploration, synaptic plasticity, and rapid acquisition of spatial memory.* Neuron, 2009. **63**(5): p. 643-656.
192. Schlomer, G.L., et al., *Developmental differences in early adolescent aggression: A gene× environment× intervention analysis.* Journal of youth and adolescence, 2015. **44**(3): p. 581-597.
193. Dmitrieva, J., et al., *Gender-specific expression of the DRD4 gene on adolescent delinquency, anger and thrill seeking.* Social cognitive and affective neuroscience, 2011. **6**(1): p. 82-89.
194. Buchmann, A.F., et al., *Moderating role of FKBP5 genotype in the impact of childhood adversity on cortisol stress response during adulthood.* European Neuropsychopharmacology, 2014. **24**(6): p. 837-845.
195. Dragan, W.L. and W. Oniszczenko, *The association between dopamine D4 receptor exon III polymorphism and intensity of PTSD symptoms among flood survivors.* Anxiety, Stress, & Coping, 2009. **22**(5): p. 483-495.
196. Watanabe, K., et al., *A global overview of pleiotropy and genetic architecture in complex traits.* Nature genetics, 2019. **51**(9): p. 1339-1348.
197. Stolf, A.R., et al., *Effects of DRD2 splicing-regulatory polymorphism and DRD4 48 bp VNTR on crack cocaine addiction.* Journal of Neural Transmission, 2019. **126**(2): p. 193-199.

198. Sharp, H., et al., *Frequency of Infant Stroking Reported by Mothers Moderates the Effect of Prenatal Depression on Infant Behavioural and Physiological Outcomes.* PLOS ONE, 2012. **7**(10): p. e45446.
199. Noble M, W.G., Dibben C, Smith G, McLennan D, et al., *The English Indices of Deprivation 2004 (revised).* Report to the Office of the Deputy Prime Minister. London: Neighbourhood Renewal Unit, 2004.
200. Rabbitt, P.M.A., et al., *The University of Manchester Longitudinal Study of Cognition in Normal Healthy Old Age, 1983 through 2003.* Aging, Neuropsychology, and Cognition, 2004. **11**(2-3): p. 245-279.
201. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-265.
202. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
203. Kent, W.J., et al., *The Human Genome Browser at UCSC.* Genome Res, 2002. **12**.
204. Sham, P. and D.J.A.o.h.g. Curtis, *Monte Carlo tests for associations between disease and alleles at highly polymorphic loci.* 1995. **59**(1): p. 97-105.
205. Rodriguez, S., T.R. Gaunt, and I.N.J.A.j.o.e. Day, *Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies.* 2009. **169**(4): p. 505-514.
206. *World Health Organization. Global action Plan on the Public Health Response to Dementia 2017–2025.*
207. Lin, R., et al., *Association of common variants in TOMM40/APOE/APOC1 region with human longevity in a Chinese population.* Journal of Human Genetics, 2016. **61**(4): p. 323-328.
208. Fuior, E.V. and A.V. Gafencu, *Apolipoprotein C1: Its Pleiotropic Effects in Lipid Metabolism and Beyond.* 2019. **20**(23).
209. Roses, A.D., et al., *A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease.* Pharmacogenomics J, 2010. **10**(5): p. 375-84.
210. Ghebranious, N., et al., *Detection of ApoE E2, E3 and E4 alleles using MALDI-TOF mass spectrometry and the homogeneous mass-extend technology.* Nucleic Acids Research, 2005. **33**(17): p. e149-e149.
211. Roses, A.D., et al., *TOMM40 and APOE : Requirements for replication studies of association with age of disease onset and enrichment of a clinical trial.* Alzheimer's & Dementia, 2013. **9**(2): p. 132-136.
212. Baillie, J.K., et al., *Somatic retrotransposition alters the genetic landscape of the human brain.* Nature, 2011. **479**(7374): p. 534-537.
213. Savage, A.L., et al., *An Evaluation of a SVA Retrotransposon in the FUS Promoter as a Transcriptional Regulator and Its Association to ALS.* 2014.
214. Linnertz, C., et al., *Characterization of the Poly-T Variant in the TOMM40 Gene in Diverse Populations.* PLoS ONE, 2012. **7**(2): p. e30994-e30994.

215. Hancks, D.C. and H.H. Kazazian, Jr., *Active human retrotransposons: variation and disease.* Curr Opin Genet Dev, 2012. **22**(3): p. 191-203.
216. 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, 2014. **9**(6): p. e90833.
217. Kanekiyo, T., H. Xu, and G. Bu, *ApoE and Aβ in Alzheimer's Disease: Accidental Encounters or Partners?* Neuron, 2014. **81**(4): p. 740-754.
218. Bishop, N.A., T. Lu, and B.A. Yankner, *Neural mechanisms of ageing and cognitive decline.* Nature (London), 2010. **464**(7288): p. 529-535.
219. Erwin, J.A., M.C. Marchetto, and F.H. Gage, *Mobile DNA elements in the generation of diversity and complexity in the brain.* Nat Rev Neurosci, 2014. **15**(8): p. 497-506.
220. Quinn, J.P. and V.J. Bubb, *SVA retrotransposons as modulators of gene expression.* Mobile Genetic Elements, 2014. **4**(4): p. e32102-e32102.
221. 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, 2014. **9**(3): p. e90833.
222. Baillie, J.K., et al., *Somatic retrotransposition alters the genetic landscape of the human brain.* Nature (London), 2011. **479**(7374): p. 534-537.
223. Bundo, M., et al., *Increased L1 Retrotransposition in the Neuronal Genome in Schizophrenia.* Neuron, 2014. **81**(2): p. 306-313.
224. Li, W., et al., *Transposable Elements in TDP-43-Mediated Neurodegenerative Disorders.* PLoS ONE, 2012. **7**(9): p. e44099-e44099.
225. Aneichyk, T., et al., *Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly.* Cell, 2018. **172**(5): p. 897-909.e21.
226. Lyall, D.M., et al., *Are APOE ε genotype and TOMM40 poly-T repeat length associations with cognitive ageing mediated by brain white matter tract integrity?* Translational Psychiatry, 2014. **4**(9): p. e449-e449.
227. Hancks, D.C. and H.H. Kazazian, Jr., *SVA retrotransposons: Evolution and genetic instability.* Semin Cancer Biol, 2010. **20**(4): p. 234-45.
228. Kwon, Y.J., et al., *Structure and Expression Analyses of SVA Elements in Relation to Functional Genes.* Genomics Inform, 2013. **11**(3): p. 142-8.
229. Makino, S., et al., *Reduced Neuron-Specific Expression of the TAF1 Gene Is Associated with X-Linked Dystonia-Parkinsonism.* American journal of human genetics, 2007. **80**(3): p. 393-406.
230. Delvallée, C., et al., *A BBS1 SVA F retrotransposon insertion is a frequent cause of Bardet-Biedl syndrome.* Clinical genetics, 2021. **99**(2): p. 318-324.
231. Yamamoto, G., et al., *SVA retrotransposon insertion in exon of MMR genes results in aberrant RNA splicing and causes Lynch syndrome.* European journal of human genetics : EJHG, 2021. **29**(4): p. 680-686.

232. 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.* Proceedings of the National Academy of Sciences - PNAS, 2017. **114**(51): p. E11020-E11028.
233. Wang, L., et al., *Human population-specific gene expression and transcriptional network modification with polymorphic transposable elements.* Nucleic Acids Research, 2017. **45**(5): p. 2318-2328.
234. Eslami Rasekh, M., et al., *Genome-wide characterization of human minisatellite VNTRs: population-specific alleles and gene expression differences.* Nucleic Acids Research, 2021. **49**(8): p. 4308-4324.
235. Caspi, A. and T.E. Moffitt, *Gene-environment interactions in psychiatry: joining forces with neuroscience.* Nat Rev Neurosci, 2006. **7**(7): p. 583-590.
236. Lung, F.-W., et al., *Association of the MAOA promoter uVNTR polymorphism with suicide attempts in patients with major depressive disorder.* BMC Medical Genetics, 2011. **12**: p. 74-74.
237. Enoch, M.A., et al., *Early life stress, MAOA, and gene-environment interactions predict behavioral disinhibition in children.* Genes, brain and behavior, 2010. **9**(1): p. 65-74.
238. Carrel, L. and H.F. Willard, *X-inactivation profile reveals extensive variability in X-linked gene expression in females.* Nature, 2005. **434**(7031): p. 400-404.
239. Michelotti, E.F., et al., *Cellular Nucleic Acid Binding Protein Regulates the CT Element of the Human c- myc Protooncogene.* The Journal of biological chemistry, 1995. **270**(16): p. 9494-9499.
240. Kang, Y., et al., *Dynamic transcriptional regulatory complexes including BORIS, CTCF and Sp1 modulate NY-ESO-1 expression in lung cancer cells.* Oncogene, 2007. **26**(30): p. 4394-4403.
241. Balakrishnan, S.K., et al., *Functional and Molecular Characterization of the Role of CTCF in Human Embryonic Stem Cell Biology.* PLoS ONE, 2012. **7**(8): p. e42424-e42424.
242. Gaweska, H. and P.F. Fitzpatrick, *Structures and Mechanism of the Monoamine Oxidase Family.* Biomol Concepts, 2011. **2**(5): p. 365-377.
243. De Colibus, L., et al., *Three-dimensional structure of human monoamine oxidase A (MAO A): relation to the structures of rat MAO A and human MAO B.* Proc Natl Acad Sci U S A, 2005. **102**(36): p. 12684-9.
244. Essletzbichler, P., et al., *Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line.* Genome Res, 2014. **24**(12): p. 2059-65.
245. Musso, M., et al., *Betaine, Dimethyl Sulfoxide, and 7-Deaza-dGTP, a Powerful Mixture for Amplification of GC-Rich DNA Sequences.* The Journal of molecular diagnostics : JMD, 2006. **8**(5): p. 544-550.
246. Jensen, M.A., M. Fukushima, and R.W. Davis, *DMSO and Betaine Greatly Improve Amplification of GC-Rich Constructs in De Novo Synthesis.* PLOS ONE, 2010. **5**(6): p. e11024.

247. Henke, W., et al., *Betaine improves the PCR amplification of GC-rich DNA sequences.* Nucleic Acids Res, 1997. **25**(19): p. 3957-8.
248. Korbie, D.J. and J.S. Mattick, *Touchdown PCR for increased specificity and sensitivity in PCR amplification.* Nature Protocols, 2008. **3**(9): p. 1452-1456.
249. Gaunt, T.R., et al., *MIDAS: software for analysis and visualisation of interallelic disequilibrium between multiallelic markers.* BMC Bioinformatics, 2006. **7**(1): p. 227.
250. Wong, C.C.Y., et al., *A longitudinal study of epigenetic variation in twins.* Epigenetics, 2010. **5**(6): p. 516-526.
251. Domschke, K., et al., *Monoamine oxidase A gene DNA hypomethylation – a risk factor for panic disorder?* The international journal of neuropsychopharmacology, 2012. **15**(9): p. 1217-1228.
252. Warburton, A., et al., *A GWAS SNP for Schizophrenia Is Linked to the Internal MIR137 Promoter and Supports Differential Allele-Specific Expression.* Schizophr Bull, 2016. **42**(4): p. 1003-8.
253. Stępnicki, P., M. Kondej, and A.A. Kaczor, *Current Concepts and Treatments of Schizophrenia.* 2018. **23**(8): p. 2087.
254. Kendler, K.S. and S.R. Diehl, *The Genetics of Schizophrenia: A Current, Genetic-epidemiologic Perspective.* Schizophrenia Bulletin, 1993. **19**(2): p. 261-285.
255. Bortolato, M., et al., *The role of serotonergic system at the interface of aggression and suicide.* Neuroscience, 2013. **236**: p. 160-185.
256. Tunbridge, E.M., et al., *Which Dopamine Polymorphisms Are Functional? Systematic Review and Meta-analysis of COMT, DAT, DBH, DDC, DRD1–5, MAOA, MAOB, TH, VMAT1, and VMAT2.* Biological Psychiatry, 2019. **86**(8): p. 608-620.
257. Schildkraut, J.J., et al., *Platelet monoamine oxidase activity in subgroups of schizophrenic disorders.* Schizophrenia bulletin, 1980. **6**(2): p. 220-225.
258. Meyer-Lindenberg, A., et al., *Impact of complex genetic variation in COMT on human brain function.* Molecular psychiatry, 2006. **11**(9): p. 867-877.
259. Berger, P.A., et al., *Platelet monoamine oxidase in chronic schizophrenic patients.* The American journal of psychiatry, 1978. **135**(1): p. 95-99.
260. Meyer, J.H., et al., *Elevated Monoamine Oxidase A Levels in the Brain: An Explanation for the Monoamine Imbalance of Major Depression.* Archives of general psychiatry, 2006. **63**(11): p. 1209-1216.
261. Ziegler, C. and K. Domschke, *Epigenetic signature of MAOA and MAOB genes in mental disorders.* Journal of Neural Transmission, 2018. **125**(11): p. 1581-1588.
262. Jönsson, E.G., et al., *Association between a promoter variant in the monoamine oxidase A gene and schizophrenia.* Schizophrenia research, 2003. **61**(1): p. 31-37.
263. Qiu, H.T., et al., *Association between monoamine oxidase (MAO)-A gene variants and schizophrenia in a Chinese population.* Brain research, 2009. **1287**: p. 67-73.
264. Sun, Y., et al., *Study of a possible role of the monoamine oxidase A (MAOA) gene in paranoid schizophrenia among a Chinese population.* 2012.

265. Koks, G., et al., *Genetic interaction between two VNTRs in the MAOA gene is associated with the nicotine dependence.* 2020. **245**(8): p. 733-739.
266. Tanifuji, T., et al., *Association of two variable number of tandem repeats in the monoamine oxidase a gene promoter with schizophrenia.* Neuropsychiatric disease and treatment, 2021. **17**: p. 3315-3323.
267. Strassnig, M., et al., *Physical performance and disability in schizophrenia.* Schizophr Res Cogn, 2014. **1**(2): p. 112-121.
268. Nielsen, P.R., T.M. Laursen, and E. Agerbo, *Comorbidity of schizophrenia and infection: a population-based cohort study.* Social Psychiatry and Psychiatric Epidemiology, 2016. **51**(12): p. 1581-1589.
269. Bitter, I., et al., *Mortality and the relationship of somatic comorbidities to mortality in schizophrenia. A nationwide matched-cohort study.* European Psychiatry, 2017. **45**: p. 97-103.
270. Chong, H.Y., et al., *Global economic burden of schizophrenia: a systematic review.* Neuropsychiatric Disease and Treatment, 2016. **12**: p. 357-373.
271. Nordstroem, A.L., et al., *Burden of illness of people with persistent symptoms of schizophrenia: A multinational cross-sectional study.* Int J Soc Psychiatry, 2017. **63**(2): p. 139-150.
272. Rabinowitz, J., et al., *Association of prominent positive and prominent negative symptoms and functional health, well-being, healthcare-related quality of life and family burden: A CATIE analysis.* 2013. **150**: p. 339-342.
273. Hilker, R., et al., *Is an Early Age at Illness Onset in Schizophrenia Associated With Increased Genetic Susceptibility? Analysis of Data From the Nationwide Danish Twin Register.* EBioMedicine, 2017. **18**: p. 320-326.
274. van der Werf, M., et al., *Systematic review and collaborative recalculation of 133 693 incident cases of schizophrenia.* Psychological Medicine, 2014. **44**(1): p. 9-16.
275. Jaffe, A.E. and J.E. Kleinman, *Genetic and epigenetic analysis of schizophrenia in blood—a no-brainer?* Genome Medicine, 2016. **8**(1): p. 96.
276. Viana, J., et al., *Schizophrenia-associated methylomic variation: molecular signatures of disease and polygenic risk burden across multiple brain regions.* Human Molecular Genetics, 2016. **26**(1): p. 210-225.
277. McCarty, R., *Cross-fostering: Elucidating the effects of gene×environment interactions on phenotypic development.* Neuroscience & Biobehavioral Reviews, 2017. **73**: p. 219-254.
278. Kebir, O., et al., *Methylomic changes during conversion to psychosis.* Mol Psychiatry, 2017. **22**(4): p. 512-518.
279. Mandelli, L. and A. Serretti, *Gene environment interaction studies in depression and suicidal behavior: An update.* Neuroscience & Biobehavioral Reviews, 2013. **37**(10, Part 1): p. 2375-2397.

280. Hung, C.-F., et al., *Monoamine oxidase A gene polymorphism and suicide: An association study and meta-analysis.* Journal of Affective Disorders, 2012. **136**(3): p. 643-649.
281. Huang, Y.-y., et al., *An Association between a Functional Polymorphism in the Monoamine Oxidase A Gene Promoter, Impulsive Traits and Early Abuse Experiences.* Neuropsychopharmacology, 2004. **29**(8): p. 1498-1505.
282. Antypa, N., et al., *MAOA and MAOB polymorphisms and anger-related traits in suicidal participants and controls.* European Archives Of Psychiatry And Clinical Neuroscience, 2013. **263**(5): p. 393-403.
283. Courtet, P., et al., *Suicidal behavior: Relationship between phenotype and serotonergic genotype.* 2005. **133C**(1): p. 25-33.
284. Zohsel, K., et al., *Mothers' prenatal stress and their children's antisocial outcomes – a moderating role for the Dopamine D4 Receptor (DRD4) gene.* Journal of Child Psychology and Psychiatry, 2014. **55**(1): p. 69-76.
285. Murgatroyd, C. and D. Spengler, *Genetic Variation in the Epigenetic Machinery and Mental Health.* 2012. **14**(2): p. 138-149.
286. Parikshak, N.N., M.J. Gandal, and D.H. Geschwind, *Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders.* Nat Rev Genet, 2015. **16**(8): p. 441-458.
287. Holmboe, K., et al., *Dopamine D4 receptor and serotonin transporter gene effects on the longitudinal development of infant temperament.* Genes Brain Behav, 2011. **10**(5): p. 513-22.
288. Green, C.G., et al., *Prenatal maternal depression and child serotonin transporter linked polymorphic region (5-HTTLPR) and dopamine receptor D4 (DRD4) genotype predict negative emotionality from 3 to 36 months.* Dev Psychopathol, 2017. **29**(3): p. 901-917.
289. Linthorst, J., et al., *Extreme enrichment of VNTR-associated polymorphicity in human subtelomeres: genes with most VNTRs are predominantly expressed in the brain.* Translational psychiatry, 2020. **10**(1): p. 369-369.
290. Windhorst, D.A., et al., *Differential susceptibility in a developmental perspective: DRD4 and maternal sensitivity predicting externalizing behavior.* Developmental Psychobiology, 2015. **57**(1): p. 35-49.
291. Fan, J.B. and P. Sklar, *Meta-analysis reveals association between serotonin transporter gene STin2 VNTR polymorphism and schizophrenia.* Mol Psychiatry, 2005. **10**(10): p. 928-938.
292. Hill, J., et al., *Evidence for interplay between genes and maternal stress in utero : monoamine oxidase A polymorphism moderates effects of life events during pregnancy on infant negative emotionality at 5 weeks: MAOA by life events in pregnancy.* Genes, brain and behavior, 2013. **12**(4): p. 388-396.