Role of Variable Number Tandem Repeats (VNTRs) on gene expression in the CNS

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

Maurizio Manca

December 2016

Acknowledgements

What's the most difficult thing you've ever done in your life until now? A PhD abroad, in England. Yes, growing up in a place where the sun reigns, and then, suddenly move to the reign of clouds. It hasn't be easy at first. But, with some help, after a while, I started to think at Liverpool as a second home.

My first THANK YOU goes to Michela, undoubtedly, for supporting me, being always on my side and, over everything, she endured me. And trust me on this, it's not an easy task. Without her I wouldn't have been here in the first place. :*

Another big thank you goes to my parents and my brother, that, not without some sacrifice, believed in me and helped me starting this amazing journey always supporting me. Vi voglio bene e sempre ve ne vorrò.

A special thanks to Giovanni, one of the best friends a person can wish for. Always there when you need him. Grazie cugì.

Another great thank you to both my supervisors: Dr. Helen Sharp and Prof. John Quinn for the amazing opportunity they gave me.

Prof. Quinn, what can I say about him, probably one of the best people I have ever met. He trusted a "random" guy knocking at his door and gave him a chance to prove himself. I Hope I met your expectations, at least partially, without letting you down. I gave my 100%, always. In the darkest (scientific) moments of my PhD he was able to bring knowledge, and he always had words of encouragement in the most frustrating days.

Veri, you've been (and hope you will always be) a great friend, always ready to give everything to others. Now enjoy your new life that you fully deserve. I wish you all the best the life can ever offer you.

A big thanks also goes to Jill, Paul, Alix and Abi that, especially when I first arrived, barely speaking and understanding English, helped me feel at home and always supported me with advice and knowledge.

How not to mention the "new" group that arrived after me: Oly, Kim, Jack, Ben, Ana, Emma. Thank you for the time we spent together chatting, having fun (hopefully) and sometimes talking about science. I wish you all the best.

Well, I cried enough now. Thank you for everything.

Contents

Abbreviations

5-HTT	5-hydroxy tryptamine (serotonin)
5-HTTLPR	Serotonin-transporter-linked polymorphic region
AA	Amino Acid
AD	Alzheimer's disease
ADHD	Attention-Deficit/Hyperactivity Disorder
AMP	Adenosine monophosphate
APOE	Apolipoprotein E
ATP	Adenosine-5′-triphosphate
AVP	Arginine vasopressin
BDNF	Brain-derived neurotrophic factor
BLAST	Basic Local Alignment Search Tool
Bp	Base pairs
BS	Binding site
C-	Carboxy-
CACNA1C	Calcium Channel, Voltage-Dependent, L Type, Alpha 1C Subunit
cAMP	Cyclic adenosine monophosphate
CAS9	CRISPR proteinassociate 9
cDNA	Complementary deoxyribonucleic acid
ChIP	Chromatin immunoprecipitation
CNS	Central nervous system
CNV	Copy number variants
CoREST	Cofactor for REST
CpG	CG dinucleotides
CREB	cAMP response element-binding protein
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
CTCF	CCCTC binding-protein
d.f.	Degrees of freedom
DA	Dopamine
DAT	Dopamine transporter
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DNMT1	DNA (cytosine-5)-methyltransferase 1
DNMTs	DNA methyltransferases
dNTP	Deoxynucleotide triphosphate
DRD3	Dopamine D3
ECR	Evolutionary conserved region
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EGTA	Ethylene glycol tetraacetic acid,
ENCODE	Encyclopaedia of DNA Elements

EZH2	Enhancer of zeste homolog 2
FAD	Flavin Adenine Dinucleotide
FMR1	Fragile Xmentalretardation 1 gene
FMRP	Fragile-X mental retardation protein
FRAXA	Fragile X mental retardation syndrome
FTD	Fronto Temporal Dementia
FXS	Fragile-X sydrome
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
HDAC	Histone deacetylase complex
HeLa	Henrietta Lacks clonal cell line
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hg	Human genome
HSAN1	Hereditary sensory and autonomic neuropathy type 1
Htt	Huntingtin
HWE	Hardy-Weinberg Equilibrium
In-del	Insertion - Deletion
Kb	Kilobase
KBM-7	Chronic myelogenous leukemia (CML) 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 II	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
SCN2A	Sodium channel type II-alpha
SCZ	Schizophrenia
SH-SY5Y	Human-derived neuroblastoma cells
SIN3A	SIN3 Transcription Regulator Family Member A
SINE	Short interspersed nuclear element
SLC	Solute carrier gene family
SNP	Single Nucleotide Polymorphism
SP1	Specificity protein 1
SSR	Simple Sequence Repeats
STR	Short Tandem Repeat
SVA	SINE-VNTR-Alu element

TBE	Tris/Borate/EDTA
TBP	TATA-box-binding protein
TE	Tris-EDTA/transposable element
TF	Transcription Factor
TFBS	Transcription factor binding sites
TR	Tandem repeat
TSG	Tumour Suppressive Gene
TSS	Transcriptional start site
UCSC	University of California, Santa Cruz
UTR	Untranslated region
VNTR	Variable Number Tandem Repeat
VPA	Valproic Acid
XIST	X-inactive specific transcript gene
μ	Mu-

List of Figures

List of Tables

Abstract

It is now known that at least 80% of the human genome is composed of non-coding DNA which has a biochemical activity and is involved in a wide range of activities and mechanisms. Among these, epigenetic modifications, cis-trans gene expression regulation, transcription factor binding sequences, are the most studied.

Non-coding DNA is often characterised by a polymorphic and repetitive nature and it is composed of a high density of GC nucleotides. These polymorphic and repetitive regions within the population may represent either protective elements or risk factors, based on population studies in various diseases, for several conditions and at the same time have the power to shape our behaviours or wellbeing.

The compositions of transcription factor binding sites (BSs) and epigenetic factors at these regions act in concert with external and environmental factors to modify gene function and gene expression.

This combined effect of environmental and genetic factors capable of influence people's wellbeing or disease risk is known as Gene – Environment Interaction (GxE) and it is a key feature that allows us to adapt to our surrounding.

The data presented in this thesis will try to address some of the well characterised polymorphic variants associated with Central Nervous System (CNS) conditions, such as the Monoamine oxidase A (*MAOA*) gene, and I will show how they can modify gene expression in response to environmental stimuli. We also report two regulatory regions in the *CACNA1C* (Calcium Voltage-Gated Channel Subunit Alpha1 C) gene, strongly associated to schizophrenia by GWAS (Genome-Wide Association Study) investigations. Finally I will also report a novel polymorphic microsatellite in the promoter region of the gene that has been defined "the master

regulator" of transcription, the RE1-Silencing Transcription factor (*REST*) gene, that strongly suggests an association with Alzheimer's disease.

Therefore I demonstrate a similarity in mechanisms and in the activity of these repeat elements in the promoter regions of three key genes for CNS behaviour and illustrate the potential power of these elements as transcriptional regulatory DNA regions.

Chapter 1

General Introduction

1.1 Thesis introduction

During the early 60s the genomic loci that did not code for protein were defined by the scientific community as “junk DNA”. The term had been made public by Dr. Susumu Ohno in 1972 (Ohno, 1972). His idea was that duplications and mutations in various genes remained silent in our genome with the unique function of spacing the coding genetic regions.

In more recent years, the international Encyclopaedia of DNA Elements (Encode) recognised that at least 80% of the human genome has a biochemical function (Encode Project Consortium, 2012). It is now known that non-coding DNA is involved in epigenetic mechanisms, or constitutes cis- and trans- regulatory sequences. Part of the non-coding DNA are telomeres, introns and retrotransposons or constitute the binding sequences for transcription factors (TFs).

However, one of the most important aspects of the non-coding DNA, which I will also focus on in this thesis, is their potential role in the regulation of gene expression. Different variants of non-coding DNA sequences have been associated with variations in gene expression both biochemical and in genetic association.

In Chapters 3 and 4 I will focus on the Monoamine Oxidase A (*MAOA*) gene and the two polymorphic variants termed uVNTR (Sabol *et al.*, 1998) and dVNTR (Philibert *et al.*, 2011) and their role on *MAOA* expression.

The *MAOA* gene has been proven to be a key gene, especially through the genetic associations of the uVNTR, in several stages of the human life and in several conditions (**Table 1.3**).

This *MAOA* polymorphism has been also shown to moderate the infant emotions in relation to the mother’s stress during pregnancy (Hill *et al.*, 2013) and at the same time it plays a role in maternal sensitivity (Pickles *et al.*, 2013).

In Chapter 5 I explore another function of the non-coding DNA: its potential to operate as positive and negative regulators of gene expression. The promoter region of the GWAS (Genome-Wide Association Study) gene associated to schizophrenia (Obermair *et al.*, 2004, Gwas Consortium Bipolar Disorder Working Group, 2011) *CACNA1C* (Calcium Voltage-Gated Channel Subunit Alpha1 C) will be the subject of this analysis. I used the UCSC web browser to study two domains in close proximity that might be polymorphic, making the overall structure of the *CACNA1C* promoter region similar to the promoter region of the *MAOA* gene. Therefore, I will investigate the polymorphic nature of these regions and, if they prove to be polymorphic, I will try to build an association with SCZ, as well as validate their activity through reporter gene constructs.

Finally in Chapter 6 I explore the role of a novel polymorphic microsatellite in the *REST* gene promoter region. REST protein levels have been found dramatically reduced in Alzheimer's disease (AD) patients' brain (Lu *et al.*, 2014) and it has been hypothesised that REST has a protective role in ageing brains. Therefore, I genotyped selected cohorts in order to test this novel microsatellite might represent a risk factor for CNS diseases and in particular AD.

1.2 Gene Regulation

1.2.1 Transcription

The regulation of gene expression represents a fundamental process for every organism whether they are prokaryotes or eukaryotes. Regulating the expression of particular genes in specific moments of these organisms' life cycles makes them able to adapt to their environment, thus increasing their chances of surviving by expressing particular proteins only when needed (Raj and van Oudenaarden, 2008). The transcription process starts when the complex of general transcription factors (GTFs) binds to a particular DNA sequence, referred to as 'core' promoter, forming the pre-initiation complex (PIC) (Kornberg, 2005). In eukaryotes there are several GTFs required in order to initiate the transcription. These are TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Lee and Young, 2000, Orphanides *et al.*, 1996, Sainsbury *et al.*, 2015) (**Figure 1.1**).

After the PIC assembles, RNA polymerase II (Pol II) binds to it to initiate the transcription. This process can be additionally regulated by other proteins such as repressors or activators which are able to modulate the formation or the activity of the initiation complex. There is evidence of a mechanism in which GTFs are recruited by enhancer sequences before the transcription is initiated. Therefore, an enhancer-GTFs complex would activate the transcription by creating a loop in the DNA, closing the gap between the core promoter sequence and the enhancer-GFTs complex and only afterwards recruiting the Pol II (Szutorisz *et al.*, 2005).

The Pol II, once anchored to the DNA core promoter, creates a transcription bubble, separating the two DNA strands by melting the hydrogen bonds between nucleotides. At this point the Pol II can start the elongation process adding RNA nucleotides to the complementary DNA strand (Fuda *et al.*, 2009).

Finally, transcriptional termination occurs when Pol II reaches the 3′-end of the newly generated transcript where there are present polyadenylation signals which contain several cleavage components that allow the transcriptional machinery to terminate the mRNA (Lykke-Andersen and Jensen, 2007).

The newly formed pre-mRNA can go through a series of post-transcriptional modifications, including 5′- capping and/or splicing processes before the mature mRNA is translated into protein.

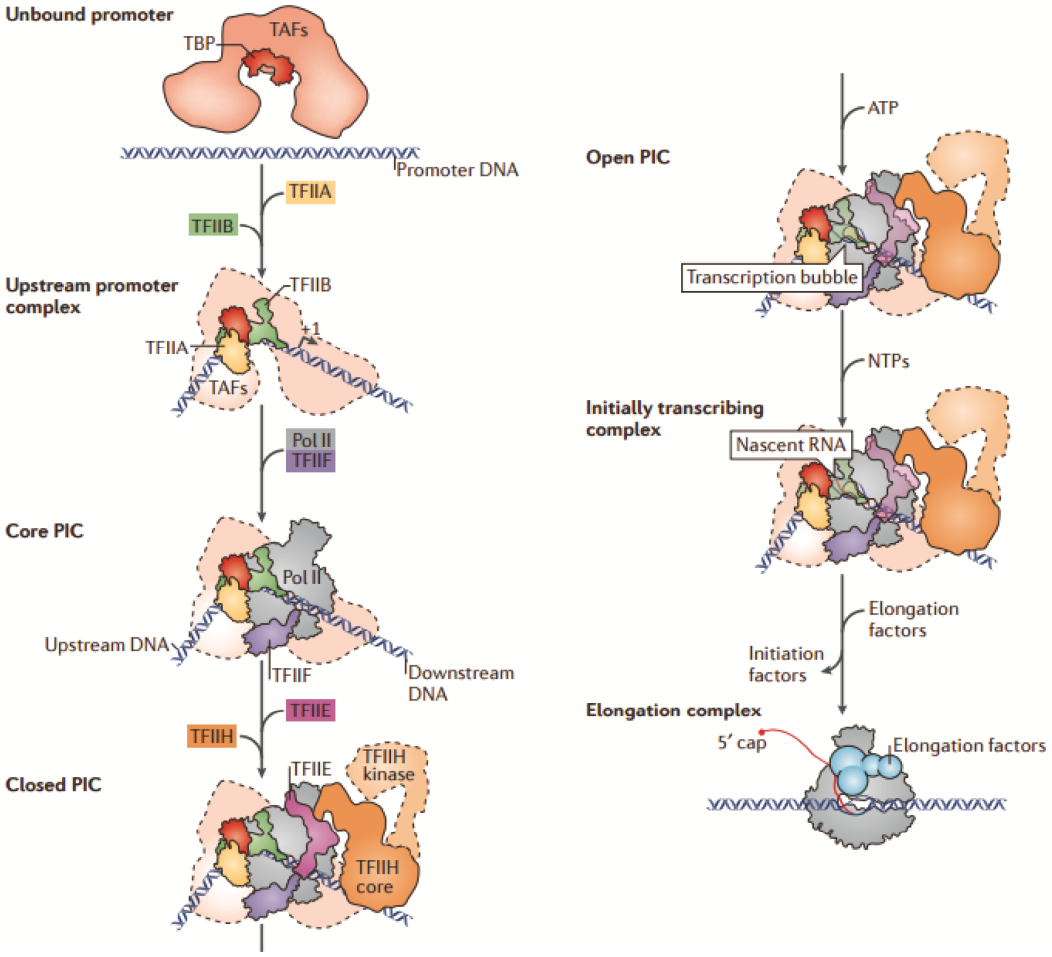


Figure 1.1 - Schematic representation of Pol II transcription initiation. Canonical model for pre-initiation complex (PIC) assembly from general transcription factors (various colours) and RNA polymerase II (Pol II; grey) on promoter DNA. The names for the intermediate complexes that form during the initiation–elongation transition are provided to the left of the images. TFIID or its TATA box-binding protein (TBP) subunit binds to promoter DNA, inducing a bend. The TBP–DNA complex is then stabilised by TFIIB and TFIIA, which flank TBP on both sides. The resulting upstream promoter complex is joined by the Pol II–TFIIF complex, leading to the formation of the core PIC. Subsequent binding of TFIIE and TFIIH complete the PIC. In the presence of ATP, the DNA is opened (forming the 'transcription bubble') and RNA synthesis commences. Finally, dissociation of initiation factors enables the formation of the Pol II elongation complex, which is associated with transcription elongation factors (blue). Image from (Sainsbury *et al.*, 2015)

1.2.2 Chromatin organization

In eukaryotes, all the genetic information for a whole organism is stored inside the nucleus of each cell. In order to fit the entire DNA length inside it, DNA is densely packed into a protein-DNA complex, referred to as chromatin.

Chromatin has several functions apart from reducing the volume of DNA inside the cells. One of the most important is regulating gene expression by different and coordinated folding and unfolding of chromatin structure organizations. There are several degrees of possible organization for the genetic material (**Figure 1.2**). The nucleosome is a fundamental, repeating unit composing the chromatin. The DNA wraps around a histone octamer composed of two H2A histones, two H2B, two H3 and two H4. One after another, several nucleosomes, form the “beads-on-a-string” structure, which is a lightly packed structure that modulates transcription processes. In order to make this structure available for transcription, two principal and reversible mechanisms take place. Histones can be modified by enzymes that can add phosphate, methyl, or acetyl groups (Fischle *et al.*, 2003). Alternatively, histones can also be displaced, allowing the transcription machinery to access DNA sequences and start the transcription (Smith and Peterson, 2005).

Chromatin is otherwise packed in more dense structures. For example, the 30nm fibres contain the linker histone H1 that makes this structure highly stable (de la Serna and Imbalzano, 2002). During meiosis and mitosis, the 30 nm fibres are further packed in the 300nm fibres that looped in themselves give rise to the characteristic chromosome (Felsenfeld and Groudine, 2003).

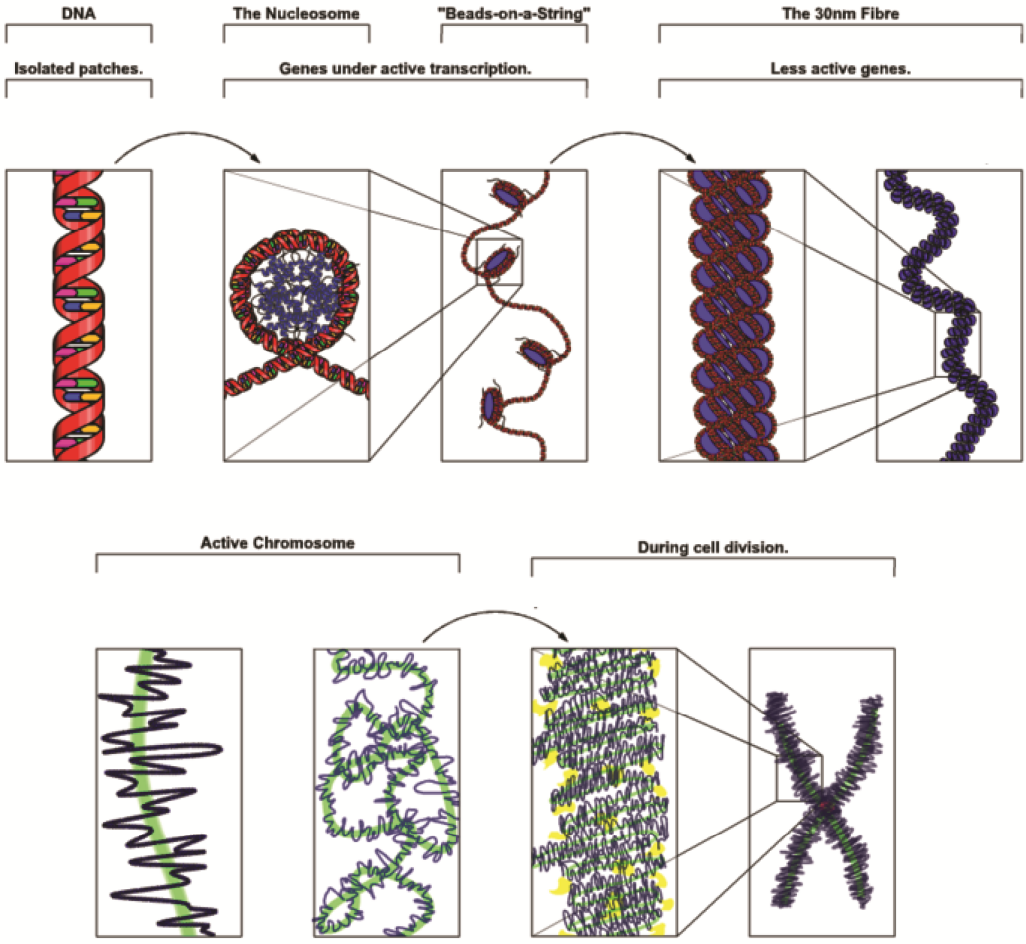


Figure 1.2 - Chromatin organization. Representation of DNA major structures. From top left: DNA double helix, single nucleosome, "beads on a string" structure (euchromatin), 30 nm fibre nucleosome (heterochromatin), section of a chromosome, condensed section of a chromosome, chromosome. Image adapted from: https://en.wikipedia.org/wiki/Chromatin#/media/File:Chromatin_Structures.png

1.2.3 Epigenetic modifications

The histone modifications, briefly mentioned in *section 1.2.2*, are part of a vast category of cellular variations that can respond to external environmental factors and are capable of influencing gene expression (Riffo-Campos *et al.*, 2015). Together, these transitory mechanisms take the name of epigenetic modifications. In general, epigenetic modifications can be defined as heritable changes in gene expression that do not involve alterations to the DNA sequences but that can switch genes on or off, modulating the transcribed protein levels. Thus enabling changing in the phenotype without altering the genotype. Age, lifestyle, disease and environment are all factors that can influence epigenetic modifications.

Specific epigenetic mechanisms include imprinting: where certain genes are expressed in a parent-of-origin-specific manner, X chromosome inactivation: the process where one of the two X chromosomes is inactivated in female mammal cells. However, three major systems are able, through interacting within themselves, to modulate gene expression: DNA methylation, histone modifications and RNA silencing (Egger *et al.*, 2004).

DNA methylation in the C^5 DNA residues has been long proven to be an epigenetic silencing mechanism (Riggs, 1975, Riggs, 2002). These cytosine residues, that can be methylated by DNA methyltransferase (DNMT) enzymes, assemble in genomic regions, usually longer than 500 bps in size, where the GC content is above 55%, referred to as CpG islands (Takai and Jones, 2002). The methylation of CpG islands in gene promoter regions is often associated with gene silencing (Phillips, 2008) and altered methylation patterns have been associated with cancer development (Luczak and Jagodzinski, 2006). CpG methylation prevents TFs and Pol II from bind the DNA, hence silencing the gene. Alternatively, methylated DNA

binding proteins, such as m5CpG binding (MeCP) and m5CpG-binding domain (MBD) proteins, can prevent the transcription machinery from binding DNA by creating spatial hindrance (Luczak and Jagodzinski, 2006).

Hypermethylation of tumour suppressive gene (TSG) promoters and hypomethylation of proto-oncogenes have both been observed during cancer development and/or metastasis (Herman and Baylin, 2003, Szyf *et al.*, 2004).

Altered methylation patterns have also been observed in relation to neurological and psychiatric disorders. An example is the hereditary sensory and autonomic neuropathy type 1 (HSAN1) condition where a mutation in the DNA (cytosine-5)-methyltransferase 1 (DNMT1) gene interferes with DNA methylation maintenance causing dementia and hearing loss in adulthood (Klein *et al.*, 2011). Another example is Fragile X Syndrome (FXS) where hypermethylation of a trinucleotide repeat expansion in the *FMR1* (fragile X mental retardation 1) gene on the X chromosome silences the transcription, causing a form of mental retardation (Devys *et al.*, 1993).

In close coordination, proteins that bind methylated DNA also interact with complexes involved in histone modifications. Histone deacetylase remove acetyl groups from lysine residues in histone tails. As a result the histones wrap DNA more tightly, silencing transcription. For example, DNMT1 has been proven to interact with the histone deacetylase HDAC1, promoting gene silencing (Fuks *et al.*, 2000). Opposite effects are produced by histone acetylases that, adding an acetyl group, make the DNA more available for transcription (Choudhary *et al.*, 2009).

Antisense non-coding RNAs, microRNAs (miRNA) and RNA interference (RNAi) have also the ability to regulate gene expression by silencing transcription through the formation of heterochromatin (Mattick and Makunin, 2005). A well

characterised example is the product of the *XIST* gene that in mammals is responsible for dosage compensation in females by silencing one of the X Chromosomes (Plath *et al.*, 2002). In addition, it has been proven that altered miRNA pathways are present in several age-dependent neurodegenerative conditions (Gascon and Gao, 2012).

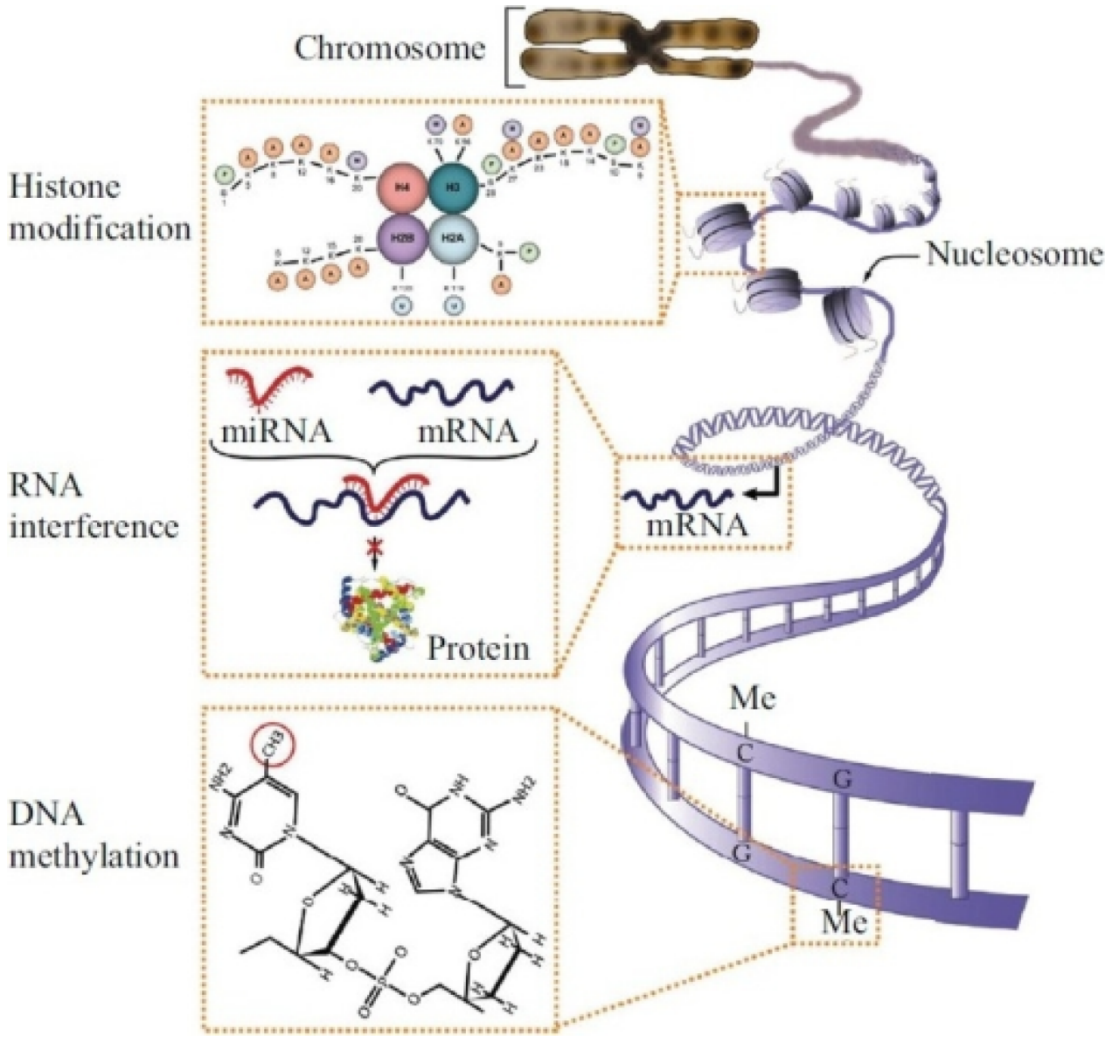


Figure 1.3 - Epigenetic modifications. Schematic of the mechanisms of the three distinct mechanisms of epigenetic regulation. Histone modifications, RNA-mediated gene silencing, and DNA methylation. Image from: Sawan *et al.* (2008)

1.3 Genetic polymorphisms

Within individual and population gene pools are present different alleles of each genetic region. These genetic variations are supported by individuals carrying the different alleles variation which are caused by differences in the nucleotide sequences and usually arise through mutations. Some examples of different alleles present in this thesis are the *MAOA* VNTR, where 6 different alleles of the 30bp repeat are present among the population, and the *REST* VNTR where at least 4 variants have been identified. These different alleles, but also many others, are capable of sustain differential expression depending on the length of the allele.

Random mutations are the source of genetic variation and are defined as permanent changes in an organism's genome. Through genetic mutations some viruses become more infective, more lethal and through the same process the immune systems are capable of defending the host from external challenges. Mutations are, however, rare events and most likely to be neutral or deleterious. Advantageous mutations however can be favoured by natural selection. Genetic variation can be categorised according to the size of the mutation, less then 1Kbp include single nucleotide polymorphisms (SNP) where a single nucleotide is modified, insertions or deletions where a larger genetic region can be deleted or inserted into the genome. Above 1Kbp there are duplications, copy number variants (CNVs) or translocations. Each of these mutations exert a different effect on the genome and, therefore, to all the related downstream pathways.

These genomic variants are often considered important genetic biomarkers to identify the susceptibility to several disorders. Among the multitude of common variants with low risk, there are some rare polymorphisms with moderate to high risk factor.

Great emphasis has been put into SNP analysis for mainly two reasons: firstly if these polymorphisms are in fact in the protein coding regions, a mutation can produce an aberrant protein that lacks function, producing a clear and easily identifiable phenotype in patients. Secondly, these single nucleotide substitutions are largely used to compare individual genotype frequencies as SNPs can be inherited in blocks, thus easily linked to a particular phenotype or disease. This feature has been exploited to identify genetic loci which may give insight on disease risk factors among the population. Linkage disequilibrium (LD) maps are a way by which the inheritance of these genetic biomarkers can be assessed or predicted. The vast majority of these single nucleotide mutations associated to a disease are located either in intronic or intergenic regions, and thus are non-coding (Hindorff *et al.*, 2009, MacKenzie *et al.*, 2013).

In recent years, the scientific community has increasingly given importance to genomic regions that are defined as Tandem Repeats (TRs) or Simple Sequence Repeats (SSRs) (Breen *et al.*, 2008). Along with a better understanding of the genomic structures and genetic regulatory mechanisms, microsatellites and short tandem repeats gained more importance in disease onsets.

In 2001 the international human genome sequencing consortium (Lander *et al.*, 2001) estimated that at least 3% of the human genome contains DNA tandem repeats and microsatellites, three times more than the protein coding sequences (~1%) in the human genome. These sequences are categorised by their repeat unit length: microsatellites are genomic region containing repetitions of 2-5bp (base pairs) or nucleotides and can be repeated 5-50 times. Longer motif lengths go under the name of minisatellites, however they are both part of the category referred to as variable number tandem repeats (VNTRs). Although their location can widely vary, they are

particularly abundant in gene promoter regions usually around 1kb upstream of the transcriptional start sites (Payseur *et al.*, 2011). VNTR structures and sequences, often offer high affinity and specific binding motifs to transcriptional factors (TFs), essential for either initiating or to simply regulate the transcription of a gene (Breen *et al.*, 2008).

Taken together, the VNTRs ability to bind TFs and the fact that their variable copy number can bind multiple copies of the same TF support their importance in modulating gene expression. These polymorphisms can affect a wide variety of factors. **Figure 1.4** summarises some of these possible effects, depending on the position they are located in the genome and their effects on downstream pathways.

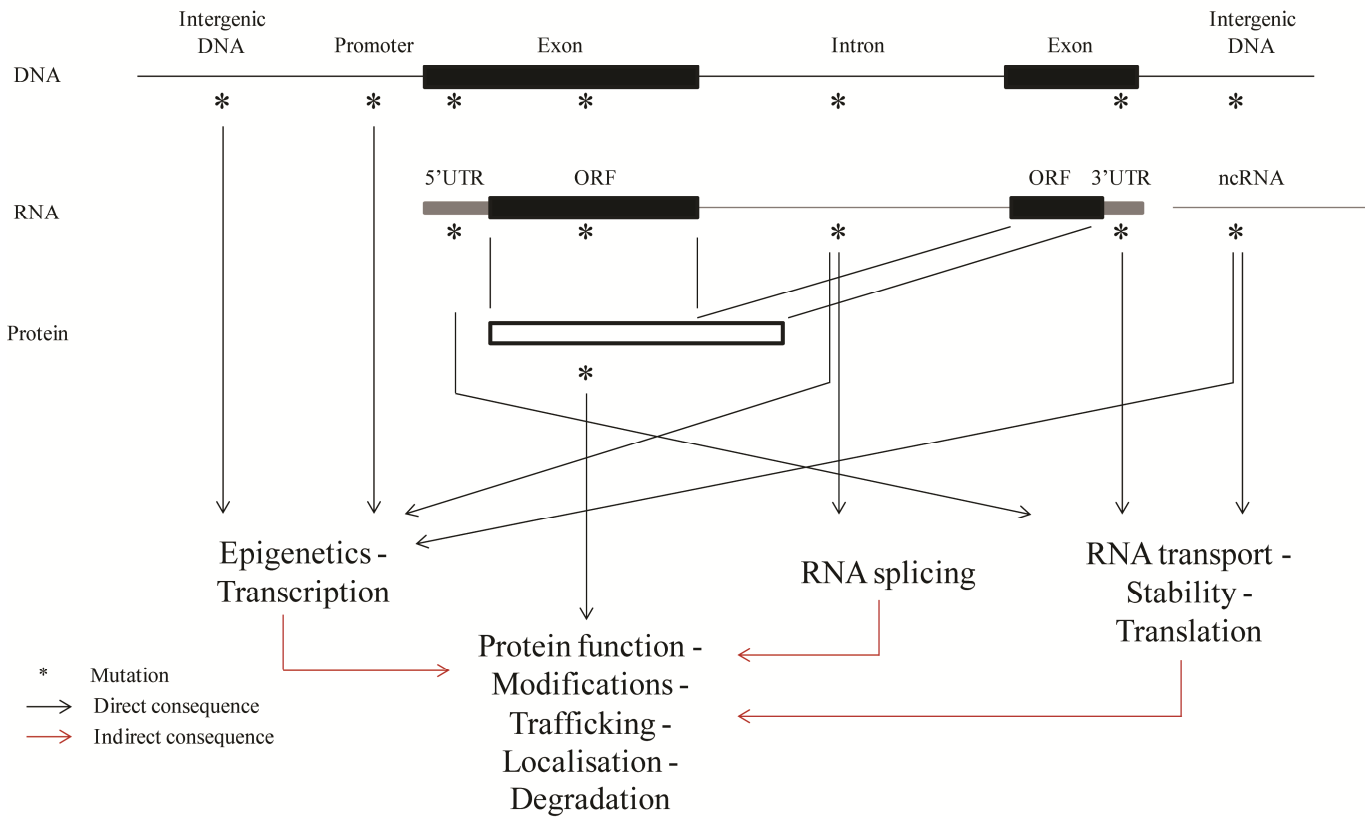


Figure 1.4 - Genetic polymorphisms and their potential effects. The positions of the polymorphisms are indicated by the stars. When located in the intergenic or promoter regions they can modulate the expression acting as trans or cis regulatory elements respectively. Within the transcribed region they can affect the translated protein sequence and therefore its activity. DNA, RNA and protein levels can be affected by these polymorphisms. Abbreviations: ncRNA, non-coding RNA; ORF, open reading frame; UTR, untranslated region. Figure adapted from (Hannan, 2010).

1.4 Variable Number Tandem Repeats (VNTRs)

Variable Number Tandem Repeats are genomic DNA sequences that can be repeated several times. One of the primary causes of mutation of these VNTRs can be accounted for by a mechanism termed replication slippage, consisting of a misalignment association during the DNA replication process (King *et al.*, 1997).

Probably the most important characteristic of VNTRs is their ability to regulate gene expression depending on the motif length. In addition, and because of this peculiarity, they have been associated to several human behavioural traits or CNS disorders. **Table 1.1** summarises some of the most characterised VNTRs in genes such as those in the serotonin transporter (*SLC6A4*) and the dopamine transporter *DAT1* (*SLC6A3*), for association to a specific disorder. The *SLC6A4* repeat length polymorphism **(**5-HTTLPR), one of the most characterised VNTRs, is located upstream of the coding region of *SLC6A4*. Of the several variants present in the population, the most common alleles contain 14 and 16 repetitions of a GC-rich sequence (Nakamura *et al.*, 2000). The shorter allele, with 14 repetitions, has been associated with lower gene expression *in-vitro* (Heils *et al.*, 1996). Association analyses have also linked the short VNTR with more severe depressive symptoms compared to the long allele carriers (Goldman *et al.*, 2010).

Another well characterised VNTR is a 30 base pairs VNTR, termed uVNTR, in the promoter region of the MAOA gene (Sabol *et al.*, 1998) and 10 base pairs VNTR, termed the dVNTR, 500 base pairs 5' from the uVNTR (Philibert *et al.*, 2011). In *section 1.8* I will talk more extensively about these VNTRs.

Table 1.1 - VNTR behavioural association

Gene	Motif (Repetitions)	Location	Association	Reference
Serotonin Transporter (*SLC6A4*), 5-HTTLPR	22 bp (14 allelic variants)	Promoter Region	Anxiety-related traits	(Sen *et al.*, 2004)
Serotonin Transporter (*SLC6A4*) STin2	17 bp (9R - 10R - 12R)	Intron	Susceptibility to bipolar disorder	(Lasky-Su *et al.*, 2005)
DNA-binding protein Jarid2	4 bp (8R to 13R)	Coding	Schizophrenia	(Pedrosa *et al.*, 2007)
Dopamine Transporter *DAT1* (*SLC6A3*)	40 bp (3R to 11R)	3′ UTR	Attention deficit hyperactivity disorder	(Cornish *et al.*, 2005)

Table adapted from Fondon *et al.* (2008)

1.5 Microsatellites or Simple Sequence Repeats (SSRs)

Microsatellites or Simple Sequence Repeats (SSRs) are short genomic sequences with a 2-5 nucleotide motifs repeated several times. They have been shown to be represented across the genome in a non-random distribution (Li *et al.*, 2007) and they have also been implicated in several processes such as chromatin organization and gene regulation (Gymrek *et al.*, 2016, Li *et al.*, 2007).

In humans, several neurological disorders have been associated with SSR expansions that frequently, as in the case of Huntington's disease (HD), disrupt the protein activity by modifying the translated sequence when the expansion is located within a coding exon. The normal huntingtin (htt) protein contains a run of 6–35 glutamines encoded by CAG repeats. Expansion of this glutamine run causes misfolding and/or abnormal cleavage of the resulting protein (Gusella and MacDonald, 2006). The altered protein is, in turn, responsible for progressive cell dysfunction and neuronal death which causes the symptoms of the pathology, such as cognitive deficits and dementia.

In the case of the Fragile X mental retardation syndrome (FRAXA), an expansion located in the 5′ UTR of the gene causes a dramatic reduction in gene expression. **Table 1.2** highlights some of these expansions where the disease outcome can be predicted by the length of the VNTR.

However, despite the fact that most of these expansions coincide with severe diseases, it has been demonstrated that the vast amount of genetic variability that arise from all the VNTRs, especially the SSRs, present in the genome may play an important role in evolution contributing to a wide morphological and phenotypic variation within species (Fondon and Garner, 2004, Kashi and King, 2006).

Table 1.2 - VNTR disease association

Disease	Motif	Expansion Length		Location	Reference
		Normal	Pathology		
Huntington's disease (HD)	CAG	6 – 35	36 – 250	Exon	(Gusella and MacDonald, 2006)
Fragile X mental retardation syndrome (FRAXA)	CGG	6 – 53	>200	5′ UTR	(Garber *et al.*, 2006)
Fragile X tremor/ataxia syndrome (FXTAS)	CGG	6 – 53	55 – 200	5′ UTR	(Loesch *et al.*, 2007)
Spinal and bulbar muscular atrophy (SBMA)	CAG	9 – 36	38 – 65	Exon	(Orr and Zoghbi, 2007)
Amyotrophic Lateral Sclerosis (ALS) (C9ORF72 gene)	GGGGCC	< 20	> 30	Intron	(DeJesus-Hernandez *et al.*, 2011, Majounie *et al.*, 2012, Renton *et al.*, 2011)

Table adapted from Fondon *et al.* (2008)

1.6 Alternative structures of repetitive elements: G-Quadruplex

In addition to the well known DNA double helix conformation there are other more unusual structures, where the organisation and stability of these alternative structures is highly dependent on nucleic acid composition and chemical properties.

One of these alternative DNA conformations takes the name of G-Quadruplex. As the name suggests, high G-rich DNA or RNA strands could be rearranged into a more stable conformation depending on the length of the sequence (Huppert, 2010). The predicted core structure of G-Quadruplexes contains four runs of guanines also known as a G-tract, each of which contains at least three guanines, separated by any other nucleotides ($G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$) (Bochman *et al.*, 2012).

Furthermore, cations such as K^+ or Na^+, have been found to intercalate in the G-Quadruplex structure stabilising it (Dapic *et al.*, 2003).

Initially thought to be a purely *in-vitro* artefact, these structures have been found in telomeres (Paeschke *et al.*, 2005) and in near proximity to TSSs possibly able to modulate transcription (Huppert, 2008, Huppert *et al.*, 2008). To highlight the importance of these G-Quadruplex structures have been demonstrated to be highly evolutionary conserved genetic regions even among yeast (Capra *et al.*, 2010).

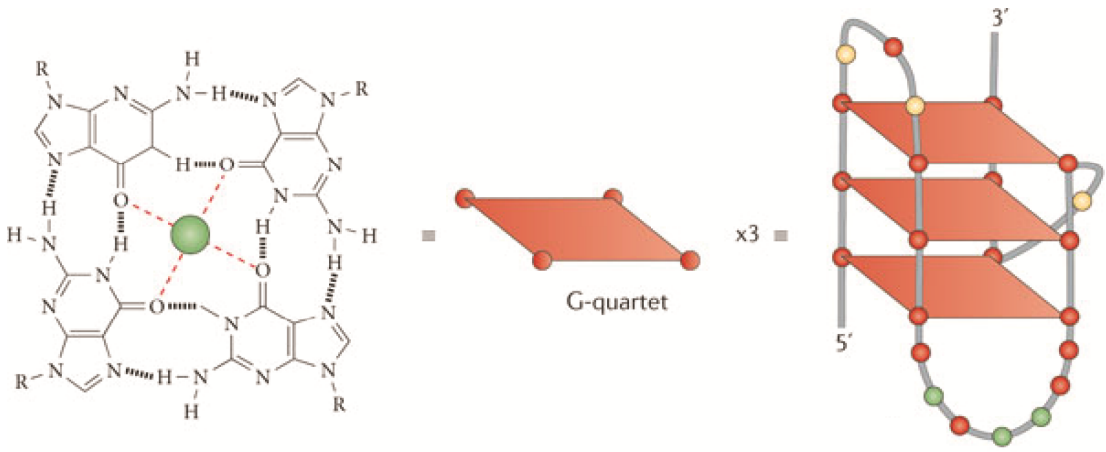


Figure 1.5 - G-Quadruplex. G-quadruplex structures are built from a core of two or more G-quartets; on the left the atomic structure of a G-quartet. Image from: (Balasubramanian *et al.*, 2011)

1.7 Gene environment interaction (G x E)

Different variants of the same gene or genetic regions shape a substantial amount of people's phenotypes. This is also true for diseases. The initiation and/or progression of these diseases and the increase of the risk or protective factors that a specific genetic variant can give. A perfect example is Huntington's disease where a genetic expansion provides a non functional protein (Warby *et al.*, 1993). However, for some conditions, genetic variations are not sufficient to explain the disease or condition onsets and often are not consistent with the Mendelian modes of inheritance within families. In addition to genetic factors, that can be considered risk genotypes for a certain condition, a broad range of environmental components, whether they are biological (e.g., infections), chemical (e.g., oxy radicals), physical (e.g., radiation), behavioural (e.g., smoking) or even life events (e.g., injury, job loss), are able to impact on the chances of disease development in people.

What has become more important in the last decades is the fact that the interplay between genetic and environmental factors is something that might explain the different outsets of some conditions in people with the same genotype but exposed to different environments.

This combined effect of environmental factors and genetic factors capable of influence people's wellbeing is known as Gene – Environment Interaction (GxE) (Moffitt *et al.*, 2005).

It has been extensively demonstrated that polymorphic genetic variants can be differentially influenced by environmental stressors and act independently on gene expression, therefore they can be considered predisposing factors in neuropsychiatric disorders or inappropriate behavioural responses (Quinn *et al.*, 2013). The *MAOA* gene, in this theses extensively discussed in chapters 3 and 4, is one of the most cited

example of gene-environment interaction. The previous defined low- and high-activity variants of the *MAOA* VNTR (uVNTR: 2, 5 repetitions and 3, 3.5 and 4 repetitions respectively), have been linked to several CNS conditions through case-control studies (**Table 1.4**).

Furthermore, environmental factors have also the potential to modify the cells' ability to adapt through epigenetic processes.

DNA methylation, histone modifications and X chromosome inactivation are some of the epigenetic mechanisms that cells use to regulate gene expression in order to develop and to adapt to the environment that surrounds them.

These epigenetic modifications are essential for normal development but they are easily influenced by environmental factors such as nutrition, stress or drugs that can modify in the medium or long term, or even permanently, gene expression.

For example, studies in both human and rodents proved that early life experiences can modify the epigenetic signatures that in turn can influence behaviour in adulthood which may also arise to neurological conditions depending on environmental stressors exposition (Hill *et al.*, 2013, Murgatroyd and Spengler, 2011, Murgatroyd *et al.*, 2015).

1.8 The Monoamine Oxidase A (MAOA) gene promoter VNTRs

Monoamine oxidases (MAO) are enzymes located in the outer mitochondrial membranes, in part responsible for the turnover of some of the most important neurotransmitters. Serotonin (5HT), dopamine (DA) and noradrenaline (NA) are, along with the dietary monoamines such as tryptamine and phenylethylamine, oxidised by these enzymes. The *MAOA* and the monoamine oxidase B (*MAOB*) genes have been mapped in the X Chromosome at 43.52 – 43.61 Mb and 43.63 – 43.74 Mb respectively and exhibit a tail-to-tail configuration beside the same intron-exon organisation (Chen *et al.*, 1992, Grimsby *et al.*, 1991). MAOs catalyse the oxidative deamination of monoamines acting in concert with the covalently bound cofactor flavin adenine dinucleotide (FAD). According to Chen *et al.* (1991) the FAD covalent binding site appears to be located in *MAOA* exon XII (Cys 406). More recent crystallography studies (De Colibus *et al.*, 2005) identified the FAD binding domain in residues 13–88, 220–294, and 400–462. This requires extensive protein folding in order for the FAD to correctly interact with MAOA and produce a functional enzymatic activity. The substrate-binding domain resides in residues 89–219 and 295–399 and the intramembrane region is located in the C-terminal of the protein in residues 463–506.

In the oxidative reaction, an oxygen atom is used to remove an amine group from a substrate molecule, resulting in the corresponding aldehyde with the production of an ammonia molecule (Gaweska and Fitzpatrick, 2011). MAOA is afterward reoxidised by molecular oxygen, which forms H_2O_2 in the process (**Figure 1.6**).

Human MAOA has been proven to crystallise as a monomer while the rat MAOA (92% of homology) and human MAOB (72% of homology) crystallise as dimers (De Colibus *et al.*, 2005).

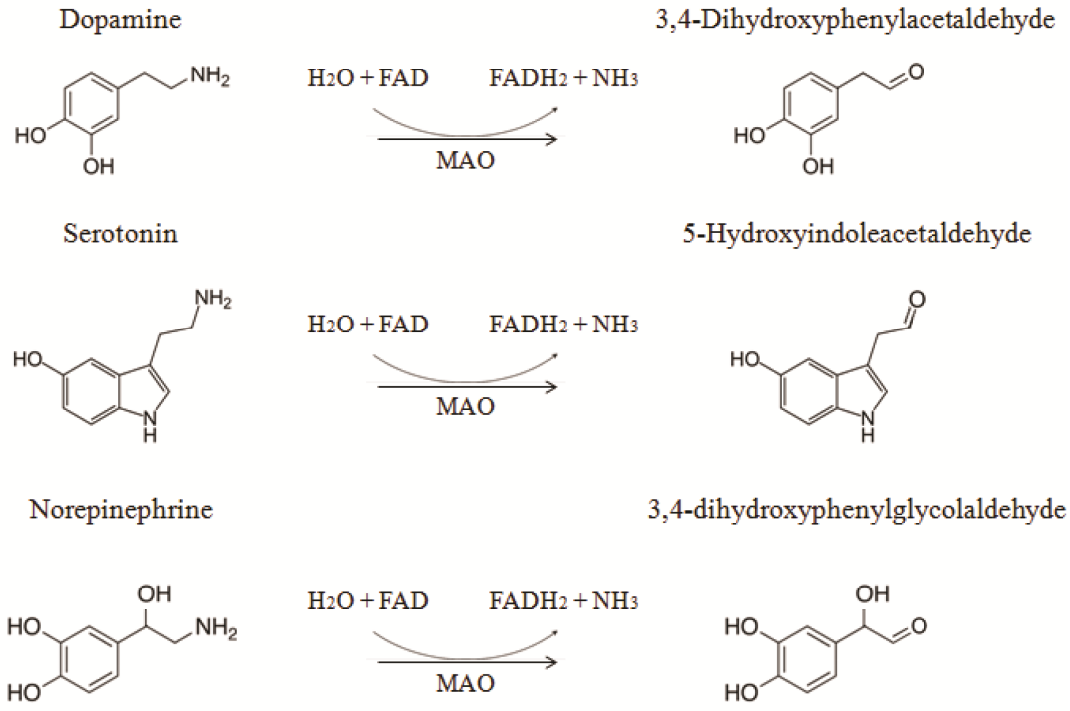


Figure 1.6 - Oxidation of monoamines catalysed by Monoamino Oxidases. Simplification of the oxidative reaction mediated by MAOA on the major monoamine neurotransmitters.

A single amino acid mutation has been found to be a possible explanation on the monomeric human MAOA state. It is a specific Glu 151 → Lys mutation uniquely found in human MAOA. This residue in located in a specifically charged region involved in the dimerisation process (Andres *et al.*, 2004).

The *MAOA* promoter region lacks a TATA box, where usually the RNA polymerase II binds in order to initiate the transcription. However the promoter is CG rich and contains multiple binding sites for SP1 transcription factor which can initiate the transcription in the absence of a TATA box (Zhu *et al.*, 1992). It has been demonstrated that *MAOA* can be positively regulated by both androgens and glucocorticoids acting through SP1 and that glucocorticoids are more powerful activators than androgens (Ou *et al.*, 2006). PET (Positron emission tomography) scan images show that MAOA activity appears to be higher in depressed patients (Meyer *et al.*, 2006), as well as mean cortisol concentrations which are generally higher than in normal control subjects (Barden, 2004). Furthermore, Manoli *et al.* (2005) found that *in-vitro* treatment with dexamethasone, a synthetic member of the glucocorticoid class of steroid drugs that is 25 times more potent than cortisol in its glucocorticoid effects, increases *MAOA* gene expression. In addition, two different VNTRs have been identified and proven able to drive expression in reporter gene constructs. The most studied is termed uVNTR and consists of a 30 base pair motif that can be repeated 2, 3, 3.5, 4 and 5 times. In 2004 Huang *et al.* (2004) described a 6 repetitions (R) variant of this same polymorphism. The 2, 3 and 5 uVNTR are defined low expression variants (MAOA-L) whilst, conversely, with a 2 to 10 fold increase in the expression, the 3.5 and 4R VNTRs are considered high expression variants (MAOA-H) (Sabol *et al.*, 1998) (**Figure 1.7 A**). The most common variants in the population are represented by the 3R and the 4R variants.

Table 1.3 - Count of alleles and polymorphism for MAOA u-VNTR in different/ethnic populations

Populations	Sample Size	2R (%)	3R (%)	3.5R (%)	4R (%)	5R (%)
Han in Taiwan (Pai *et al.*, 2007)	474	1 (0,2)	307 (64,7)		165 (34,8)	1 (0,2)
Chinese Han in Taiwan (Lu *et al.*, 2002)	77	1 (1,3)	42 (54,5)		34 (44,2)	
Japanese (Kunugi *et al.*, 1999)	125		78 (62)		47 (38)	
Asian/Pacific Islander (Sabol *et al.*, 1998)	82		50 (61)	1 (1,2)	31 (37,8)	
White/Non-Hispanic (Sabol *et al.*, 1998)	1612		539 (33,1)	8 (0,5)	1056 (64,8)	26 (1,6)

Note: Table adapted from Pai *et al.* (2007)

As illustrated in **Table 1.3** (Pai *et al.*, 2007), the trend between the most common alleles (4R and 3R) is opposite in western and eastern populations. Pai *et al.* (2007) in their study draw attention to the fact that *MAOA* uVNTR per se could be implicated in MAOA transcription and activity but may not be critical in modulating behavioural abnormalities. This is possibly due to the racial/ethnic differences in the genotype frequencies among world populations.

In earlier versions of UCSC genome browser the uVNTR was located 1.2Kb upstream of the TSS, showing only one human splicing variant containing the uVNTR within the *MAOA* promoter region. At the time of writing this thesis, the latest version of the Hg19 database in UCSC genome browser, moved the *MAOA* TSS of the reference allele upstream of the uVNTR, ultimately including it into the transcript and making it the only variant. However, at the same time, the new Hg38 database in UCSC genome browser and other online databases, such as AceView, kept separated the two TSSs of *MAOA* gene, describing two isoforms with two different 5′ UTRs. More details about the *MAOA* mRNA isoforms will be discussed in Chapter 4.

The second VNTR in the *MAOA* promoter region, located ~500bp upstream of the uVNTR, is composed by two different decamers CCCCTCCCCG (Repeat A) and CTCCTCCCCG (Repeat B) that alternate themselves up to 7R (ABABABA); after that only the repeat A is present (ABABABAAAA). The 9 and 10 repetitions are the most common in the population, 8R, 11R and 12R have also been genotyped (**Figure 1.7 B**). In gene reporter assays, similarly to the uVNTR, the 9R and 10R have opposite potency in driving the reporter gene expression, where the 9R is stronger than the 10R and the other genotypes fall in between them (Philibert *et al.*, 2011).

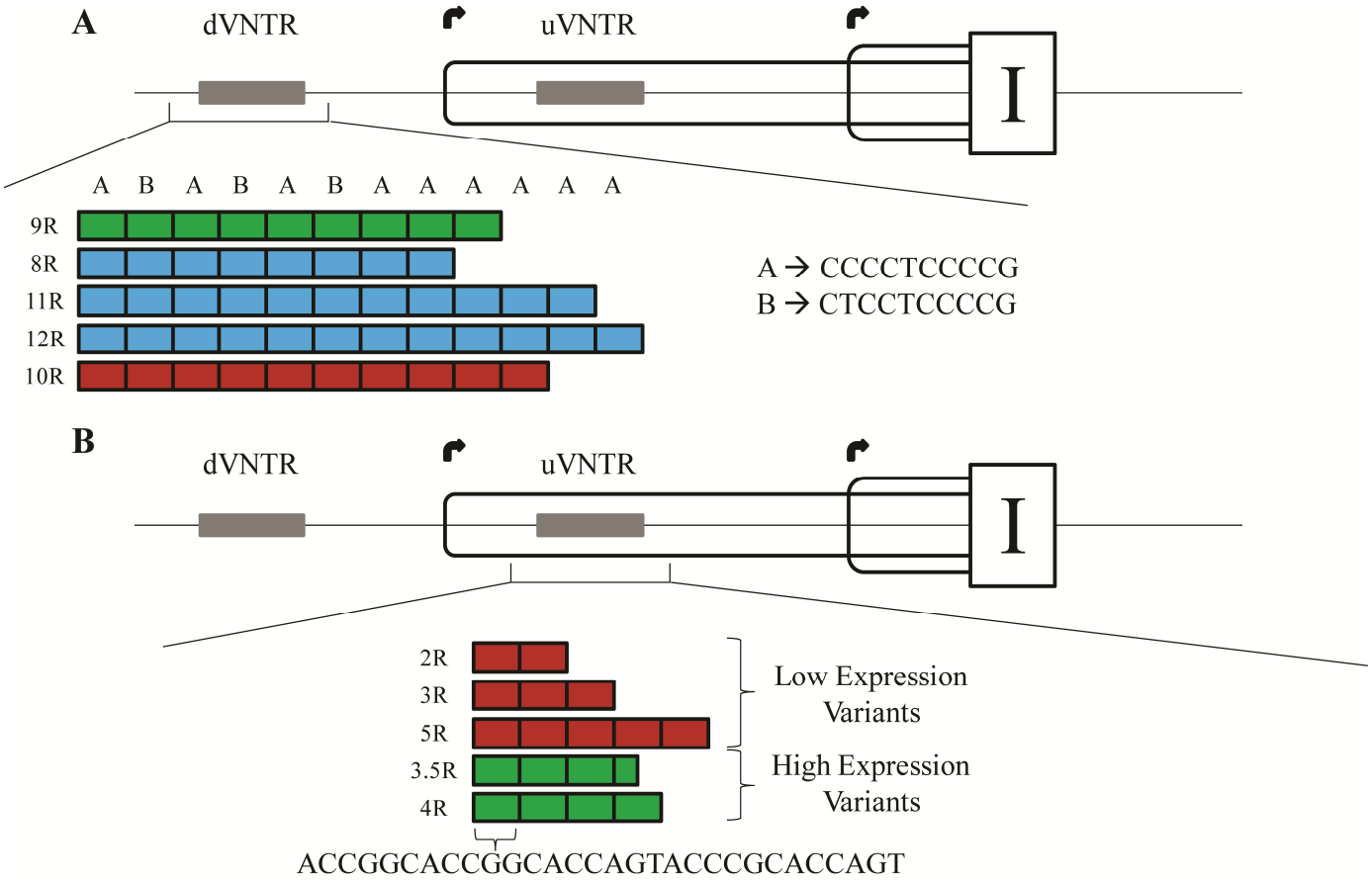


Figure 1.7 – Monoamine Oxidase A (*MAOA)* promoter region VNTRs. Graphic representation of the *MAOA* u- and dVNTR. **A**. *MAOA* dVNTR genotyped alleles: 8R to 12R. The structure forming the dVNTR is represented above the coloured squares, each representing a repetition. In green high expression, in red low expression, in blue the other copy variants as reported by Philibert *et al.* (2011). **B**. *MAOA* uVNTR genotyped alleles: 2R to 5R. The 30bp motif underneath the coloured squares. In red the low expression variants and in green the high expression variants as first reported by Sabol *et al.* (1998).

However, the vast majority of the literature refers only to the uVNTR as the major mediator of *MAOA* expression and a possible biomarker for stress-related illnesses, including major depressive disorder, addiction, and violent behaviour (Fan *et al.*, 2010, Philibert *et al.*, 2011, Reif *et al.*, 2014). For example, the high activity variant of the uVNTR has been associated with suicide attempts in female subjects with bipolar disorder (Ho *et al.*, 2000) and with suicide in depressed male subjects (Du *et al.*, 2002). Male children with the high expression variant, who experienced early life stressors, are less likely to develop antisocial behaviour compared to the children with the low activity variant (Caspi *et al.*, 2002).

Table 1.4 summarises some of the most recent or most relevant publications that associate, or try to associate, *MAOA* uVNTR to a wide range of CNS conditions and behavioural traits. Despite these efforts, the results are often inconclusive and conflicting. Limiting factors cited by the authors are often the small sample size, gender differences, ethnic differences in the case of genotype analysis or construct structure in case of the expression assays. An important factor that is usually excluded from meta-analysis and genotype analysis are female heterozygous for the uVNTR. *MAOA* is an X chromosome gene, therefore the ratio between males and females are different as different is the number of X chromosomes present in males and females. The female *MAOA* locus has been proposed as one of the loci that can escape X inactivation (Ji *et al.*, 2015, Joo *et al.*, 2014, Mugford *et al.*, 2014).

This could be consistent with the observation that *XIST,* a gene which contributes to the control of X chromosome inactivation, has recently been reported to be over expressed in major affective disorders in females (Ji *et al.*, 2015). Differential X chromosome inactivation could therefore have significant influence on the regulation of *MAOA* in females.

This further layer of complexity in the regulation of the *MAOA* gene expression in females could in part explain why many studies of the *MAOA* G×E only report on males due to uncertainties with respect to which allele is active in heterozygous uVNTR females (Caspi *et al.*, 2002, Fergusson *et al.*, 2012, Kim-Cohen *et al.*, 2006).

Studies which have included both genders have had more complex findings often showing differences in the uVNTR association with behaviour (Aslund *et al.*, 2011, Nikulina *et al.*, 2012) such as pronounced effects on ADHD and anxiety only in females whereas the opposite is true for autism, bipolar disorder and aggressive behaviour only in males (Melas *et al.*, 2013, Reif *et al.*, 2012).

Table 1.4- Summary of *MAOA* uVNTR in various conditions

Condition	Gender	Associated *MAOA* genotype	Reference
ADHD and ADHD spectrum disorders	M/F	3 - no association	(Verma *et al.*, 2014)
	M	3	(Wargelius *et al.*, 2012)
Aggression	M	3	(Chester *et al.*, 2015)
Alcohol Consumption	M/F	M Low + - F High + + = sexual abuse	(Nilsson *et al.*, 2011)
	M	4	(Tikkanen *et al.*, 2010)
Anxiety disorder	M/F	3R more present in F with anxiety disorder - 4R in F with agoraphobia	(Samochowiec *et al.*, 2004)
Autism Spectrum Disorder	M	Children 4 - mother's 4/4 more present in ADHD children	(Cohen *et al.*, 2003)

Bipolar Affective Disorder	M/F	3 in males	(Lin *et al.*, 2008)
Climacteric symptoms	F	no association	(Grochans *et al.*, 2013)
Conduct behaviours	M	3	(Caspi *et al.*, 2002, Frazzetto *et al.*, 2007)
Depression	F	3 + child adversities	(Melas *et al.*, 2013)
Major Depressive Disorder	M/F	4 [H] risk factor in F	(Schulze *et al.*, 2000)
	M/F	3 [L] increase risk through cis-acting regulation	(Zhang *et al.*, 2010)
	M/F	4 more frequent in F - F 3 /3more responsive to fluoxetine	(Yu *et al.*, 2005)
Mood disorder	M/F	no association	(Kunugi *et al.*, 1999)
Panic Disorder	M/F	4R only in females	(Reif *et al.*, 2012)
Post-Partum Depression	F	MAOA 4/4	(Grochans *et al.*, 2015)
Psychiatric Disorders	M/F	no association	(Liu *et al.*, 2015)
Schizophrenia	M/F	no association	(Alvarez *et al.*, 2010, Qiu *et al.*, 2009)
Suicide Behaviour	M/F	no association	(Hung *et al.*, 2012)

Note: Gender: M=males; F=females; Genotype: with 3 and 4 are intended the 3R and 4R genotype of the uVNTR; H and L represent the High and Low expression genotype – High are 3.5R and 4R, Low are 2R, 3R and 5R

1.9 The Calcium Voltage-Gated Channel Subunit Alpha1 C *(CACNA1C)* gene

Calcium channels are a category of ion channels selectively permeable to calcium ions. They can be divided in two categories: voltage- or ligand-gated depending on the activation methodology. The voltage-gated channels activate after a cellular depolarization while the ligand-gated channel activate after a ligand binds to the channel. The voltage-gated calcium channels are found in the membrane of excitable cells such as neurons and muscles. Of the protein complex that forms the functional voltage calcium channel, the α1 sub-unit forms the selective permeable pore. It also determines the activation speed of the channel and therefore the function that exerts. Ten different α1 sub-units have been identified, each of which is encoded by a different gene termed *CACNA1 A* to *H* and *CACNA1S* (**Figure 1.8**).

However, the auxiliary sub-units α2, β, δ and γ are necessary to stabilise and make the channel fully functional.

The *CACNA1C* gene is located on chromosome 12 and codes for the α1C sub-unit (Ca_v 1.2) of the L-type voltage-dependent calcium channel (LTCC). The Ca_v 1.2 subunit is mainly expressed in the cardiac muscle tissue and in the brain (Christel and Lee, 2012, Striessnig *et al.*, 2014). The inward calcium current that the activation of the LTTC brings, has been associated with formation of spatial memory and learning (Moosmang *et al.*, 2005, White *et al.*, 2008), neuronal differentiation, through the regulation of axonal growth and guidance (Rosenberg and Spitzer, 2011) and gene regulation through the activation of the transcription factor cAMP response element-binding protein (CREB) (Weick *et al.*, 2003, Wheeler *et al.*, 2008).

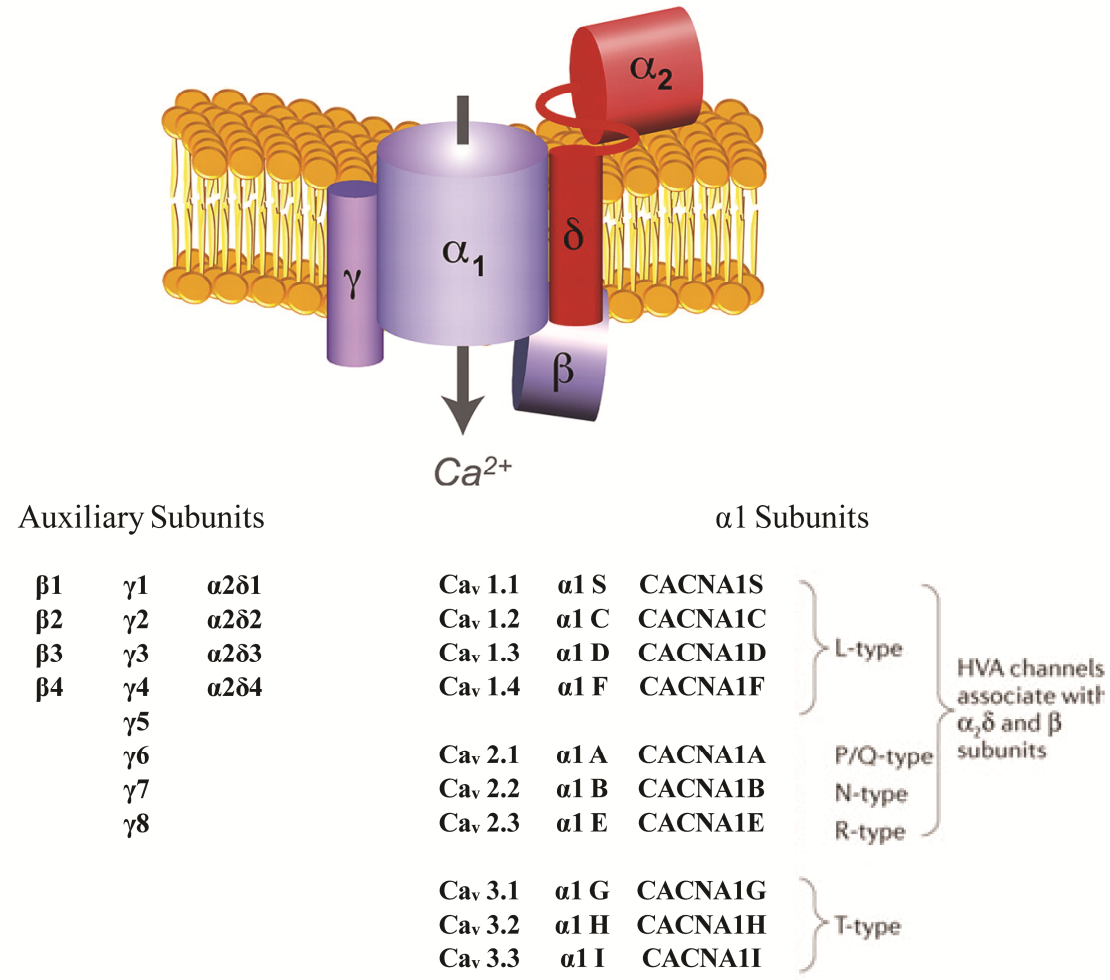


Figure 1.8 - Voltage-dependent calcium channel. Graphic representation of the high voltage-activated calcium channel complex consisting of the main pore forming α1-subunit plus auxiliary subunits β-, γ-, and α2-δ. Image adapted from (Dolphin, 2012, Khosravani and Zamponi, 2006).

CACNA1C has been considered a potential candidate gene for several neurological disorders. An intronic SNP within the *CACNA1C* gene (rs1006737) has been initially associated with bipolar disorder (Sklar *et al.*, 2008) followed by another intronic SNP (rs4765913) with an even stronger association (Gwas Consortium Bipolar Disorder Working Group, 2011).

Subsequently, *CACNA1C*, has been associated to SCZ (Bigos *et al.*, 2010) and the previous associated SNP (rs1006737), associated with bipolar disorder, has also been found as a possible risk factor for SCZ (Nyegaard *et al.*, 2010) and possible regulator of dorsolateral prefrontal cortex activity (Paulus *et al.*, 2014).

It has been proposed that in SCZ modulation of Ca^{2+}-mediated signalling caused by differential *CACNA1C* and glutamatergic transmission could lead to alterations in the rate of axon outgrowth and path finding, leading to the inefficient neuronal wiring observed in SCZ (Fromer *et al.*, 2014, Purcell *et al.*, 2014).

Based on these specific and recurrent associations of the *CACNA1C* gene with neurological disorders and specifically with SCZ I performed a preliminary bioinformatic analysis on the *CACNA1C* promoter region and I identified two putative polymorphic VNTRs according to UCSC web browser that will be discussed in chapter 5.

1.10 The RE1-Silencing Transcription factor *(REST)* gene

RE1-Silencing Transcription factor (*REST*) (Chong *et al.*, 1995), also known as Neuron-Restrictive Silencer Factor (*NRSF*) (Schoenherr and Anderson, 1995) is a zinc finger transcription factor member of the Kruppel-type zinc fingers family. The gene maps to 4q12, the major transcript is composed of 4 exons and the translated protein contains 9 zinc-finger domains in its N-terminal (Chong *et al.*, 1995).

It exerts its activity by binding on a specific motif in the promoter region of its target genes (NRSE - Neuron-Restrictive Silencer Element) required for its activity as gene regulator (Bruce *et al.*, 2004, Tapia-Ramirez *et al.*, 1997). It acts indirectly by recruiting co-factors that in turn perform epigenetic changes on the target genes or recruit other proteins.

The first co-factor was identified in 1999 by Andres *et al.* (1999), specifically binding at the C-terminal of the REST protein, CoREST. This factor was shown to be required to reduce the expression of the type II sodium channel gene (Andres *et al.*, 1999, Ballas *et al.*, 2001). Another co-factor, SIN3A, binds at the N-terminal of the REST protein (Huang *et al.*, 1999, Naruse *et al.*, 1999, Roopra *et al.*, 2000) and in a complex with a histone deacetylase mediates gene repression. SIN3A has also been proven to bind in different development stages compared to CoREST suggesting a more dynamic repression pattern than previously thought (Ballas *et al.*, 2005, Grimes *et al.*, 2000). An alternative mechanism by which REST regulates gene expression is through a complex with CoREST and MeCP2 (Lunyak *et al.*, 2002) where it appears to be a silencing mechanism through methylation of the CpG islands across the genetic target region. The initial role given to REST was to act outside the CNS to repress neuronal genes in non-neuronal tissues (Chong *et al.*, 1995, Schoenherr and Anderson, 1995).

However, REST mRNA has been found in adult neurons (Palm *et al.*, 1998) and its levels respond to ischemic or epileptic insults (Calderone *et al.*, 2003, Palm *et al.*, 1998). A study performed by our group also associated a *REST* exonic VNTR polymorphism in combination to a BDNF polymorphism in increased general cognitive ability (Miyajima *et al.*, 2008). Furthermore it has been proven that REST negatively regulates hippocampal neurogenesis in adult rat brains. Its dysregulation leads ultimately to depletion of the neuronal stem cells (NSC) (Gao *et al.*, 2011). A more recent study showed a protective effect of REST in the aging process of normal adult brains from age-related insults. A dysregulation of this process and REST loss from the neuronal nuclei has been observed in the early stages of Alzheimer's Disease (AD) (Lu *et al.*, 2014). Interestingly, a comparison of AD patients, all meeting the pathological criteria, showed significantly different levels of nuclear REST whether they displayed severe cognitive impairment or mild / no cognitive impairment (Lu *et al.*, 2014). Other disorders such as fronto-temporal dementia (FTD) and dementia with Lewy bodies showed a depletion of REST from the nuclei. This depletion has been associated in autophagosomes together with other misfolded proteins (Lu *et al.*, 2014). Another study (Goetzl *et al.*, 2015) also confirms a reduction of the transcription factor REST in the plasma neuronal exosomes in AD patients.

1.11 Project Main Aim

- Address the functional significance of genetic polymorphisms involved in transcriptional regulation within the promoter regions of candidate genes implicated in CNS dysfunction and/or as biomarkers for clinical predisposition to disease

1.12 Specific Objectives

- Investigate the *MAOA* gene polymorphic VNTRs and their potential role as gene expression regulators.
- Investigate the GWAS candidate gene for schizophrenia *CACNA1C* for polymorphic variants within the promoter region and their possible implications in a GxE interaction contest.
- Perform an association analysis of the novel polymorphic microsatellite in *REST* promoter region in case - control cohorts for schizophrenia, fronto-temporal dementia and Alzheimer's disease.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Commonly used Buffers and Reagents

5X TBE Buffer was made adding 108 g Tris Base (Sigma), 55 g Boric acid (Sigma), 5.84 g EDTA (Sigma) and was made up to 2L using d.H_2O.

LB Agar was made adding 40 g/L in d.H_2O (Fluka Analytical). LB agar was autoclaved the same day and stored at room temperature.

LB Broth was made adding 25 g/L in d.H_2O (Fluka Analytical). LB broth was autoclaved the same day and stored at room temperature.

2.1.2 Chromatin Immunoprecipitation (ChIP) buffers

Tris-EDTA (TE) buffer was made with 10 mM Tris and 0.1 mM EDTA (pH 8.0) made up to the final volume with sterile d.H_2O. The pH was adjusted to be between 7.5 and 8.0.

Cell lysis buffer was made with 10 mM hepes (pH 7.9), 1.5 mM $MgCl_2$, 10 mM KCl and 0.5% NP-40 and made up to the final volume with sterile d.H_2O. it was supplemented with 10 µl/ml of 100X protease inhibitor cocktail (PIC, Sigma) immediately before use and stored on ice.

Nuclear lysis buffer was made with 50 mM Tris-HCl (Tris was adjusted to pH 8.1 with HCl), 1% SDS and 10 mM EDTA and made up to the final volume with sterile d.H_2O. It was supplemented with 10 µl/ml 100X PIC immediately before use and stored on ice.

Sonication buffer was made with 50 mM hepes (pH 7.5), 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate and 0.1% SDS and made up to the final volume with sterile d.H_2O. It was supplemented with 10 µl/ml 100X PIC immediately before use and stored on ice.

ChIP dilution buffer was made with 16.7 mM Tris–HCl, 167 mM NaCl, 1.1% Triton X-100, 0.01% SDS and 1.2 mM EDTA made up to the final volume with sterile d.H_2O. It was supplemented with 10 µl/ml 100X PIC immediately before use and stored on ice.

Low-salt wash buffer was made with 20 mM Tris–HCl, 150 mM NaCl, 0.1% SDS, 1% Triton X-100 and 2 mM EDTA and made up to the final volume with sterile d.H_2O.

High-salt wash buffer was made with 20 mM Tris–HCl, 500 mM NaCl, 0.1% SDS, 1% Triton X-100 and 2 mM EDTA and made up to the final volume with sterile d.H_2O.

LiCl wash buffer was made with 10 mM Tris–HCl, 250 mM LiCl from a 10M stock, 1% Igepal, 1% sodium deoxycholate and 1 mM EDTA and made up to the final volume with sterile d.H_2O.

Elution buffer was made with 50 mM Tris–HCl, 1 mM EDTA, 1% SDS and 50 mM $NaHCO_3$ and made up to the final volume with sterile d.H_2O.

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

SDS 10% :Sodium dodecyl sulfate (Sigma) 10% w/v made up to the final volume with sterile d.H_2O.

Proteinase K: Proteinase K solution (Qiagen)

2.1.3 Drug Treatment Solutions

Cortisone (Sigma) was 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) was dissolved in sterile filtered d.H_2O to make a 1 M stock solution 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).

Sodium valproate (Sigma) was dissolved in sterile filtered dH_2O to make a 1 M stock solution 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 Human DNA Samples

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 all 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 selected at random 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 detailed medical and psychiatric history assessment using a semi-structured interview. Participants that did not meet the exclusion criteria were invited to a comprehensive interview 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 in accordance to the ethical standards outlined in the Declarations of Helsinki.

2.1.5 Human cell lines

SH-SY5Y

Human-derived neuroblastoma cell-line obtained from the American Type Culture Collection (ATCC).

HAP1

HAP1 is a near-haploid human cell line that was derived from the male chronic myelogenous leukemia (CML) cell line KBM-7 (Carette *et al.*, 2011).

HAP1 cell line have been purchased from © 2017 Horizon Discovery Group plc, Company Registration Number 08921143 (https://www.horizondiscovery.com/)

2.1.6 Cell culture media

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 ug/ml streptomycin.

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 ug/ml streptomycin.

Freezing media for SH-SY5Y

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

Freezing media for HAP1

Medium A: IMDM + 20 % FBS

Medium B: IMDM + 20 % FBS + 20 % DMSO

2.1.7 PCR primers

Table 2.1 - PCR primers used for gene expression profiling, genotyping and ChIP

Gene	Region	Forward (5′ - 3′)	Reverse (5′ - 3′)	Product Size	Template
MAOA	uVNTR	TCCGAATGGAGCGTCCGTTC	ACAGCCTGACCGTGGAGAAG	324 bp	gDNA - cDNA
	dVNTR	GGGTTAAGCGCCTCAGCTTG	CTGCTTCCTTAAGTCCACTCTTG	365 bp	gDNA
	Exon I	CGGGTATCAAAAGAAGGATCG		298 bp	cDNA
	Exon IIA		CCAGGAGCTGCTTTCCTCTGATGC		
	Exon III	TACGTAGATGTTGGTGGAGCTT		440 bp	cDNA
	Exon VI		AGAATATCCGAGTGGTGCCC		
CACNA1C	P1 construct	GCCCGATCCGAAACGAGACTCTG	GGCTTCCTCGAATCTT	350 bp	gDNA
	P2 construct	GCCACATCTGGAAGCGTTCAGC		600 bp	gDNA
	P3 construct	GCTCCCTTTGACAGCAGAGAGC		1039 bp	gDNA
REST	VNTR	GGCACTCCTTGCTTGGTAGAGG	GTCCTGTGTTGGAATGTGCGGC	178 bp	gDNA

2.1.8 DNA constructs and commercial vectors

Table 2.2 - Reporter gene and expression constructs generated for use in *in-vitro* luciferase assays

Name	Vector	Orientation	RE sites of insertion	Primers for amplification (5′ – 3′)	Application
CACNA1C P1	pGL3B	Forward	HindIII	Fw: GATCTGCGATCTAAGTGCCCGATCCGAAACGA Rev: ACAGTACCGGAATGCCAAGATTCGAGGAAGCC	Luciferase
CACNA1C P2	pGL3B	Forward	HindIII	Fw: GATCTGCGATCTAAGTGCCACATCTGGAAGCG Rev: ACAGTACCGGAATGCCAAGATTCGAGGAAGCC	Luciferase
CACNA1C P3	pGL3B	Forward	HindIII	Fw: GATCTGCGATCTAAGTGCTCCCTTTGACAGACA Rev: ACAGTACCGGAATGCCAAGATTCGAGGAAGCC	Luciferase
REST VNTR	pGL3B	Forward	HindIII	Fw: GATCTGCGATCTAAGTGGCACTCCTTGCTTG Rev: ACAGTACCGGAATGCCGCCGCACATTCCAAC	Luciferase

Note: Underlined sequences indicate vector overlapping sequences for direct cloning into the specified vector.

2.1.9 ChIP grade antibodies

Table 2.3 - Antibodies used for ChIP in human SH-SY5Y cell line

Antibody	Host Species	Species Reactivity	Immunogen	Company and Catalogue Number
Anti-H3	Rabbit (Polyclonal)	Human, mouse, rat	Synthetic peptide corresponding to the carboxy-terminal 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	Synthetic peptide corresponding to a sequence from the C-terminus of isoform a of human hnRNP K.	Abcam, ab70492
Anti-CNBP	Goat (Polyclonal)	Human, rat	Synthetic peptide: GESGHLARECTIE, corresponding to C terminal of Human CNBP.	Abcam, ab48027

2.2 Methods

2.2.1 Designing PCR 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 dimerisation 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 **Table 2.1**.

2.2.2 General Cloning Methods

PCR primer design for direct cloning into commercial vectors

Primers were designed as outlined in *section 2.2.1* with a minimum length of 15 bp. 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 are incorporated at the ends of the target DNA sequence following PCR amplification. To ensure efficient DNA cleavage by the restriction enzymes, a sequence of 15 bp overlapping the 5′ or 3′ plasmid's arm was also included in the 5′ of the primer. Primers used for direct cloning is listed in **Table 2.2** and PCR methods outlined below.

PCR using a proof-reading polymerase

Phusion High-Fidelity DNA Polymerase (NEB) was used in PCR for the amplification of DNA targets. The Phusion DNA Polymerase master mix is outlined below

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) 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 denature, 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 are listed in **Table 2.1** along with primer sequences and expected product sizes. PCR products were analysed by agarose gel electrophoresis.

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 gels supplemented with GelRed (1:10,000 dilution) (Cambridge Bioscience) or 0.5 µl per 10 ml of ethidium bromide (Sigma 10 mg/ml) and measured against a 100 bp or 1 Kb DNA ladder (Promega). The voltage (standard is 5 V/cm) and time for which the gel was run was dependent on the percentage of the gel and fragment size. The DNA was visualised using a UV transillumintor (BioDoc-it Imaging System).

Recovery of DNA from agarose-gels

Following separation through 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 manufacturer's instruction. The purified DNA was eluted in 30 µl Elution Buffer.

Restriction digest and DNA purification

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

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 (*section 2.2.2*).

Ligation – 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 a single 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 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.

Transformation of chemically competent E. Coli cells

Following ligation, the resulting plasmids were transformed into chemically competent E. Coli Sub-cloning Efficiency™DH5-α cells (Invitrogen), following manufacturer's instruction. Briefly, 5 µl of the ligation reaction or 10 ng plasmid DNA was added to 50 µl of competent DH5-α cells and incubated on ice for 30 minutes. The cells were then subjected to 'heat-shock' for 20 seconds at 42°C in a water bath followed by 2 minutes incubation on ice. Next, 950 µl of pre-warmed LB broth was added to the cells and the culture incubated with constant shaking (225 rpm) at 37°C for 1 hour. Following this, 200 µl of the culture was spread evenly onto LB agar plates supplemented with 100 µg/ml ampicillin and incubated overnight at 37°C.

2.2.3 Isolation of plasmid DNA from bacterial cultures

Mini-preparation of plasmid DNA

In order 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 (*section 2.2.2*) were transferred to 5 ml LB broth supplemented with 100 µg/ml ampicillin and cultured overnight at 37°C on a shaker at 225 rpm. DNA was isolated from the resulting bacterial culture using the QIAprep Spin Miniprep Kit (QIAGEN), according to manufacturer's guidelines. Purified plasmid DNA was eluted in 50 µl nuclease-free water into fresh 1.5 ml microcentrifuge tubes and subject to restriction enzyme digestion to check for the correct size and orientation of the insert (*section 2.2.2*).

Maxi-preparation of plasmid DNA

A Plasmid Maxi Kit (Qiagen) was used to purify 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. A 200 µl aliquot of the 5 ml starter culture from *section 2.2.2* or starter culture grown from a small scraping of a glycerol stock (*section 2.2.2*) was grown overnight at 37°C with shaking (225 rpm) in 100 ml of LB broth supplemented with the appropriate antibiotic to generate a sufficient quantity of bacteria for extraction of the plasmid DNA. DNA purification was carried out according to manufacturer's instruction for high-copy plasmids. The resulting DNA pellet was resuspended in 200-500 µl of EB buffer and quantified using a Nanodrop 8000 and then stored at -80^{o}C for long-term storage or at -20^{o}C for working stocks.

Sequencing

Plasmid DNA with cloned inserts and PCR products were sequenced externally by Dundee DNA Sequencing and Service or Source Bioscience Life Sciences. The samples and primers were supplied as required by the companies.

Glycerol stocks

Glycerol stocks of transformed bacteria were made for long term storage. A 1.4 ml volume 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 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.

pGL3-Basic (pGL3B) constructs

REST VNTR fragments containing either the 7-copy 9-copy or 12-copy variant of the *REST* VNTR were amplified from human DNA and ligated into the pGL3B vector, which lacks a minimal promoter sequence, at the HindIII restriction sites with Gibson Isothermal Assembly.

The same procedure applies for the *CACNA1C* promoter constructs.

2.2.4 Cell Culture

Culturing of SH-SY5Y and HAP1 cells

Human SH-SY5Y neuroblastoma and human HAP1 cells were maintained in culture media outlined in *section 2.1.6* at 37°C, 5% CO_2, in T175 tissue culture flasks until 70-80% confluent. To passage cells, media was removed and the cells washed down with 1X sterile PBS (Sigma) pre-warmed at 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 detached 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 mixed into a single cell suspension through pipetting. Between 1-2 ml (approximately 1-2.4 million cells depending on cell type) of the cell 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 mycoplasm infection every six months using MycoAlert Mycoplasma Detection kit (Lonza).

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 previously up to the cells being washed down with 10 ml of media. Prior to the coverslip being 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 visualised under a light microscope on the 10X objective. The number of cells within the 25 larger squares bounded by three parallel lines were 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.

Freezing cells for storage in liquid nitrogen

For long term storage, cell lines were frozen in freezing media in liquid nitrogen. The cells were grown in T175 flasks until 70-80% confluent and then passaged as described previously, but the cells were washed from the surface of the flask using 10 ml of freezing media and the cell suspension split across cryovials with 1.8 ml in each. The cryovials were immediately placed into a Mr Frosty with isopropanol at -80^0C for 24 hours and then transferred to liquid nitrogen.

Drug treatments

Drug treatments were performed using concentrations previously optimised in our lab or reported in the literature to be effective in cell culture. These are detailed in *section 2.1.3*. Drugs were diluted in appropriate volumes of complete cell culture media and added to the cells for the specified time. For each drug treatment, n=4. Basal (untreated) and drug vehicle control cells were also included. For mRNA expression profiling (*section 2.2.7*), RNA extractions were performed immediately after the drug treatment. For luciferase and over-expression assays, drug treatments were performed 4 hours post-transfection.

2.2.5 *D*elivery of plasmid DNA to cultured cells

Single transfection assays

SH-SY5Y cells were seeded into 6-well plates at approximately 400,000 cells per well and transfected with either 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 before being processed for RNA extraction.

Co-transfection assays

For luciferase assays, SH-SY5Y cells were seeded in 24-well plates at approximately 100,000 cells per well and transfected with 1 μg plasmid DNA and 10 ng pMLuc2 (Novagen) (internal control for transfection efficiency) using TurboFect (Thermo Scientific). Transfected cells were processed 48 hours post-transfection using the Dual-Luciferase Reporter Assay System (Promega).

2.2.6 Luciferase Reporter Gene Assays

Cellular lysis

At 48 hours post-transfection, tissue culture media was removed from the cells and the cells 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 incubated at room temperature on a rocking platform for 15 minutes. A 20 µl aliquot of the cell lysate was then transferred to an opaque 96-well plate for analysis.

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 setup 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 is run, which measures the bioluminescence from the reaction catalysed by the firefly and renilla luciferase enzymes. The LARII is 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.

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 internal control reduces 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 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.

2.2.7 mRNA expression analysis

***In-vitro* RNA extraction**

Total RNA was extracted using TRIzol reagent (Invitrogen) following manufacturer's instruction. Briefly, SH-SY5Y and HAP1 cells were plated out 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 10 cm^2 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 (per 1ml of TRIzol reagent) 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^oC 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 100% molecular grade isopropanol (per 1ml of TRIzol reagent) 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 removed leaving behind the RNA pellet.

To purify the RNA, 1 ml of 75% ethanol (per 1 ml of TRIzol reagent used in the initial step) 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.

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 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 through 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.

First strand cDNA synthesis

cDNA was synthesised from total RNA, extracted using methods outlined in *sections 2.2.5*, 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)	Xµl	-
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.

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 manufacturer's guidelines. The GoTaq Flexi DNA Polymerase master mix is:

Component	**Volume (n=1)**	**Final Concentration**
5X Green GoTaq Flexi buffer	2.5 µl	1X
$MgCl_2$ (25 mM)	0.9 µl	4 mM
PCR nucleotide mix (10 mM of each dNTP)	0.5 µl	0.4 mM
Forward primer (20 µM)	0.1 µl	0.2 µM
Reverse primer (20 µM)	0.1 µl	0.2 µM
GoTaq DNA polymerase (5u/µl)	0.1 µl	0.05 U/µl
Nuclease free water	X µl	-
cDNA template (1:10 dilution)	1 µl	-
Final volume	12 µl	-

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 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 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 prior to gel electrophoresis (*section 2.2.2.3*) or at -20°C for long-term storage.

2.2.8 Bioinformatic Analysis

ECR (Evolutionary Conserved Regions) Browser

Conservation of transcription factor consensus binding sequences were addressed based on the TRANSFAC 4.0 database (Matys, 2003) available through the rVista 2.0 tool 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 include 6 levels, from family of transcription factors to splice variants), 4; and similarity of the sequence to the matrix, 1.

NCBI

Sequence alignments between human and other species genomes were performed using the basic local alignment search tool of nucleotide databases (BLASTN) (Altschul *et al.*, 1997), available at NCBI (http://blast.ncbi.nlm.nih.gov/ Blast.cgi). BLAST finds regions of local similarity between sequences through making comparisons of nucleotide or protein sequences to sequence databases, calculating the statistical significance of matches. BLAST can be used to infer functional and

evolutionary relationships between sequences as well as help to identify members of gene families.

2.2.9 Genotyping

VNTR analysis

The REST VNTR was genotyped on a QIAxcel Advanced System (Qiagen) electrophoresis platform. DNA amplification was performed using 10 ng genomic DNA following the protocol described in *section 2.2.5* and the human REST VNTR primer set detailed in **Table 2.1**. For analysis on the QIAxcel Advanced System platform, 12 µl of PCR product was subjected to capillary electrophoresis following the manufacturers' protocol. The appropriate run method was the 0M800 associated with QIAxcel DNA High Resolution Gel Cartridge, with an injection time of 10 seconds, a separation voltage of 3kV and separation time of 800 seconds. The QIAxcel DNA High Resolution Gel Cartridge is designed for high-resolution (3–5 bp) genotypings. 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 a 2% agarose gel as described in *section 2.2.2*. Expected fragment sizes ranged between 178-190 bp. Sequenced samples 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 2.2.10.*

2.2.10 Chromatin Immunoprecipitation (ChIP)

***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 (see *section 2.2.3*). 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 *section 2.1.8*, **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. Details of the antibody used per IP are outlined in **Table 2.3**. The protein–DNA complexes were added to 40 µl Dynabeads™ (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. DNA bound magnetic beads were subjected to 5 minute wash steps (performed in a 4°C cold room) with rotation to remove non-specific DNA and proteins associated with the Dynabeads™. Firstly, 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 complex was eluted by adding 50 µl of elution buffer containing 50 µg/ml proteinase K to the magnetic bead/protein–DNA

complexes and the supernatant transferred to a new microcentrifuge tube and mixed at 65°C for 2 h to release the protein-bound DNA and reverse the cross-linking. The samples were then incubated at 95°C for 10 min to denture the proteins and inactivate the proteinase K and the DNA recovered from the sample through spin column purification using the Wizard® SV Gel and PCR Clean-Up System (Promega) in a volume of 20 µl. The DNA was quantified using a Nanodrop 8000 and analysed by PCR (*section 2.2.5*).

2.2.11 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 manufacturer's instruction. Briefly, a total of 300 ng of sonicated genomic 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 are 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 transferred to a new microcentrifuge tube and the beads discarded. Samples were subjected to PCR analysis (*section 2.2.5*) and then stored at -20°C.

2.2.12 Statistical Analysis

Clump analysis

Significance-testing of allele frequency and genotype data for the REST VNTR between cases and controls of the schizophrenia cohort (Materials *section 2.1.4*) and AD and FTD cases and healthy controls 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) that have the same marginal totals as the one under consideration and counting the number of times that a chi-squared value associated with the real table is achieved by the randomly simulated data. 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 and directly tests the hypothesis that several alleles are more common among the cases than among the controls (Sham and Curtis, 1995). The Clump program generates four test statistics using Monte Carlo methods to evaluate the significance of chi-squared values by assessing how many times 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

Hardy-Weinberg Equilibrium (HWE)

As a measure of quality control in our cohorts, 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 goodness-of-fit of the observed genotype frequency against the expected under HWE (Rodriguez *et al.*, 2009).

Chapter 3

Distinct chromatin structures at the *MAOA* promoter correlate with allele specific expression in female cells.

3.1 Introduction

It is thought 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, which mediates response to stress that the individual has, or is experiencing. This gene and the uVNTR polymorphism are the most often cited example of a genotype by environment interaction in which an individual's genotype moderates the effect of environmental experience to alter mental health outcomes (Caspi *et al.*, 2002). Simplistically the function of the uVNTR is modified by the complement of active transcription factors in the cell, altered in response to the environment, that are able to recognise and bind to the uVNTR thus allowing the domain to act as a transcriptional regulator. However two recent advances in understanding the regulation of this gene have questioned this simple hypothesis. Firstly, bioinformatic data now indicates that there are at least two transcriptional start sites (TSS) for the *MAOA* gene. Furthermore, the newly described transcriptional start site now places the uVNTR in the 5′ UTR (5′ untranslated region) of the MAOA gene (**Figure 3.1**). Secondly, the *MAOA* locus has been proposed as one of the loci that can escape X inactivation (Carrel and Willard, 2005, Joo *et al.*, 2014, Mugford *et al.*, 2014).

To address the mechanisms which might underpin the sex differences in studies of *MAOA* G×E, we defined regulatory mechanisms at the *MAOA* promoter. One of the *MAOA* mRNA isoforms, according to UCSC web browser, now contains the uVNTR in the 5′ UTR. This allowed us to perform the analysis of allele specific expression and chromatin structure in the female cell line SH-SY5Y, which is heterozygous for this VNTR. We were able to demonstrate that both alleles, at least in the population of cells, can support expression of *MAOA*.

However, using chromatin immunoprecipitation (ChIP) analysis we demonstrate that each allele can support distinct transcription factor, histone binding and methylation patterns; the clear allelic difference in binding to the promoter and the expression from both alleles would be supportive of the model that the *MAOA* gene is within a genomic region that can escape X inactivation.

3.2 Aims

- Identify and assess the expression pattern of the *MAOA* mRNA in the neuroblastoma cell line SH-SY5Y
- Analyse the differences in the expression led by the *MAOA* VNTRs
- Analyse the GxE interaction in the SH-SY5Y cell line on both expression levels and chromatin interactions
- Analyse the methylation pattern in the *MAOA* gene promoter region

3.3 Results

3.3.1 Bioinformatic analysis

3.3.1.1 Human *MAOA* Gene Promoter Architecture

The uVNTR was originally described as being 1.2kb upstream of the TSS of the MAOA gene (Sabol *et al.*, 1998), however bioinformatic analysis from data in Hg19 and Hg38 of the human genome using ENCODE data on the USCS genome browser (http://genome.ucsc.edu) (Kent *et al.*, 2002)(Kent *et al.*, 2002) demonstrated an additional TSS which is upstream of the uVNTR. This now locates the uVNTR in the 5′ UTR of the *MAOA* gene generating two possible *MAOA* isoforms, both active and that will encode for two different MAOA proteins (**Figure 3.1**). The ENCODE data shows those transcription factors determined by chromatin immunoprecipitation assay in a variety of cell lines to bind to this promoter region: of interest was the strong binding of CTCF, a transcription factor that we have previously implicated in regulation of the human serotonin transporter gene (*SLC6A4*) via in part regulatory VNTRs (Ali *et al.*, 2010, Klenova *et al.*, 2004, Stevens *et al.*, 2012, Vasiliou *et al.*, 2012). These VNTRs in the *SLC6A4* gene are associated with risk for neuropsychiatric disorders and respond to many pharmaceutical challenges such as lithium and cocaine *in-vitro* (Haddley *et al.*, 2012, Klenova *et al.*, 2004, Roberts *et al.*, 2007, Vasiliou *et al.*, 2012). Less than 500bp upstream from the uVNTR lies the dVNTR which has been demonstrated to have differential regulatory activity based on copy number of the repeat in reporter gene assay and different copy variants have been correlated to antisocial personality disorder (Philibert *et al.*, 2011). Bioinformatic data suggests that *MAOA* locus has two different TSSs and that in part differential expression of *MAOA* gene may be directed by these distinct VNTR regulatory motifs.

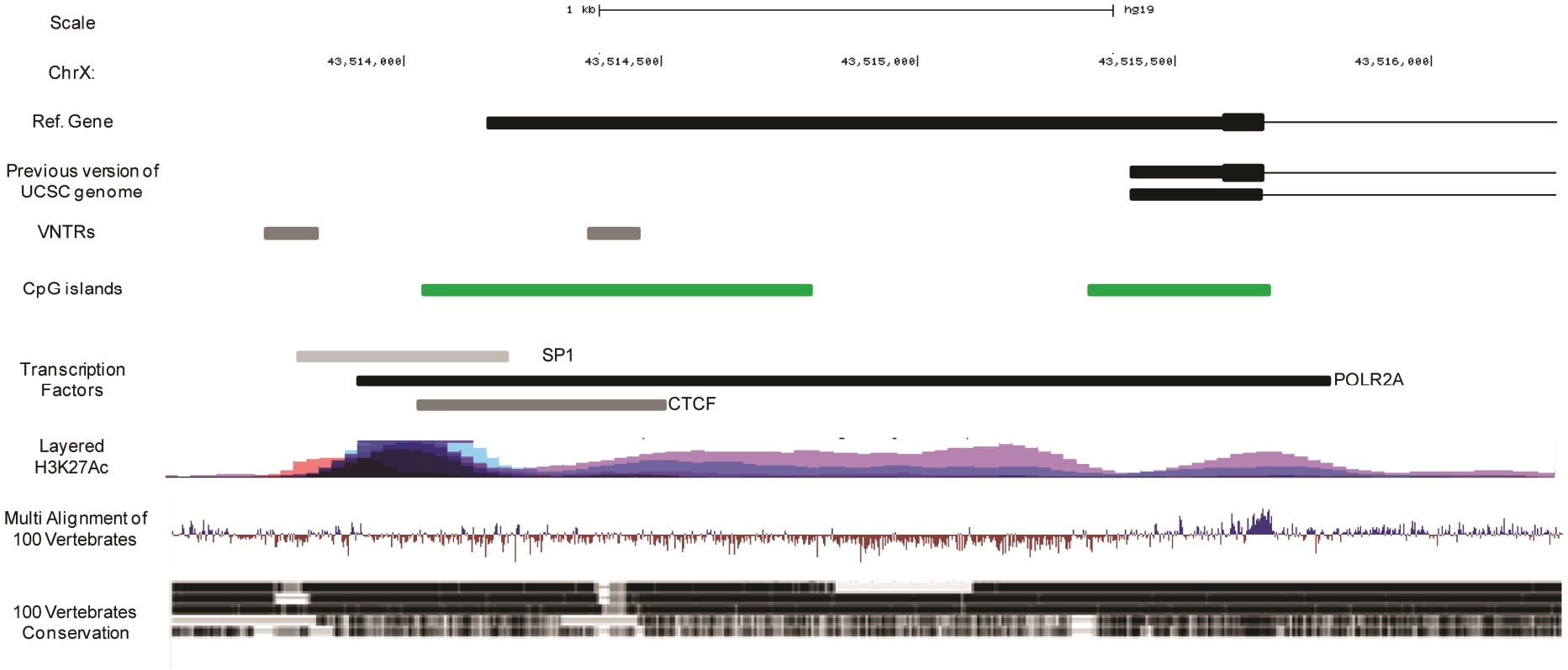


Figure 3.1 – Monoaine Oxidase A (*MAOA)* promoter region. Graphic representation of *MAOA* promoter region based on UCSC genome web browser (Hg19). From top to bottom: reference gene sequence; previous version of UCSC genome; tandem repeats: from left to right dVNTR, uVNTR; CpG islands; ENCODE transcription factor data (Data version: ENCODE Mar 2012 Freeze) indicating those factors demonstrated to bind to this region in specific cell lines, in short the darker the line the more confidence in the association; ENCODE data for H3K27Ac mark (often found near active regulatory elements) on 7 cell lines; multi alignment (of 100 vertebrate) representing the degree of homology between primates and other species; vertebrate conservation (100 species).

However the primary sequence of both the u- and dVNTR are very different and this would suggest they could respond to different transcription factor pathways.

Furthermore the dVNTR sequence predicts the formation of a G-Quadruplex structure as described in Chapter 1, *section 1.6.* As illustrated in **Figure 3.2** the uVNTR sequence does not appear to contain potential binding sites for transcription factors while the dVNTR contains several binding motifs for SP1, hnRNPK and CNBP transcription factors overlapping the repeats of the copy number variation.

A similar scenario has been proposed using the C-Myc gene as a model in which the double helix DNA would change its conformation modulating gene expression (Brooks and Hurley, 2009). Binding from the transcription factor SP1 would increase transcription of the gene, or at least promoting it, which in turn would increase the supercoiling upstream of the binding site allowing the downstream region to take single-stranded conformation. This would allow the formation of the G-Quadruplex in the purine rich strand and the i-motif in the pyrimidine rich strand potentially silencing the transcription (Sun and Hurley, 2009). Alternatively single strand DNA binding proteins such as hnRNPK and CNBP would bind to the purine rich and pyrimidine rich respectively maintaining the transcription active. Based on this assumption and on the presence of the binding motifs in the dVNTR core sequence we will test the hypothesis of a similar mechanism working on the *MAOA* promoter. Thus a different genotype, containing different copies of this VNTR, will modify the number of binding sites for these transcription factors which, in turn, might modulate the transcribed *MAOA* mRNA.

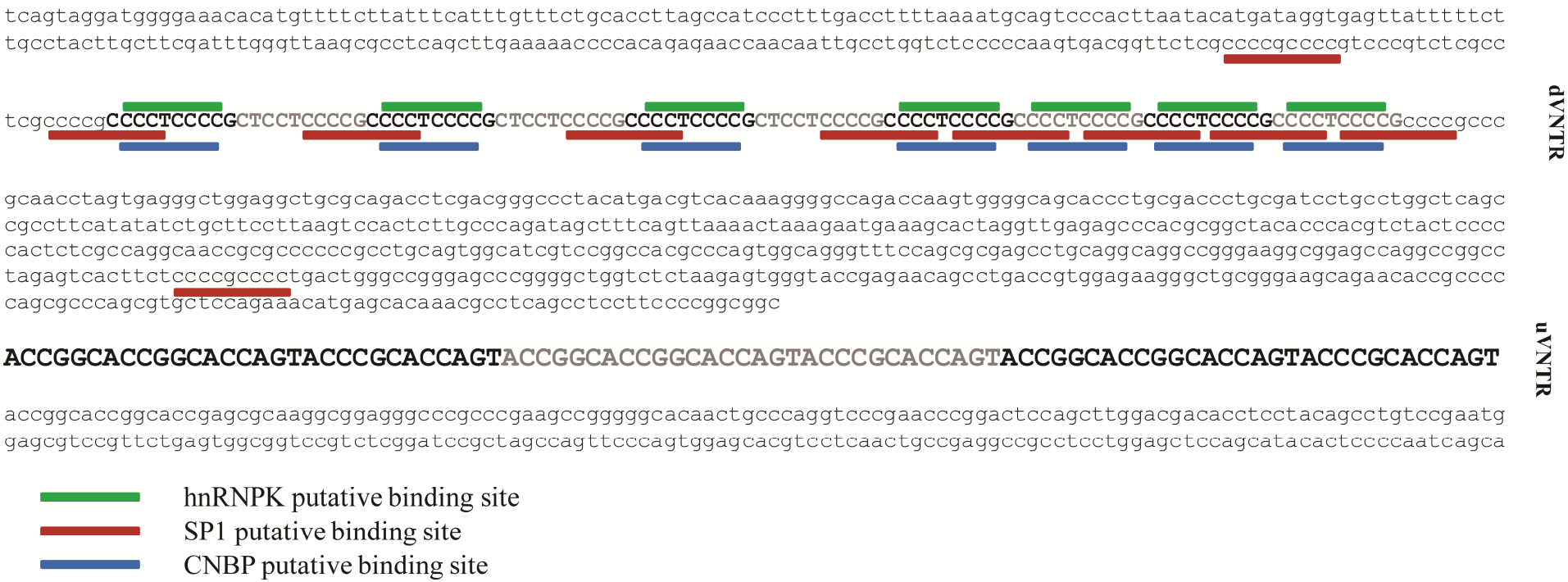


Figure 3.2 - Bioinformatic analysis on transcription factor binding sites over *MAOA* dVNTR and uVNTR genetic region. Genetic locus comprehensive of the *MAOA* dVNTR and uVNTR. Underlined in red the predicted SP1 binding motif (Raiber *et al.*, 2012); in green the predicted hnRNPK binding motif (Michelotti *et al.*, 1996) and in blue the predicted CNBP binding motif (Armas *et al.*, 2013).

3.3.1.2 *MAOA* VNTRs conservation analysis

DNA sequences of the genomic regions encompassing the *MAOA* u- and dVNTR were taken from UCSC browser through the *Multi Alignment of 100 Vertebrates* tool for each of species displayed in **Figure 3.3** and aligned with Clustal Omega multiple sequence alignment online tool (http://www.ebi.ac.uk/Tools/msa/clustalo/). **Figure 3.3 A** shows the dVNTR sequence in humans and non-human primates, the dVNTR is not present in rodents. The core sequences (A and B), as reported by Philibert *et al.* (2011), can be found in all species analysed except for marmoset which sets the separation line between the primates order in old world monkeys with the new world ones that we analysed. The described arrangement ABABABAAA can only be found in humans. In chimps, our genetic closest relatives in this analysis, it can be found two A sequences and one B with an A AB arrangement where between the A sequences is present an imperfect B sequence. Gorillas have an A B structure separated by a partial A sequence. Orangutans, surprisingly, show the higher number of copies after humans with a total of three A sequences and four B sequences with a structure B AB ABAB with a partial A sequence after each B copy variant (**Table 3.1**). Both upstream and downstream flanking regions are highly conserved among species. **Figure 3.3 B** points out the *MAOA* uVNTR sequence in humans and non-human primates. The complete 30bp *MAOA* uVNTR sequence that in human can be found from a minimum of 2 repetitions up to 6 repetitions (Huang *et al.*, 2004, Sabol *et al.*, 1998), goes back in time until chimps, where a single complete repeat can be found. Once again, the flanking sequences, upstream and downstream the repetitive element are highly conserved among species even in the most phylogenetic distant ancestors. These conservation data may be in line with a recent report affirming that diversity among great apes in the X chromosome is lower than in autosomes (Nam *et al.*, 2015).

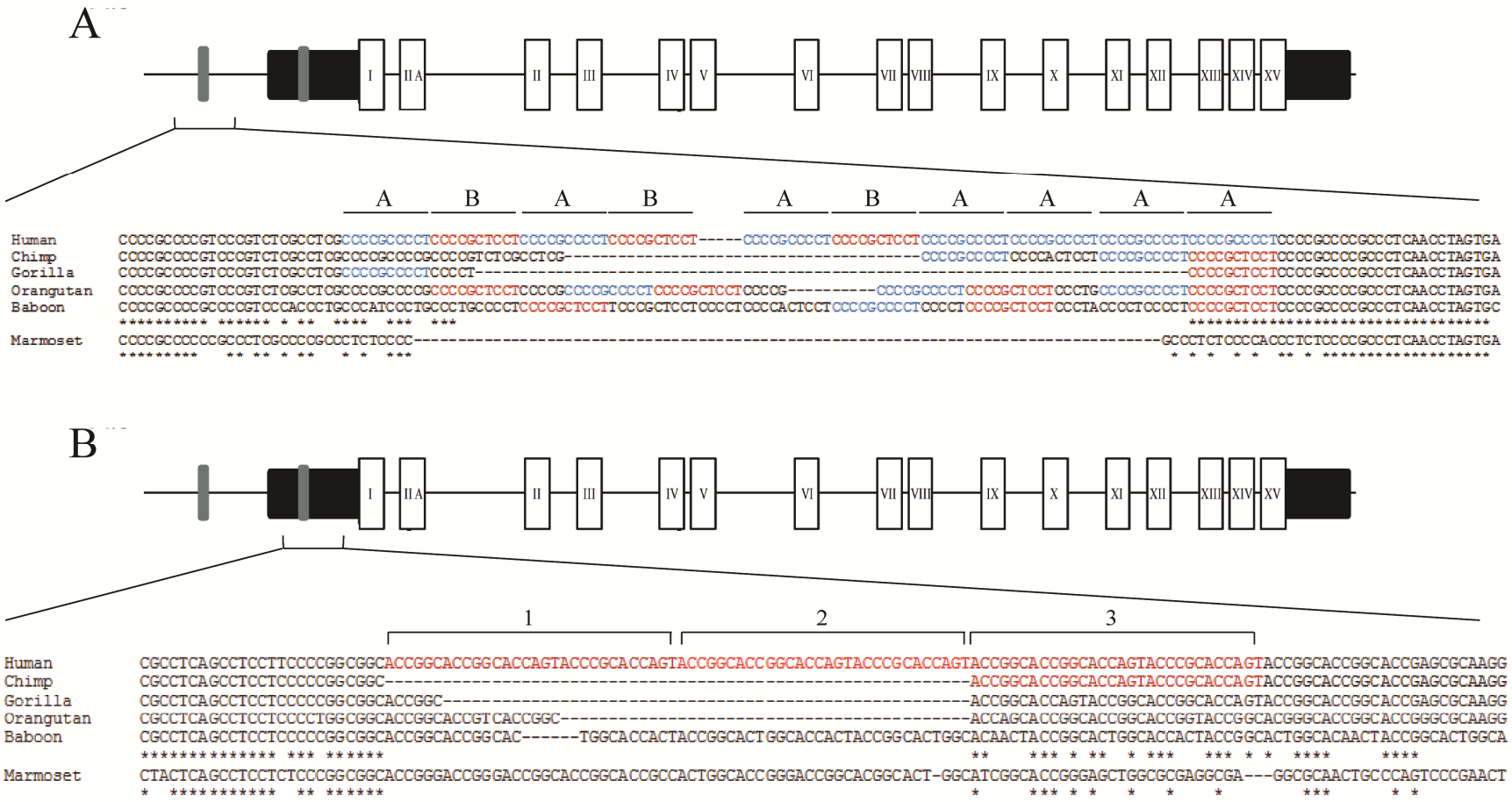


Figure 3.3 – Monoamine Oxidase A (*MAOA)* VNTRs conservation in human and non-human primates. **A.** Conservation analysis of the *MAOA* dVNTR. At the top graphic illustration of the *MAOA* gene. Underneath the alignment of the sequences for humans and non-human primates. **B.** Conservation analysis of the *MAOA* uVNTR. At the top graphic illustration of the MAOA gene. Underneath the alignment of the sequences for humans and non-human primates.

Table 3.1 - MAOA dVNTR conservation

Human	A	B	A	B	A	B	A	A	A	A
Chimp	*A*	*B*	B	A	*A*	A	B			
Gorilla	A	A	B							
Orangutan	A	B								
Baboon	*A*	B	*B*	*A*	A	A	B	*A*	B	

A Repeat	CCCCGCCCCT
B Repeat	CCCCGCTCCT
A ; B	Perfect Repeat
A ; *B*	Imperfect Repeat
A ; B	Partial Repeat

Note: summary of the dVNTR repeats as present in humans and different non-human primates as illustrated in ***Figure3.3***

3.3.2 Allelic specific regulation of *MAOA* expression

We used the human neuroblastoma cell line SH-SY5Y to address *MAOA* expression. SH-SY5Y is a well characterised cell line which was originally isolated from a four year old female. Genotyping of this cell line demonstrated that it was heterozygous for the uVNTR with 3 and 4 copies of the repeat element (**Figure 3.4 A**). This allowed identification of allelic specific expression when initiated from the newly described TSS which places the uVNTR in the 5′ UTR. The *MAOA* uVNTR primer set that we used for the gDNA genotype can also be used to assess the mRNA expression since this genomic region, comprehensive of the uVNTR, has been

transcribed in the *MAOA* mRNA, thus converted into cDNA after the RT-PCR process.

Under basal growth conditions we were able to detect mRNA corresponding to expression from both alleles using the uVNTR length as the distinguishing feature (**Figure 3.4 A**). Likewise, analysis of cDNA prepared from saliva from females heterozygous for the uVNTR demonstrated that *in-vivo* expression from both alleles could be observed in the same individual (**Figure 3.4 B**). In our experience PCR analysis of VNTRs can lead to different molar intensities of product from each allele. Therefore we made no attempt to correlate intensity of PCR product with expression levels, e.g the Longer (L – 4R) allele being more intense than the Shorter (S – 3R) allele or vice versa, in the same experiment. We observed a similar bias in the PCR of genomic DNA in several of our subsequent ChIP analysis. However we were able to compare intensity between experiments for the same primer set; this demonstrated that in SH-SY5Y cells in response to exposure to sodium valproate, a mood stabiliser, differential *MAOA* allelic expression was observed. Specifically the ratio of L- to S- allele expression was reversed (**Figures 3.4 A - 3.5 A**). Similarly to sodium valproate, exposure to ethanol, lithium chloride and cortisone at different concentrations (20, 50 and 100 nM) exerted a similar response. **Figure 3.5** highlights the different expression levels obtained by the subsequent analysis of the treatments over the *MAOA* expression. This switch in relative levels of expression cannot be explained by PCR allele preference and reflects differential regulation of the alleles in response to the stimulus. As the *MAOA* gene is on the X chromosome the expression of both alleles could have been expected to be affected by the inactivation of one allele. The expression of both alleles does not have to be consistent with *MAOA* escaping X chromosome inactivation but could be construed as a mosaic

effect of random silencing of one allele in any particular cell and thus the population would appear to express both alleles. However there have been reports of large sections of the X chromosome escaping X inactivation which includes regions encompassing the *MAOA* gene (Carrel and Willard, 2005, Joo *et al.*, 2014, Mugford *et al.*, 2014).

To explore the allelic expression further we analysed by ChIP, transcriptional and epigenetic variation at the *MAOA* promoter in SH-SY5Y cells in response to valproate using again the size difference of the uVNTR to identify allelic specific changes. RNA polymerase II was demonstrated to be present on both alleles, consistent with expression from both alleles (**Figure 3.6 A**). This was in stark contrast to histones which were found only on the promoter of the shorter, 3 copy repeat, uVNTR allele (**Figure 3.6 B**). The lack of histones on the 4 copy uVNTR promoter was observed with two different antibodies, H3 which identifies all histone 3 proteins and H3K9me3 which is indicative of an inactive histone complex consistent with promoter inactivation and heterochromatin formation (Hublitz *et al.*, 2009). The methylation status of both alleles was addressed by their ability to associate with methylation binding domain protein (MBD). This demonstrated, under basal growth conditions, a clear allelic specificity in that the 4 copy uVNTR promoter, which did not associate with the active histone mark, was predominantly methylated whereas the 3 copy uVNTR promoter was unmethylated (**Figure 3.7**). We observed a small unmethylated proportion of the 4 copy allele. The methylation pattern was altered when the cells were exposed to sodium valproate, most noticeably in the methylated fractions, in that there was now a significant proportion of the 3 copy allele which was methylated.

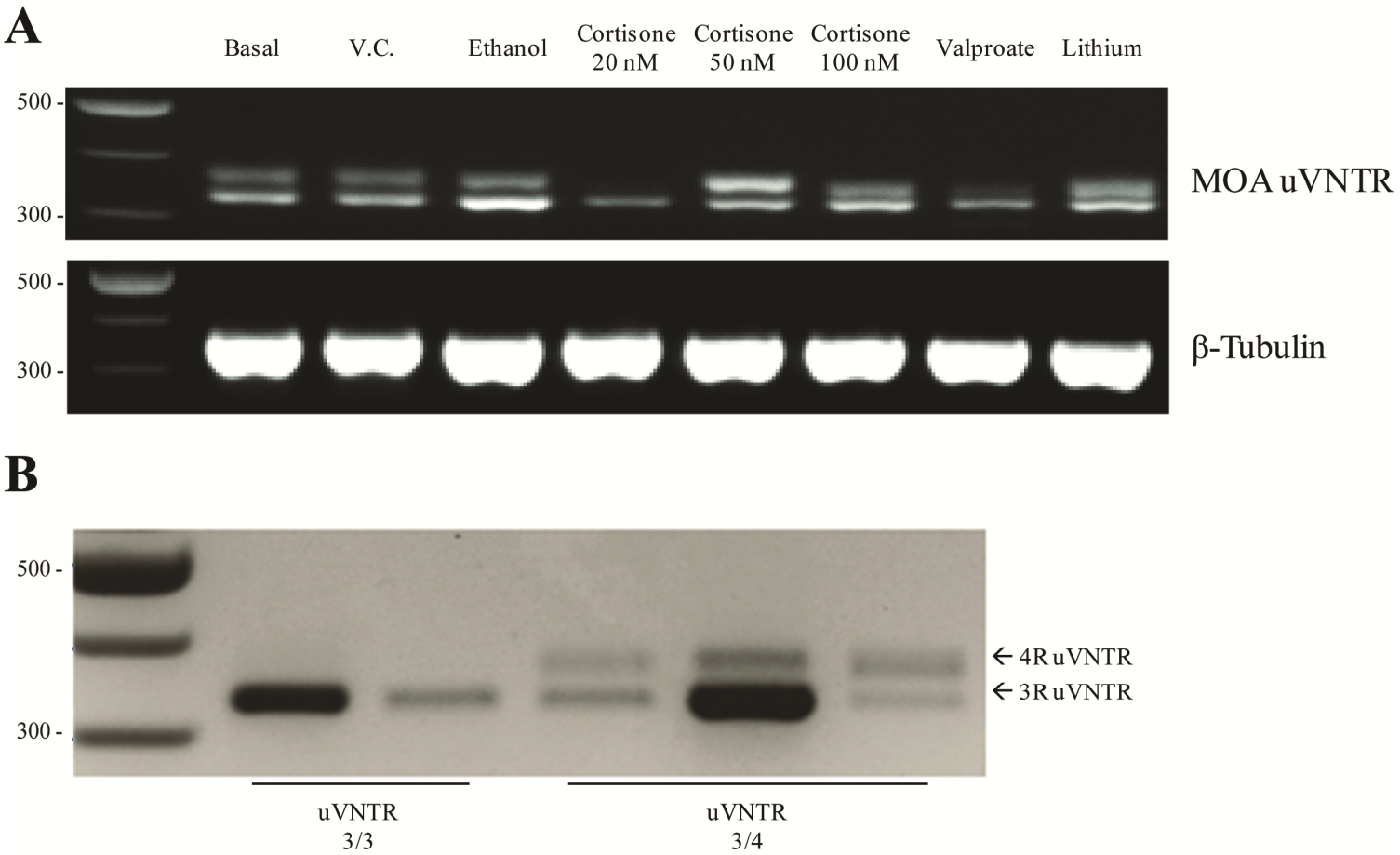
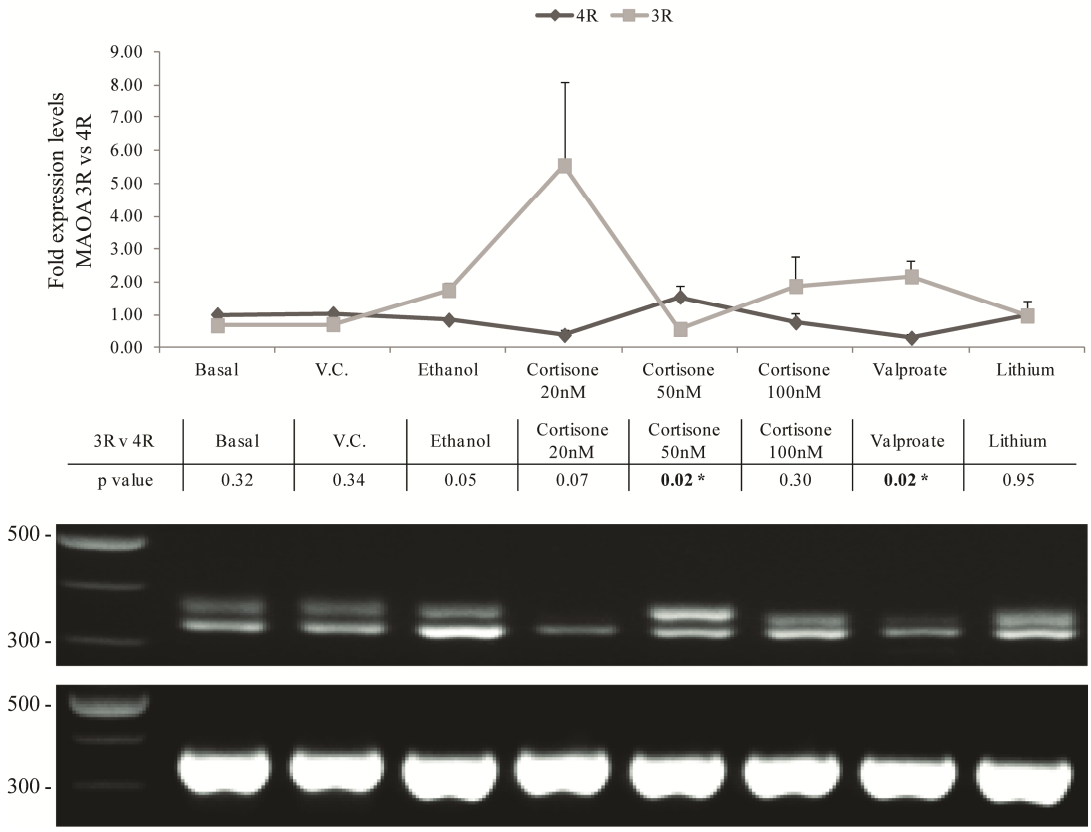


Figure 3.4 - 1.5% agarose gel of Monoamine Oxidase A (*MAOA)* uVNTR PCR amplification. Amplicon size 324bp for 3R VNTR, 354bp for 4R VNTR. **A**. SH-SY5Y cell line cDNA: lanes from left to right: 100bp ladder marker; basal condition; vehicle control: H_2O 1 hour ; sodium valproate treatment 1 hour; PCR negative control. **B**. human female samples cDNA: lanes from left to right: 100bp ladder marker; 2x homozygous 3R/3R; 3x heterozygous 3R/4R.



3R v 4R	Basal	V.C.	Ethanol	Cortisone 20nM	Cortisone 50nM	Cortisone 100nM	Valproate	Lithium
p value	0.32	0.34	0.05	0.07	**0.02** *	0.30	**0.02** *	0.95

Figure 3.5 - Monoamine Oxidase A (*MAOA)* cDNA expression. A. Line graph illustrating the *MAOA* mRNA fold expression level comparing the uVNTR 3R vs uVNTR 4R. The table represent the Student's T-Test (two-tailed type 2). **B.** The agarose gels illustrate a sample of the PCRs used to generate the graph. At the top *MAOA* uVNTR, right below the housekeeping gene β-Tubulin. Values for both 3R and 4R uVNTR have been normalised for β-Tubulin before calculating the fold increase. For each condition N=3. * = $p<0.05$.

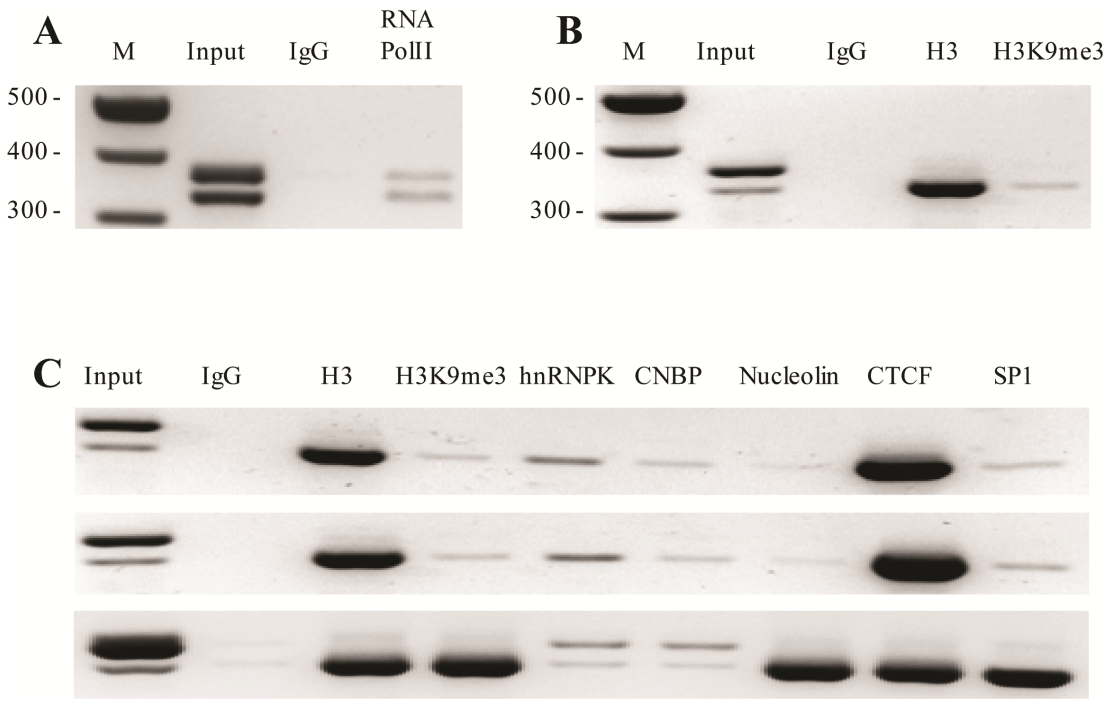


Figure 3.6- ChIP analysis of the Monoamine Oxidase A (*MAOA)* VNTRs alleles. Amplicon size 324bp for 3R VNTR, 354bp for 4R VNTR. **A.** *In-vivo* interaction of active RNA pol II (CTD phospho Ser5) under basal conditions, lanes from left to right: 100bp ladder marker; input DNA (1% sheared chromatin) acts as a positive control for PCR; IgG: –ve control for non-specific background binding; active RNA pol II (CTD phospho Ser5) 10ug DNA. **B.** *In-vivo* interaction of Histone 3 and active Histone mark H3K9me3 with the MAOA alleles under basal conditions. Lanes from left to right: 100bp ladder marker; input DNA (1% sheared chromatin); IgG; Histone 3; active Histone mark H3K9me3. **C.** ChIP analysis of the *in-vivo* interaction of factors associated with alternative DNA structure formation with MAOA alleles. Under Basal conditions (top) , following 1hr exposure to H_20 (middle), following 1hr exposure to 2 µM sodium valproate (bottom). Lanes from left to right: input DNA (1% sheared chromatin); IgG; Histone 3; active Histone mark H3K9me3; Single Stranded DNA (ssDNA) binding protein hnRNP K; ssDNA binding protein CNBP; G4 binding protein nucleolin; transcription factor CTCF; transcription factor Sp1.

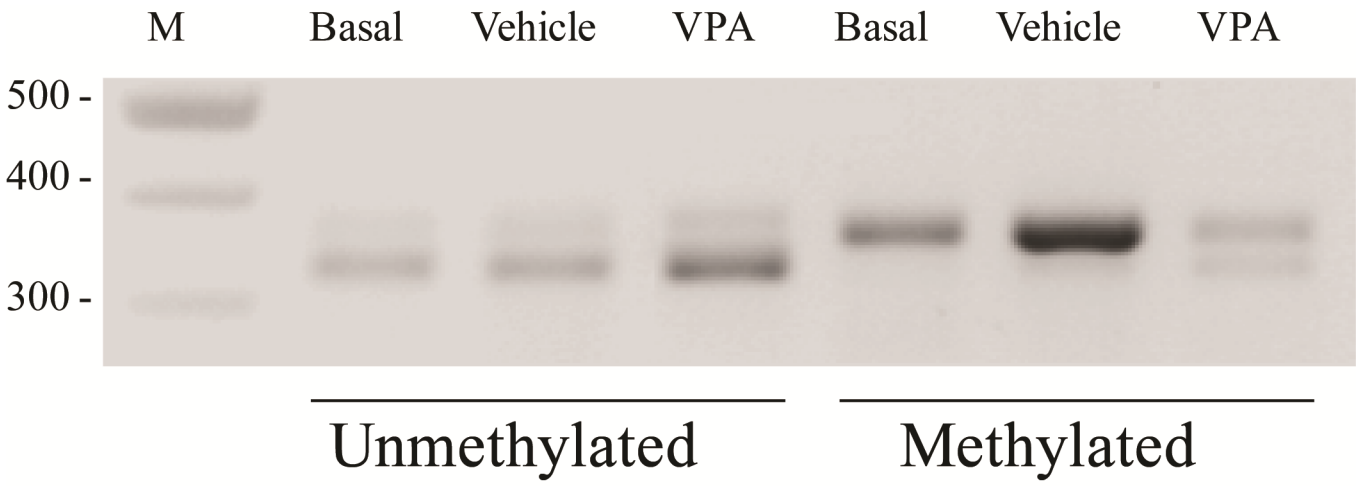


Figure 3.7- Methylation at the Monoamine Oxidase A (*MAOA)* promoter. PCR amplification of sheared gDNA post CpG MethylQuest treatment. Amplicon size 324bp for 3R VNTR, 354bp for 4R VNTR. Lanes from left to right: 100bp ladder marker; basal condition; vehicle control: H_2O ; sodium valproate treatment of unmethylated DNA: not bound to methyl domain binding protein (CpG MethylQuest™ Protein). Basal condition; vehicle control: H_2O ; sodium valproate treatment of methylated DNA bound to methyl domain binding protein (CpG MethylQuest™ Protein).

To complement the distinct epigenetic marks on the MAOA alleles we addressed the potential binding of transcription factors to the promoter (**Figure 3.6 C**). The selection of transcription factors for analysis was based mainly on a) the CpG island, and b) our previous study on CTCF and Sp1 interactions on VNTRs in several genes, including the *SLC6A4* gene (Haddley *et al.*, 2012, Klenova *et al.*, 2004, Roberts *et al.*, 2007, Vasiliou *et al.*, 2012). We therefore addressed the binding to the promoter by ChIP analysis of 4 transcription factors: CTCF, Sp1, HnRNP K, CNBP and the transcriptional coactivator protein nucleolin. These proteins have previously been shown to bind similar sites or affect the binding of one another to regulatory regions (Lobanenkov *et al.*, 1990, Michelotti *et al.*, 1995, Witcher and Emerson, 2009). As for histones, under basal growth conditions, transcription factor binding was only observed on the short, 3 copy uVNTR allele (**Figure 3.6 C**). Upon exposure to sodium valproate two major changes were observed; 1) binding of the single stranded nucleic acid binding proteins hnRNP K and CNBP on the longer 4 copy uVNTR and 2) a significant increase in the intensity (equated to binding) for the inactive histone (H3K9me3), nucleolin and Sp1 to the short allele.

Previous models of chromatin structure have proposed that constitutively active genes have proximal promoters with no or few nucleosome which allows access for transcription factors to initiate the transcriptional event (Choi and Kim, 2008), whereas transcription of stimulus inducible genes would require movement or removal of well-positioned nucleosomes from the promoter to allow the transcription factor to interact with the consensus binding site (Li *et al.*, 2007). Based on this hypothesis and the findings presented above, our data would be consistent with the short 3 copy allele being at least in part, stimulus inducible in our cell model. We undertook preliminary testing of this hypothesis by exposing SH-SY5Y cells to

challenges we have previously used to elicit responses in modulation of transcription factors and methylation state of cell lines (Haddley *et al.*, 2012, Roberts *et al.*, 2007, Vasiliou *et al.*, 2012).

3.4 Discussion

The *MAOA* uVNTR is one of the most studied GxE interactions in the literature and the accepted hypothesis is that the different repeat copy numbers regulate the differential expression of *MAOA* in response to the environment. This is a relatively straightforward explanation for the male gender which has only one allele for this gene. However, the situation is more complex in females for several reasons including a) gene dosage, b) the potential for *MAOA* to escape X-inactivation and c) the female, unlike the male, can be heterozygous at this locus which could correlate with distinct expression profiles to differentially modulate *MAOA* gene expression. The *MAOA* gene is one of the approximately 15% of X chromosome genes which may escape X inactivation (Carrel and Willard, 2005, Joo *et al.*, 2014, Mugford *et al.*, 2014). However there is still much debate about X inactivation *in-vivo* and there may be no reason why X-inactivation of the second allele in females itself is not dependent on GxE interactions. Consistent with this, the gene termed a master regulator of X chromosome inactivation, *XIST*, is over expressed in females with major affective disorders (Ji *et al.*, 2015). This would help, in part to explain, the differences in association seen for specific mental health issues when male and females are compared when the uVNTR is assessed as a risk factor, such examples include: *MAOA* uVNTR moderating the relationship between childhood maltreatment and dysthymia only in females, gender differences including pronounced effects on ADHD and anxiety in females whereas the opposite is true for

autism, bipolar disorder and aggressive behaviour in males (Melas *et al.*, 2013, Reif *et al.*, 2012). Indeed many studies of *MAOA* G×E report only on males due to compounding effects including female heterozygosity for the uVNTR (Caspi *et al.*, 2002, Fergusson *et al.*, 2012, Kim-Cohen *et al.*, 2006). Thus, it is likely that both sex-based and disorder-based differences exist.

We investigated the molecular mechanisms underlying *MAOA* expression in the human neuroblastoma cell line SH-SY5Y which is both female and heterozygous for the uVNTR. This heterozygosity was crucial for our study to correlate mRNA expression from each individual *MAOA* allele with transcription factor binding and epigenetic marks at the proximal promoter. Our data suggested that the *MAOA* gene might escape X inactivation in SH-SY5Y cells and that both alleles were active, albeit by very distinct molecular mechanisms. Most noticeably although both alleles had different chromatin structures (methylation and transcription factor interactions), when ChIP was performed for an active RNA polymerase (anti-RNA pol II CTD phospho Ser5 antibody) binding was observed at both alleles which was consistent with the cDNA studies. This demonstrates that both alleles were active despite the distinct chromatin architectures (**Figure 3.6 A**). The methylation of one allele which apparently did not repress *MAOA* expression (**Figure 3.7**) would be consistent with recent work on the X chromosome which indicated that methylation does not necessarily correlate with gene repression but rather functions as a parameter affecting a region's ability to escape X inactivation (Joo *et al.*, 2014). Our data adds to the importance of gender in regulation of *MAOA* expression via epigenetic mechanisms, e.g. the gender specific changes of methylation in a longitudinal study of twins (Wong *et al.*, 2010) and hypomethylation in the pathogenesis of panic disorder particularly in female patients (Domschke *et al.*, 2012).

To further investigate the regulation of the *MAOA* promoter we analysed binding for CTCF, Sp1, hnRNP K, CNBP and nucleolin. Our data provided evidence that these proteins recognise the proximal *MAOA* promoter and more importantly demonstrated distinct patterns of binding over each specific allele which could be modulated by inducible gene expression (**Figure 3.6 C**). Our transcription factor binding profile to the promoter is consistent with our previous findings of distinct chromatin architecture over the heterozygous alleles which correlated with differential expression in response to sodium valproate (**Figure 3.4**). That mRNA expression is increased from the 3 copy allele in response to multiple challenges (ethanol, cortisone, sodium valproate and lithium - **Figure 3.5**), might represent the fact that this allele is the one most likely to respond initially to transcription factor changes as minimum methylation is present so allowing ease of access for the transcription factors to recognise their DNA binding sites. However the complex genomic architecture at each *MAOA* allele remains to be resolved in future experiments.

In females the differences in epigenetic marks and transcription factor binding over the *MAOA* promoter *in-vivo,* which we have defined *in-vitro,* may be generated by inconsistent X-inactivation and the action of other environmental challenges that modulate promoter activity including the uVNTR. This would be consistent with the over expression of *XIST* in major affective disorders in females (Ji *et al.*, 2015). Combined the data suggests a requirement for the reappraisal of the regulation of X chromosome gene expression, and specifically *MAOA* in females that might contribute to mental health issues. However we have demonstrated with valproate exposure and other challenges that the level of expression could still be altered. In this manner GxE interactions in females could be quite distinct from those in males

in which a more conventional chromatin structure over the promoter might be expected. Therefore females may be even more variable with respect to both X-linked gene expression and subsequent GxE interactions. Our data reappraises the regulatory mechanisms operating at the *MAOA* promoter to control gene expression and furthermore whether a gene dosage model is important for gender effects on mental health. The gender effect could modulate CNS function and development at any time in the foetus, child or adult as it would still be dependent on transcription factor expression to modulate the regulator functions of the proximal *MAOA* promoter. The same mechanism may operate on other genes on the X chromosome implicated in mental health disorders. It may also be relevant in conditions such as Klinefelter Syndrome (XXY) and triple X syndrome as in general these patients show an increased incidence of psychiatric conditions (DeLisi *et al.*, 2005, Otter *et al.*, 2010). The complexity of X inactivation to modulate the action of risk factors involved in predisposition to a disorder may also have significant consequences for genome wide association studies on the X chromosome.

Chapter 4

MAOA isoforms in the haploid HAP1 cell line

4.1 Introduction

The *MAOA* gene expresses at least two mRNA isoforms that are reported in the major online web browsers. These two isoforms lead to translation of two distinct MAOA proteins whose function might be modified by the difference in the amino acid sequence that these two isoforms possess (**Figure 4.1**). The effect and activity of the minor *MAOA* isoform has not previously been reported to our knowledge and we will try to address its mRNA and protein expression in this chapter.

The HAP1 cell line we utilised for these experiments is a semi-haploid cell line derived from KBM7 cell line. The HAP1 cell line possess a single set of chromosomes with the exception of a 30 Mbs (megabases) fragment of chromosome 15 which is integrated on the long arm of chromosome 19. Carette *et al.* (2011) in an unsuccessful attempt to induce pluripotency in KBM7 cells for their experiment, generated the HAP1 cell line. This haploid cell model is used as a tool by the company Horizon that through genome editing with Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) - Cas9 technique is able to knockout (KO) specific genetic sequences.

The single set of chromosomes possessed by this cell line makes it ideal to completely and specifically knock out genomic regions of interest either to assess gene expression or protein function and interested pathways in which these proteins might be involved.

We exploited this major advantage of the HAP1 cell line purchasing from Horizon *MAOA* KO cell lines for the uVNTR, the dVNTR and both VNTRs, in order to assess the effect that the two VNTRs in the *MAOA* promoter region, might have in the expression of the two different isoform start sites.

4.2 Aims

- Identify and assess the expression pattern of the *MAOA* mRNA and protein isoforms due to the different TSSs in the HAP1 cell line clones
- Analyse the differences in the expression led by the *MAOA* uVNTR and dVNTR in the different KO cell lines for the two TSSs
- Validate the GxE interaction, observed for the *MAOA* gene in Chapter 3, in the HAP1 cell line

4.3 Results

4.3.1 Bioinformatic analysis on *MAOA* gene

4.3.1.1 *MAOA* isoforms description

Different accredited web browser such as UCSC (https://genome.ucsc.edu/), AceView (http://www.ncbi.nlm.nih.gov/ieb/research/acembly/), UniProt (http://www.uniprot.org/) and Ensemble (http://www.ensembl.org/) report two major isoforms of the MAOA mRNA which in turn translate two different MAOA protein isoforms, along with different unspliced and untranslated oligonucleotides.

Figure 4.1 defines these mRNAs (**Figures 4.1 A** and **4.1 C**) along with a novel isoform identified in this study (**Figure 4.1 B**). We will refer to these isoforms as Long MAOA (L), Medium MAOA (M) and Short MAOA (S) isoforms according to the expected protein weight.

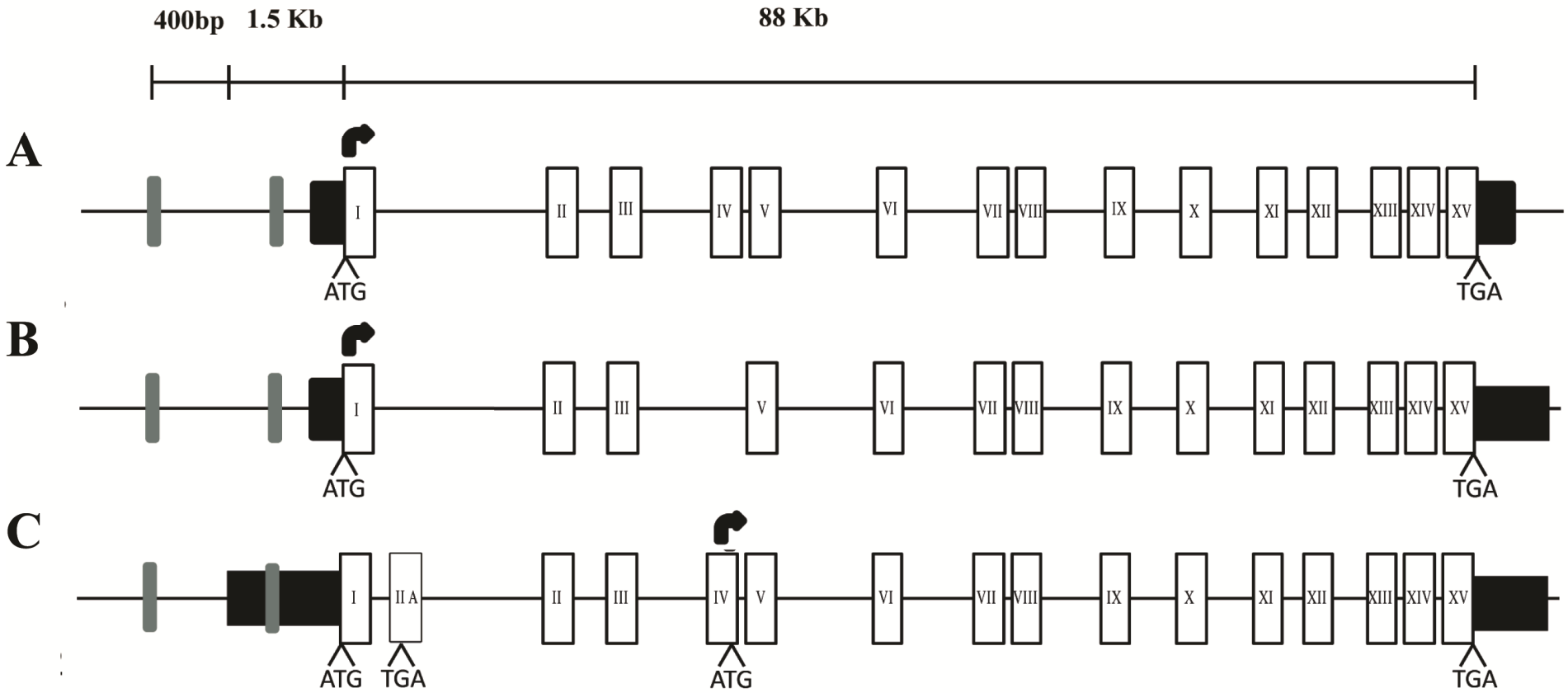
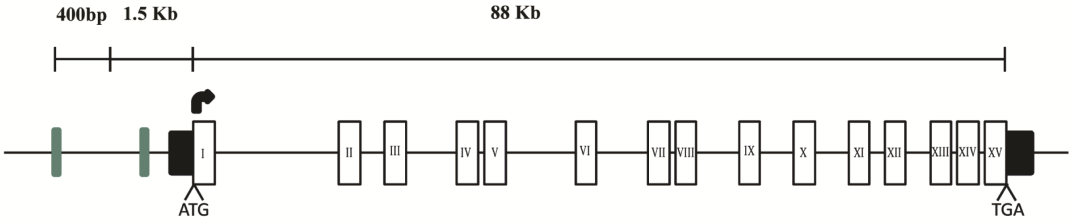


Figure 4.1 - Monoamine Oxidase A (*MAOA)* isoforms. Graphic representation of the *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19. Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site. Gray bars are the dVNTR and the uVNTR from left to right respectively. **A.** Major *MAOA* isoform (Long). **B.** *MAOA* splicing variant missing Exon IV (Medium). **C.** *MAOA* minor isoform (Short).

The *MAOA* L variant (**Figure 4.1 A**) represents the *MAOA* major isoform, it is comprised of a 4090 bp mRNA and is composed of 527 amino acids (AA). The Hg19 database and AceView web browser, at the time of writing, identify the 5′ UTR of this isoform as identical to the isoform S, where the mRNA sequence comprehensive of the 5′ UTR is 5330 bp long (**Figures 4.2 – 4.3**). In the older database versions this spanned from bp +1 to +181. Therefore MAOA uVNTR, since its first characterization in 1998 (Sabol *et al.*, 1998) was considered more than 1kb away from the first, shorter, transcriptional start site.

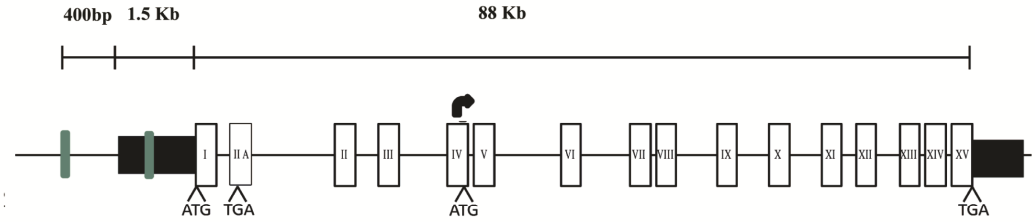
On the contrary the *MAOA* isoform S (**Figure 4.1 C**) has a longer 5′ UTR (1446bp), thus contains the uVNTR within the UTR. The major feature of this isoform is the presence of an alternative exon (Exon IIA) that produces a truncated version of the MAOA protein, exactly 133 AA shorter than the Long isoform. As illustrated in **Figure 4.1 C**, the alternative Exon IIA contains a (TGA) stop codon in the reading frame which makes the translational start site shift to Exon IV where the next in frame methionine codon (ATG) is located. The resulting expected protein is 394 AA long. The *MAOA* S isoform lacks residues 1-133, thus an important piece of the FAD binding domain. Because the cofactor FAD is necessary to the MAOA protein to exert its activity, this protein should be inactive. Unfortunately the literature does not report any evidence for this particular protein isoform. Only the mRNA has been confirmed for this isoform and the putative protein sequence deduced.

Finally the *MAOA* M isoform (**Figure 4.1 B**) has been identified in my current studies by PCR of cDNA (**Figure 4.21**). This isoform lacks Exon IV, from residue 102 to 137, which overlaps with a section of the substrate binding domain. Again no literature on this isoform is present and not even the splicing variant has been identified before.




```
   1 caggagcgtg tcagccaaag catggagaat caagagaagg cgagtatcgc gggccacatg
  61 ttcgacgtag tcgtgatcgg aggtggcatt tcaggactat ctgctgccaa actcttgact
 121 gaatatggcg ttagtgtttt ggttttagaa gctcgggaca gggttggagg aagaacatat
 181 actataagga atgagcatgt tgattacgta gatgttggtg gagcttatgt gggaccaacc
 241 caaaacagaa tcttacgctt gtctaaggag ctgggcatag agacttacaa agtgaatgtc
 301 agtgagcgtc tcgttcaata tgtcaagggg aaaacatatc catttcgggg cgcctttcca
 361 ccagtatgga atcccattgc atatttggat tacaataatc tgtggaggac aatagataac
 421 atggggaagg agattccaac tgatgcaccc tgggaggctc aacatgctga caaatgggac
 481 aaaatgacca tgaaagagct cattgacaaa atctgctgga caaagactgc taggcggttt
 541 gcttatcttt ttgtgaatat caatgtgacc tctgagcctc acgaagtgtc tgccctgtgg
 601 ttcttgtggt atgtgaagca gtgcgggggc accactcgga tattctctgt caccaatggt
 661 ggccaggaac ggaagtttgt aggtggatct ggtcaagtga gcgaacggat aatggacctc
 721 ctcggagacc aagtgaagct gaaccatcct gtcactcacg ttgaccagtc aagtgacaac
 781 atcatcatag agacgctgaa ccatgaacat tatgagtgca aatacgtaat taatgcgatc
 841 cctccgacct tgactgccaa gattcacttc agaccagagc ttccagcaga gagaaaccag
 901 ttaattcagc ggcttccaat gggagctgtc attaagtgca tgatgtatta caaggaggcc
 961 ttctggaaga agaaggatta ctgtggctgc atgatcattg aagatgaaga tgctccaatt
1021 tcaataacct tggatgacac caagccagat gggtcactgc ctgccatcat gggcttcatt
1081 cttgcccgga aagctgatcg acttgctaag ctacataagg aaataaggaa gaagaaaatc
1141 tgtgagctct atgccaaagt gctgggatcc caagaagctt tacatccagt gcattatgaa
1201 gagaagaact ggtgtgagga gcagtactct gggggctgct acacggccta cttcctcct
1261 gggatcatga ctcaatatgg aagggtgatt cgtcaacccg tgggcaggat tttctttgcg
1321 ggcacagaga ctgccacaaa gtggagcggc tacatggaag gggcagttga ggctggagaa
1381 cgagcagcta gggaggtctt aaatggtctc gggaaggtga ccgagaaaga tatctgggta
1441 caagaacctg aatcaaagga cgttccagcg gtagaaatca cccacacctt ctgggaaagg
1501 aacctgccct ctgtttctgg cctgctgaag atcattggat tttccacatc agtaactgcc
1561 ctggggtttg tgctgtacaa atacaagctc ctgccacggt cttgaagttc tgttcttatg
1621 ctctctgctc actggttttc aataccacca agaggaaaat attgacaagt ttaaaggctg
1681 tgtcattggg ccatgtttaa gtgtactgga tttaactacc tttggcttaa ttccaatcat
1741 tgttaaagta aaaacaattc aaagaatcac ctaattaatt tcagtaagat caagctccat
1801 cttatttgtc agtgtagatc aactcatgtt aattgataga ataaagcctt gtgatcactt
1861 tctgaaattc acaaagttaa acgtgatgtg ctcatcagaa acaaaaaaaa aaaaaaaaaa
1921 aa
```

Figure 4.2 - Monoamine Oxidase A (*MAOA*) mRNA Long isoform. mRNA sequence of the canonical MAOA protein. Long isoform in this manuscript. NCBI Accession number: BC008064 (Version BC008064.2) (http://www.ncbi.nlm.nih.gov/nuccore/BC008064). At the top illustration of MAOA gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19 (GENECODE v24 track). Black boxes represent 5′ and 3′ UTRs, white boxes the exons. Curved black arrow sets the translational start site for the conventional MAOA protein (long isoform).




```
   1 cccactcctg tgcctacgac ccaggagcgt gtcagccaaa gcatggagaa tcaagagaag
  61 gcgagtatcg cgggccacat gttcgacgta gtcgtgatcg gaggtggcat ttcaggccaa
 121 gagaagagat tggcatgaac tgttgctttg ccagttccag ccagtaccag caaaaacata
 181 catgtatgtt ggagcatcag aggaaagcag ctcctggtgg agagactatc tgctgccaaa
 241 ctcttgactg aatatggcgt tagtgttttg gttttagaag ctcgggacag ggttggagga
 301 agaacatata ctataaggaa tgagcatgtt gattacgtag atgttggtgg agcttatgtg
 361 ggaccaaccc aaaacagaat cttacgcttg tctaaggagc tgggcataga gacttacaaa
 421 gtgaatgtca gtgagcgtct cgttcaatat gtcaagggga aaacatatcc atttcggggc
 481 gcctttccac cagtatggaa tcccattgca tatttggatt acaataatct gtggaggaca
 541 atagataaca tggggaagga gattccaact gatgcaccct gggaggctca acatgctgac
 601 aaatgggaca aaatgaccat gaaagagctc attgacaaaa tctgctggac aaagactgct
 661 aggcggtttg cttatctttt tgtgaatatc aatgtgacct ctgagcctca cgaagtgtct
 721 gccctgtggt tcttgtggta tgtgaagcag tgcgggggca ccactcggat attctctgtc
 781 accaatggtg gccaggaacg gaagtttgta ggtggatctg gtcaagtgag cgaacggata
 841 atggacctcc tcggagacca agtgaagctg aaccatcctg tcactcacgt tgaccagtca
 901 agtgacaaca tcatcataga gacgctgaac catgaacatt atgagtgcaa atacgtaatt
 961 aatgcgatcc ctccgacctt gactgccaag attcacttca gaccagagct tccagcagag
1021 agaaaccagt taattcagcg tcttccaatg ggagctgtca ttaagtgcat gatgtattac
1081 aaggaggcct tctggaagaa gaaggattac tgtggctgca tgatcattga agatgaagat
1141 gctccaattt caataacctt ggatgacacc aagccagatg ggtcactgcc tgccatcatg
1201 ggcttcattc ttgcccggaa agctgatcga cttgctaagc tacataagga aataaggaag
1261 aagaaaatct gtgagctcta tgccaaagtg ctgggatccc aagaagcttt acatccagtg
1321 cattatgaag agaagaactg gtgtgaggag cagtactctg ggggctgcta cacggcctac
1381 ttccctcctg ggatcatgac tcaatatgga agggtgattc gtcaacccgt gggcaggatt
1441 ttctttgcgg gcacagagac tgccacaaag tggagcggct acatggaagg ggcagttgag
1501 gctggagaac gagcagctag ggaggtctta aatggtctcg ggaaggtgac cgagaaagac
1561 atctgggtac aagaacctga atcaaaggac gttccagcgg tagaaatcac ccacaccttc
1621 tgggaaagga acctgccctc tgtttctggc ctgctgaaga tcattggatt ttccacatca
1681 gtaactgccc tggggtttgt gctgtacaaa tacaagctcc tgccacggtc ttgaagttct
1741 gttcttatgc tctctgctca ctggttttca ataccaccaa gaggaaaata ttgacaagtt
1801 taaaggctgt gtcattgggc catgtttaag tgtactggat ttaactacct ttggcttaat
1861 tccaatcatt gttaaagtaa aaacaattca aagaatcacc taattaattt cagtaagatc
1921 aagctccatc ttatttgtca gtgtagatca actcatgtta attgatagaa taaagccttg
1981 tgatcacttt ctgaaattca caaagttaaa cgtgatgtgc tcatcagaaa caatttctgt
2041 gtcctgtttt tattcccttc aatgcaaaat acatgatgat ttcagaaaca aagcatttga
2101 ctttctgtct gtggaggtgg agtaggtgaa ggcccagcct gtaactgtcc ttttttcttcc
2161 cttaggcaat ggtgaactgt cattacagag cctagaggct cacagcctcc tggaggaagc
2221 agcctccact ttggatcagg aaatagtaaa ggaaagcagt gttgggggta gcggcatgca
2281 gaccctcaga ccagaatggg gacatcttgt ggtctgctgc ctcaggaatc tcctgaccac
2341 ttgtagtccc tccgacttct ctagacatct agtctcagtg cta
```

Figure 4.3 - Monoamine Oxidase A (*MAOA*) mRNA Short isoform. mRNA sequence of the short MAOA protein. Short isoform in this manuscript. NCBI Accession number: AK293926 (Version AK293926.1). At the top illustration of MAOA gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19 (GENECODE v24 track). Black boxes represent 5′ and 3′ UTRs, white boxes the exons. Curved black arrow sets the translational start site for the short MAOA protein (short isoform).

A 3D representation of the putative MAOA protein isoforms is illustrated in **Figure 4.4 B – D**. The reconstruction for each isoform has been obtained from http://raptorx.uchicago.edu website. The input sequences can be found in the **Figures 4.5 – 4.5 – 4.6**. **Figure 4.4 A** illustrate the representation of MAOA protein obtained after crystallisation (De Colibus *et al.*, 2005) highlighting in blue the FAD binding domains and in red the substrate binding domains.

The comparison between the **Figure 4.4 A** and **4.4 B** both representing the major MAOA isoform (L) shows high level of similarity which increase the reliability about the possible structure of the other MAOA isoforms deduced in our study.

As expected, several structural changes can be seen in the M and S isoform (**Figure 4.4 C – D**). In the M isoform the FAD binding domain appear intact whilst the structural change can be seen in the substrate binding domain appearing substantially reduced compared to the L isoform. On the contrary the S isoform (**Figure 4.4 D**) suffered a radical structural change undergoing the deletion of a third of the residues that compose the FAD binding domain.

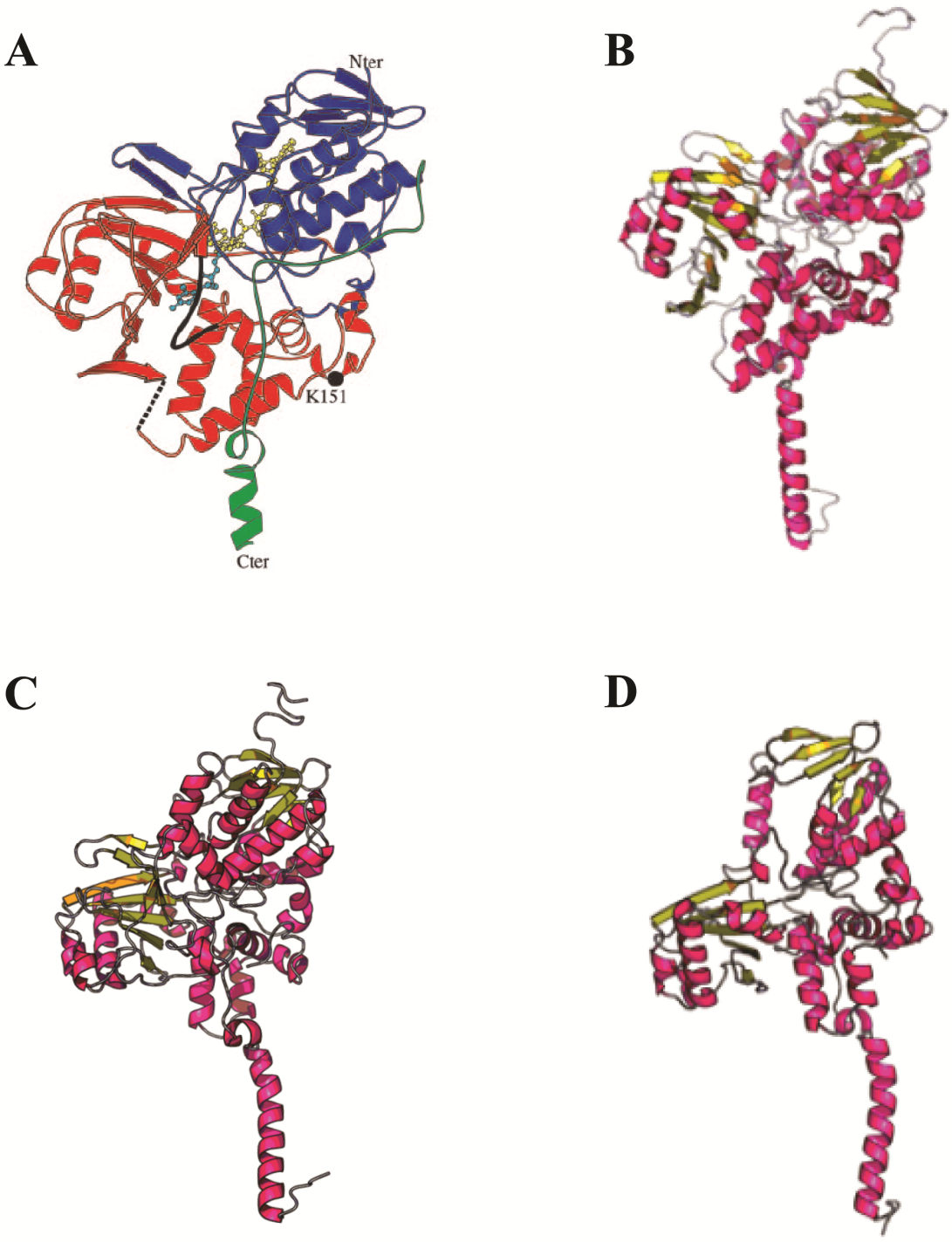
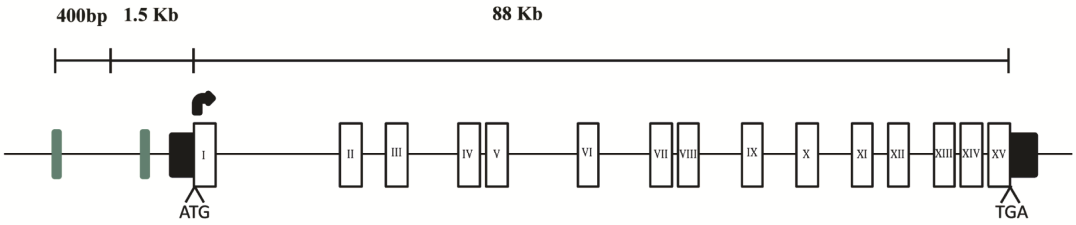
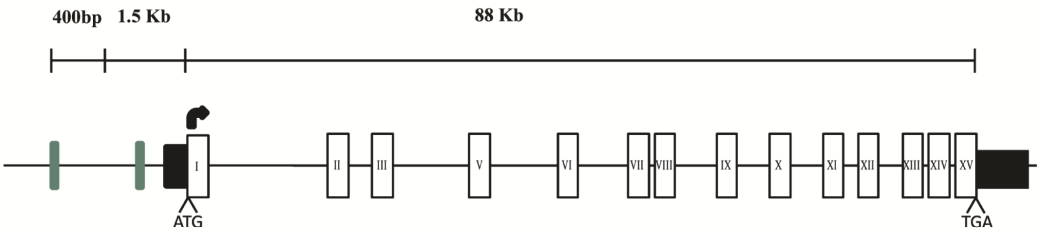


Figure 4.4 - Prediction of Monoamine Oxidase A (*MAOA*) protein isoforms. 3D predicted protein molecular structures of MAOA protein based on the mRNA sequences obtained from online databases (B-D) or sequencing data (C) using http://raptorx.uchicago.edu website. **A**. Image from De Colibus *et al.* (2005): in blue the FAD binding domain, in red substrate binding domain, membrane region (C-terminal) in green. In yellow and cyan ball-and-stick representation are FAD and clorgyline respectively. Black coil is the active site loop. **B**. Representation of the MAOA L isoform. **C**. **D**. Putative reconstruction of isoform M and S respectively.



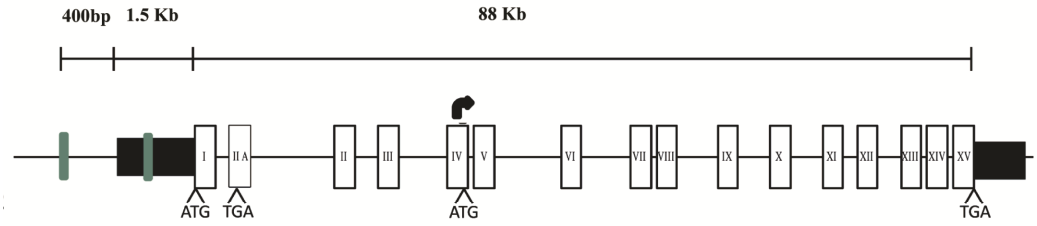
MENQEKASIAGHMFDVVVIGGGISGLSAAKLLTEYGVSVLVLEARDRV
GGRTYTIRNEHVDYVDVGGAYVGPTQNRILRLSKELGIETYKVNVSER
LVQYVKGKTYPFRGAFPPVWNPIAYLDYNNLWRTIDNMGKEIPTDAPW
EAQHADKWDKMTMKELIDKICWTKTARRFAYLFVNINVTSEPHEVSAL
WFLWYVKQCGGTTRIFSVTNGGQERKFVGGSGQVSERIMDLLGDQVKL
NHPVTHVDQSSDNIIIETLNHEHYECKYVINAIPPTLTAKIHFRPELP
AERNQLIQRLPMGAVIKCMMYYKEAFWKKKDYCGCMIIEDEDAPISIT
LDDTKPDGSLPAIMGFILARKADRLAKLHKEIRKKKICELYAKVLGSQ
EALHPVHYEEKNWCEEQYSGGCYTAYFPPGIMTQYGRVIRQPVGRIFF
AGTETATKWSGYMEGAVEAGERAAREVLNGLGKVTEKDIWVQEPESKD
VPAVEITHTFWERNLPSVSGLLKIIGFSTSVTALGFVLYKYKLLPRS

Figure 4.5 - Monoamine Oxidase A (*MAOA*) Protein Long isoform. Amino acid sequence of the MAOA long isoform protein used also to generate the image in **Figure 4.4**. At the top illustration of MAOA gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19 (GENECODE v24 track). Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site for the long MAOA protein (long isoform).



MENQEKASIAGHMFDVVVIGGGISGLSAAKLLTEYGVSVL
VLEARDRVGGRTYTIRNEHVDYVDVGGAYVGPTQNRILRL
SKELGIETYKVNVSERLVQYVKIPTDAPWEAQHADKWDKM
TMKELIDKICWTKTARRFAYLFVNINVTSEPHEVSALWFL
WYVKQCGGTTRIFSVTNGGQERKFVGGSGQVSERIMDLLG
DQVKLNHPVTHVDQSSDNIIIETLNHEHYECKYVINAIPP
TLTAKIHFRPELPAERNQLIQRLPMGAVIKCMMYYKEAFW
KKKDYCGCMIIEDEDAPISITLDDTKPDGSLPAIMGFILA
RKADRLAKLHKEIRKKKICELYAKVLGSQEALHPVHYEEK
NWCEEQYSGGCYTAYFPPGIMTQYGRVIRQPVGRIFFAGT
ETATKWSGYMEGAVEAGERAAREVLNGLGKVTEKDIWVQE
PESKDVPAVEITHTFWERNLPSVSGLLKIIGFSTSVTALG
FVLYKYKLLPRS

Figure 4.6 - Monoamine Oxidase A (*MAOA*) protein Medium isoform. Amino acid sequence of the MAOA medium isoform protein used also to generate the image in **Figure 4.4**. At the top illustration of MAOA gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19 (GENECODE v24 track). Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site for the medium MAOA protein (medium isoform).



MGKEIPTDAPWEAQHADKWDKMTMKELIDKICWTKTARR
FAYLFVNINVTSEPHEVSALWFLWYVKQCGGTTRIFSVT
NGGQERKFVGGSGQVSERIMDLLGDQVKLNHPVTHVDQS
SDNIIIETLNHEHYECKYVINAIPPTLTAKIHFRPELPA
ERNQLIQRLPMGAVIKCMMYYKEAFWKKKDYCGCMIIED
EDAPISITLDDTKPDGSLPAIMGFILARKADRLAKLHKE
IRKKKICELYAKVLGSQEALHPVHYEEKNWCEEQYSGGC
YTAYFPPGIMTQYGRVIRQPVGRIFFAGTETATKWSGYM
EGAVEAGERAAREVLNGLGKVTEKDIWVQEPESKDVPAV
EITHTFWERNLPSVSGLLKIIGFSTSVTALGFVLYKYKL
LPRS

Figure 4.7 - Monoamine Oxidase A (*MAOA*) protein Short isoform. Amino acid sequence of the MAOA short isoform protein used also to generate the image in **Figure 4.4**.At the top illustration of MAOA gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19 (GENECODE v24 track). Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site for the short MAOA protein (short isoform).

4.3.1.2 Conservation analysis of *MAOA* isoforms – interested exons

I performed a conservation analysis on the Exon IIA region (chrX:43,517,110-43,517,217) and the 5′ and 3′ flanking regions in humans, non-human primates and rodents. The sequences from each species were extrapolated from UCSC web browser Hg19.

The alignment of the sequences obtained from the 100 vertebrates conservation tool in UCSC web browser shows few differences in non-human primates compared to the human sequence (**Figure 4.8**). Thus the TGA stop codon in the human Exon

IIA sequence, that generates the truncated MAOA protein, could be expressed in these species as well. However in non human primates, according to UCSC and ECR web browser, this genomic region is not considered an exon that can be spliced in the *MAOA* sequence.

In rodents this sequence appears significantly different. Both mouse and rat sequences do not show the TGA stop codon in their reading frame and the overall sequence presents significant changes compared to human and non-human primates.

MAOA Exon IV, as illustrated in **Figure 4.9**, shows a very high degree of conservation among species. Some point mutations are present in the DNA sequences but they are all synonymous mutations except for one that in rodents and marmoset encode for a methionine instead of an isoleucine. As stated before, this exon is part of the substrate binding domain in the fully functional MAOA protein. In order to maintain its activity it has to be expected an high degree of conservation of this genomic region among species.

This analysis is in accordance with Andres *et al.* (2004), where the *MAOA* gene sequence has been compared in human and non human primates. Specifically for one of the regions analysed, comprising Exons IV – V and Exons VII – VIII, humans show divergence levels from other species no higher than 1.55% and the lower level is with chimpanzees for a 1.06% divergence. This is also in accordance with Ebersberger *et al.* (2002), where in their study they analysed divergence levels of parts of all the chromosomes in humans and chimpanzees and they concluded that the whole X chromosome has accumulated the least amount of differences (1.0%) compared to the autosomes.

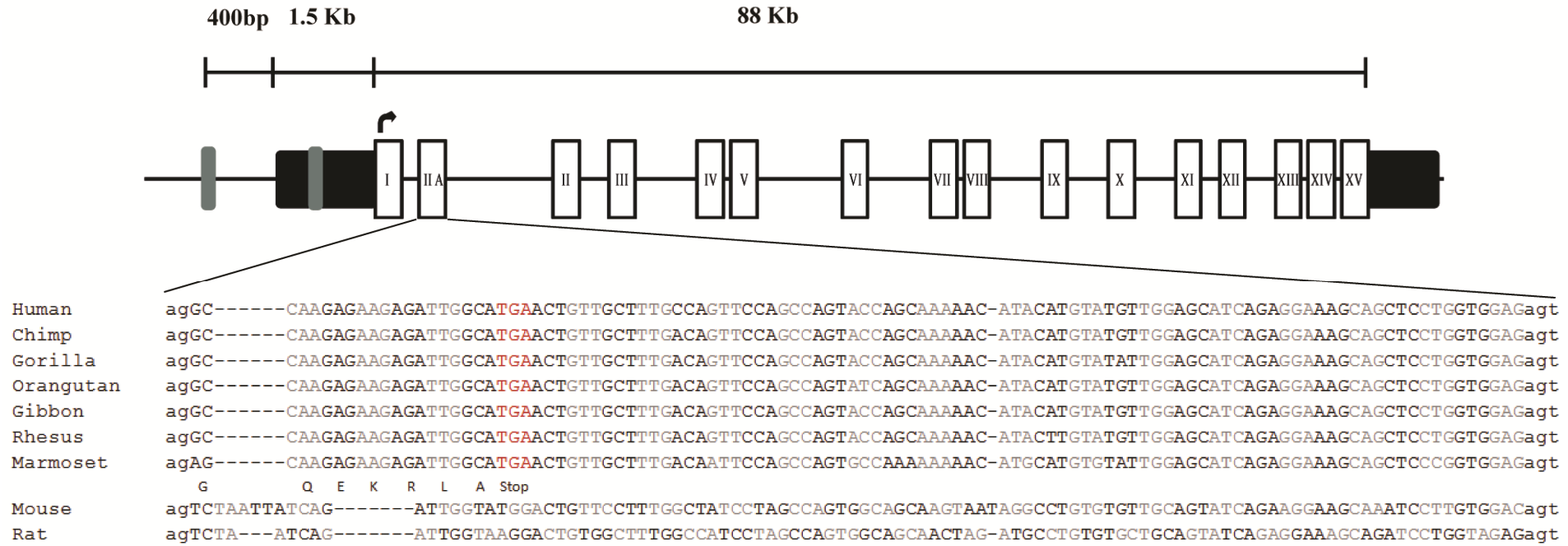


Figure 4.8 - Monoamine Oxidase A (*MAOA)* Exon IIA conservation analysis. At the top graphic representation of the *MAOA* gene. White boxes: exons, black boxes: untranslated regions (UTRs), gray boxes: dVNTR and uVNTR from left to right respectively. Underneath alignment of *MAOA* Exon IIA in different species. Alternate codons are highlighted in gray or black. Amino acids translation is represented by corresponding letters underneath DNA sequences. TGA stop codon is termed as Stop.

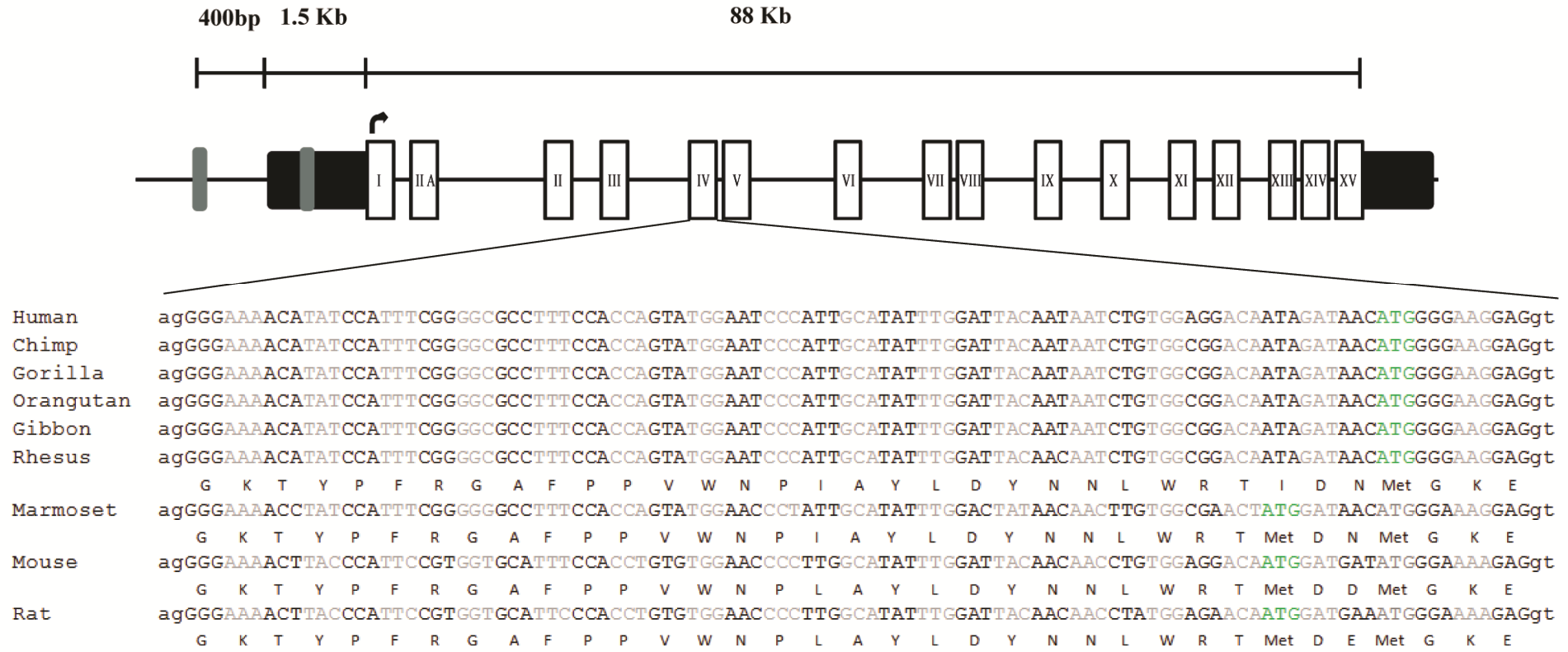


Figure 4.9 - Monoamine Oxidase A (*MAOA)* Exon IV conservation analysis. At the top graphic representation of the *MAOA* gene. White boxes: exons, black boxes: untranslated regions (UTRs), gray boxes: dVNTR and uVNTR from left to right respectively. Underneath alignment of *MAOA* Exon IV in different species. Alternate codons are highlighted in gray or black. Amino acids translation is represented by corresponding letters underneath DNA sequences. ATG codons are termed as Met.

4.3.2 HAP 1 cell lines characterisation

In total we obtained 9 cell line clones from Horizon. As explained in **Figure 4.10** from the parental cell line (P) were generated, after a first CRISPR/Cas9 deletion, four different single KO cell lines. Two of them have the uVNTR deleted from the *MAOA* promoter region and still possess the dVNTR. These are clone 9_F4 and 9_E2 termed A and B respectively for simplicity. The other two clones have the dVNTR deleted and still possess the uVNTR and they are clones 13_B5 and 13_B1 termed C and D respectively.

Using clones B and C as parental cell line to generate a second mutation, the double KO clones du_B5 (E), du_F3 (F), ud_D8 (G) and ud_F3 (H) were generated applying a second, *de-novo*, CRISPR/Cas9 deletion. These second deletion clones are E and F using clone B as parental cell line and clones G and H using clone C as the parental cell line.

The deletions were confirmed by PCR using the primer sets for *MAOA* uVNTR and dVNTR respectively as shown in **Figure 4.12 A – B**. SH-SY5Y cell line has been used as positive control for the PCR and also to identify the genotype of the HAP1 cell line for both VNTRs since SH-SY5Y carries the two most common alleles for both VNTRs. The 3R and 4R for the uVNTR and the 9R and 10R for the dVNTR. HAP1 cell line, after the comparison with SH-SY5Y, turned out to carry the 3R and 10R polymorphism for uVNTR and dVNTR respectively.

Both the agarose gel and the virtual gel generated by the Qiaxcel confirmed the deletion for all the cell line for both uVNTR and dVNTR single KO and the four double KO cell lines.

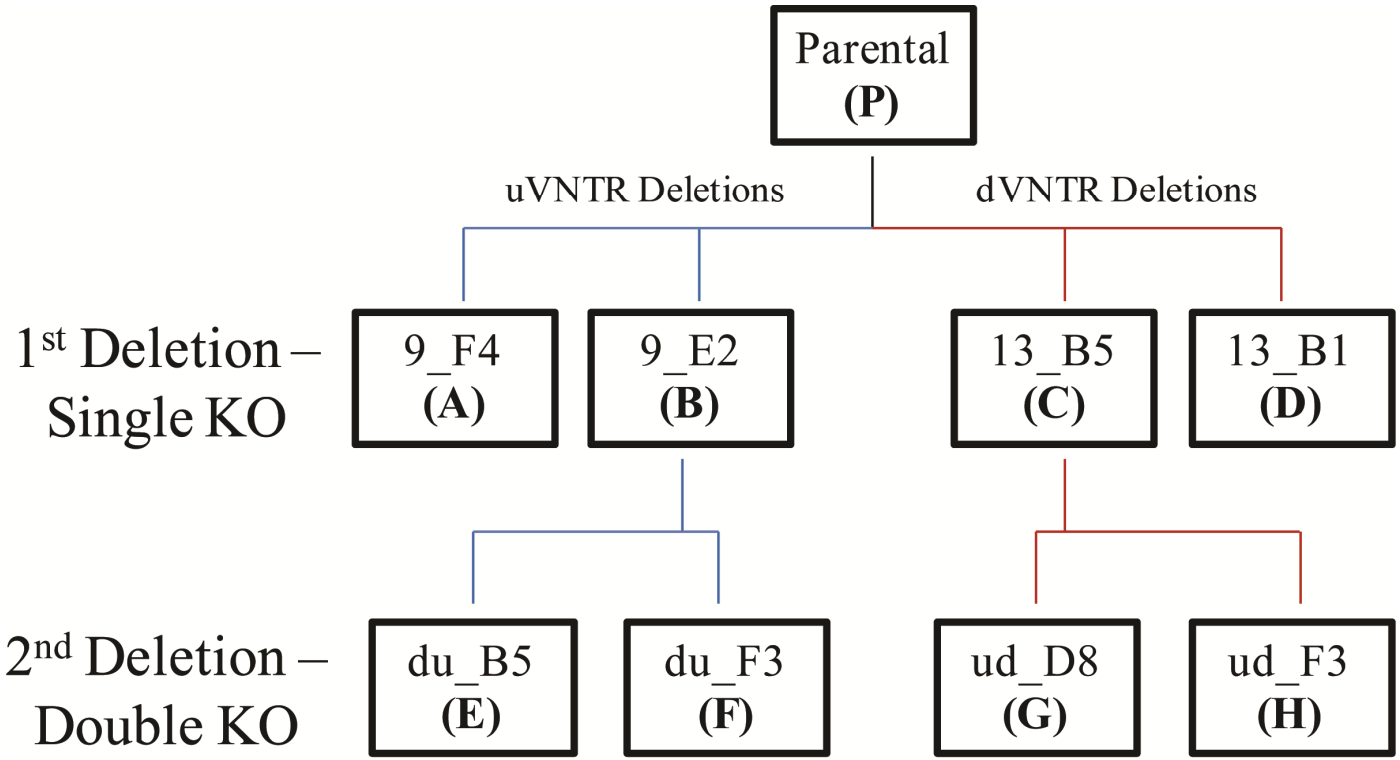


Figure 4.10 - HAP1 cell line genetic tree. Genetic tree explaining the different HAP1 cell line clones. Parental cell line (P) at the top has been used to generate the single KO cell lines for uVNTR, on the left side (blue branch) and for dVNTR on the right side (red branch). Clones B and C has been used as parental cell line to generate a second KO that brought to the double KO cell lines (E to H) which lack both uVNTR and dVNTR.

Figure 4.11 confirms that the DNA deletion has been successful and the actual sizes in the gels correspond to the expected ones (data file information from Horizon). However, a substantial difference of 43 bp can be observed between clones E, F and clones G, H for the uVNTR locus whilst there is no difference in the dVNTR one. The sequences alignment in **Figure 4.12** confirms this difference between clones observed in the gel, thus excluding the possibility of the agarose gel artefact.

The most plausible explanation is the parental cell line that each double KO set derives from. Again, regarding the uVNTR locus, clones G and H show a difference of 4 bp between themselves while clones E and F have the same sequence. No difference has been observed in the dVNTR locus among any of the clones.

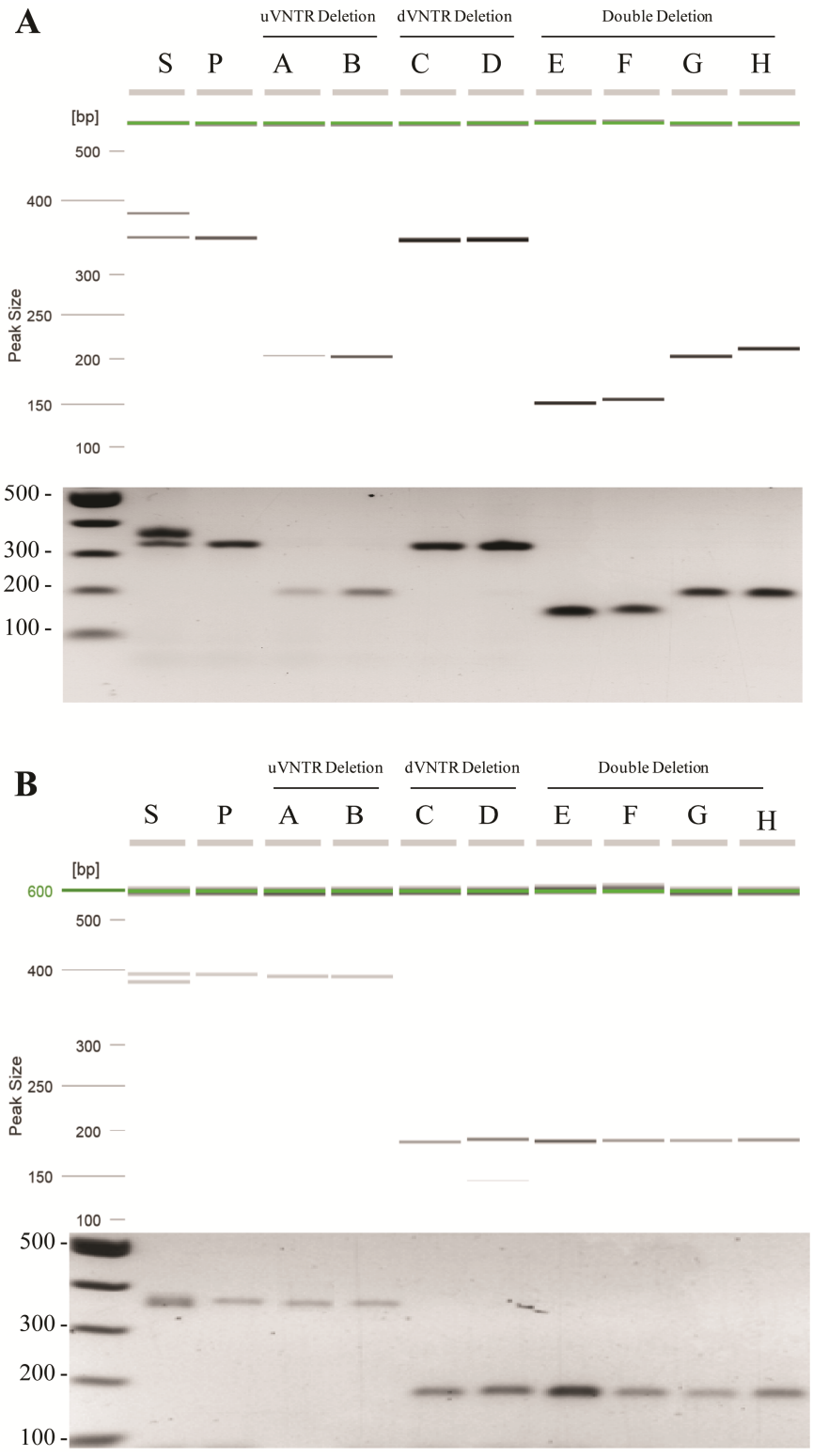


Figure 4.11 - HAP1 cell line genotype. Genotype of HAP1 cell line clones confirming the deletion of the interested genetic region. **A** - PCR reaction amplifying uVNTR genetic region. **B** - reaction amplifying dVNTR genetic region. Letters on top refer to **Figure 4.1** for the clones.

```
Clone_E      ------------------------------NTGCGGGAAGCAGAACACCGCCCCCAG
Clone_F      GGGTACCGAGAACAGCCTGACCGTGGAGAAGGGCTGCGGGAAGCAGAACACCGCCCCCAG
Clone_G      GGGTACCGAGAACAGCCTGACCGTGGAGAAGGGCTGCGGGAAGCAGAACACCGCCCCCAG
Clone_H      GGGTACCGAGAACAGCCTGACCGTGGAGAAGGGCTGCGGGAAGCAGAACACCGCCCCCAG
UCSC         GGGTACCGAGAACAGCCTGACCGTGGAGAAGGGCTGCGGGAAGCAGAACACCGCCCCCAG
                                           ****************************

Clone_E      CGCCCAGCGTGCTCCAGAAACATGAGCACAAACGCCTCAGCCTCCTTCCCCGGCGGCACC
Clone_F      CGCCCAGCGNGCTCCANAAACATGAGNACAAACGCCTCAGCCTCCTTCCCCGGCGGCACC
Clone_G      CGCCCAGCGTGCTCCAGAAA-----------------------------------------
Clone_H      CGCCCAGCGTGCTCCA--------------------------------------------
UCSC         CGCCCAGCGTGCTCCAGAAACATGAGCACAAACGCCTCAGCCTCCTTCCCCGGCGGCACC

Clone_E      GGAGCCGGGGGCACAACTGCCCAGGTCCCGAACCCGGACTCCAG----------------
Clone_F      GGAGCCGGGGGCACAACTGNNCAGGNCCCGAACCCGGACTCCAG----------------
Clone_G      ------------------------------------------------------------
Clone_H      ------------------------------------------------------------
UCSC         GGCACCGGCACCAGTACCCGCACCAGTACCGGCACCGGCACCAGTACCCGCACCAGTACC

Clone_E      ------------------------------------------------------------
Clone_F      ------------------------------------------------------------
Clone_G      ------------------------------------------------------------
Clone_H      ------------------------------------------------------------
UCSC         GGCACCGGCACCAGTACCCGCACCAGTACCGGCACCGGCACCGAGCGCAAGGCGGAGGGC

Clone_E      ---------------------------------------------------CTTGGACGAC
Clone_F      ---------------------------------------------------CTTGNACGAC
Clone_G      ---------GCCGGGGGCACAACTGCCCAGGTCCCGAACCCGGACTCCAGCTTGGACGAC
Clone_H      ---------GCCGGGGGCACAACTGCCCAGGTCCCGAACCCGGACTCCAGCTTGGACGAC
UCSC         CCGCCCGAAGCCGGGGGCACAACTGCCCAGGTCCCGAACCCGGACTCCAGCTTGGACGAC
                                                           ***********

Clone_E      ACCTCCTACAGCCTGTCCGAATGGAGCGTCCGTTCTGAGTGGCGGTCCGTCTCGGATCCG
Clone_F      ACCTCCTACAGCCTGTNCNAATGNANCGNCCGTTCTGAGTGGCGNTCCGTCTCGGATCCG
Clone_G      ACCTCCTACAGCCTGTCCGAATGGAGCGTCCGTTCTGAGTGGCGGTCCGTCTCGGATCCG
Clone_H      ACCTCCTACAGCCTGTCCGAATGGAGCGTCCGTTCTGAGTGGCGGTCCGTCTCGGATCCG
UCSC         ACCTCCTACAGCCTGTCCGAATGGAGCGTCCGTTCTGAGTGGCGGTCCGTCTCGGATCCG
             ***************** * **** * ** ***************** ***************

Clone_E      CTAGCCAGTTCCCAGTGG------------------------------------------
Clone_F      CTAGCCAGTTCCCAGTGGAGCA-------------------------------------
Clone_G      CTAGCCAGTTCCCAGTGGAGCACGTCCTCNACTGCCGAGGCCGCCTCCTGGAGCTCCAGC
Clone_H      CTAGCCAGTTCCCAGTGGAGCACGTCCTCAACTGCCGAGGCCGCCTCCTGGAGCTCCAGC
UCSC         CTAGCCAGTTCCCAGTGGAGCACGTCCTCAACTGCCGAGGCCGCCTCCTGGAGCTCCAGC
             ******************
```

Figure 4.12 - HAP1 double KO alignment. Alignment of the HAP1 double KO clones E, F, G and H after sequencing of each clone.

4.3.3 *MAOA* gene expression

To assess any effect of uVNTR and dVNTR deletions on the *MAOA* expression in HAP1 cell line, the RNA has been extracted and converted to cDNA. Different primer sets have been used in order to assess the expression of the different predicted *MAOA* mRNA isoforms. The primer set that amplifies from Exon III to Exon VI of *MAOA* cover the expression of both long and short *MAOA* isoforms, thus giving insight on the total *MAOA* RNA that has been expressed.

The comparison between the parental cell line and the uVNTR single KO, illustrated in **Figure 4.13 A**, highlights a slight difference in the expression supported by the two different clones: clone A shows a slight increase of the *MAOA* expression while clone B presents a small reduction. Both of them are not significantly different from the expression of the parental cell line according to the Student's T-Test. However, the difference between the two single KO clones is statistically significant ($p<0.05$). **Figure 4.13 B** with its representation of a cartoon mechanism model summarises the effect of the uVNTR deletion on the expression of this mRNA which appears to be completely neutral.

However, the dVNTR KO clones C and D appear to support *MAOA* expression differently from the uVNTR KO clones. Both cell lines display a significant reduction ($p<0.05$) of expression compared to the parental cell line (**Figure 4.14 A**). This trend strongly suggests a major role of the dVNTR in the regulation pattern of this *MAOA* isoform expression compared to the major accredited risk factor uVNTR. The model represented in **Figure 4.14 B** displays the positive regulatory effect of the dVNTR on this *MAOA* isoform.

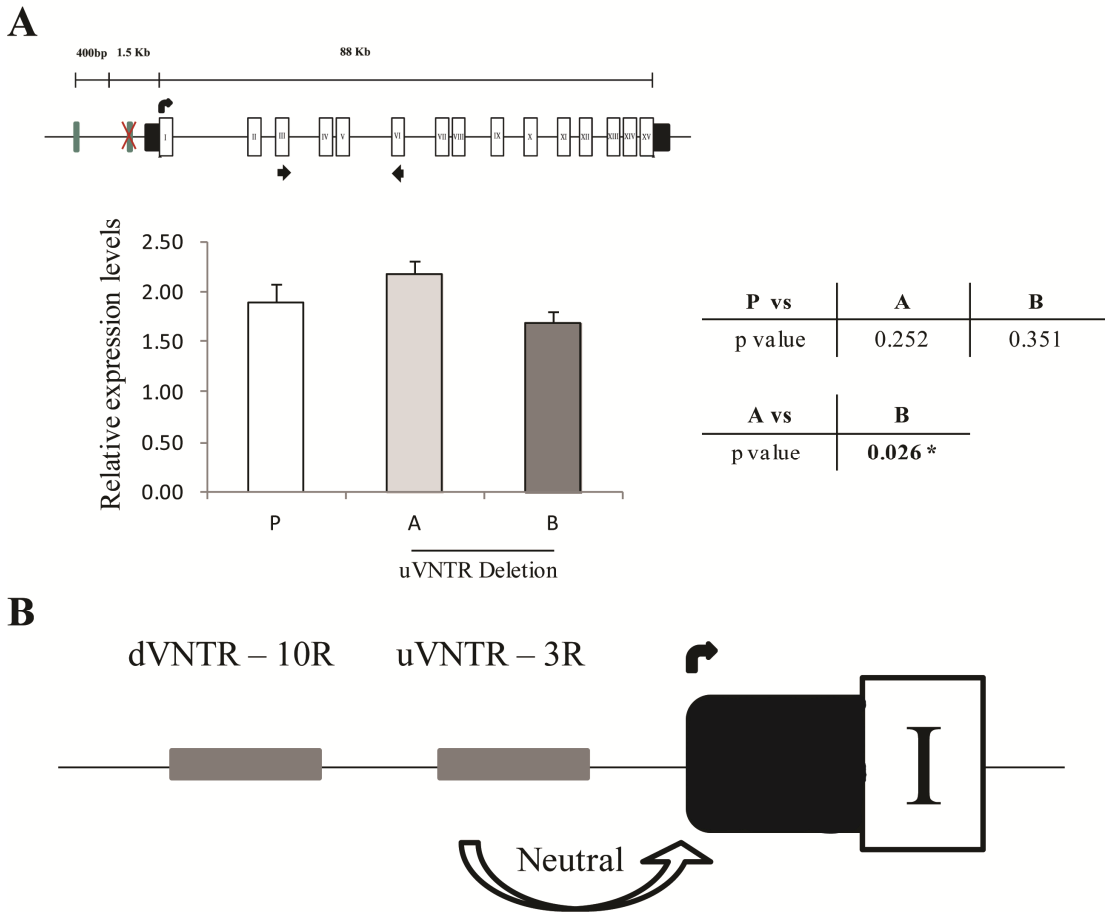


Figure 4.13 – Monoamine Oxidase A (*MAOA*) expression in HAP1 cell line – uVNTR deletion clones. A. Relative expression level of *MAOA* mRNA at basal condition. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19 (GENECODE v24 track). Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the conventional MAOA protein (long isoform). Black straight arrows show the forward and reveres primers respectively. P is the parental cell line, A and B are the uVNTR single deletion clones: 9_F4 and 9_E2 respectively. The table at the side provides the p-values obtained with the two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001. For each clone N=4. All values were normalised to β-Actin. **B.** Illustration of the *MAOA* promoter gene and the effect on the isoform expression of the uVNTR. Curved black line represents the TSS.

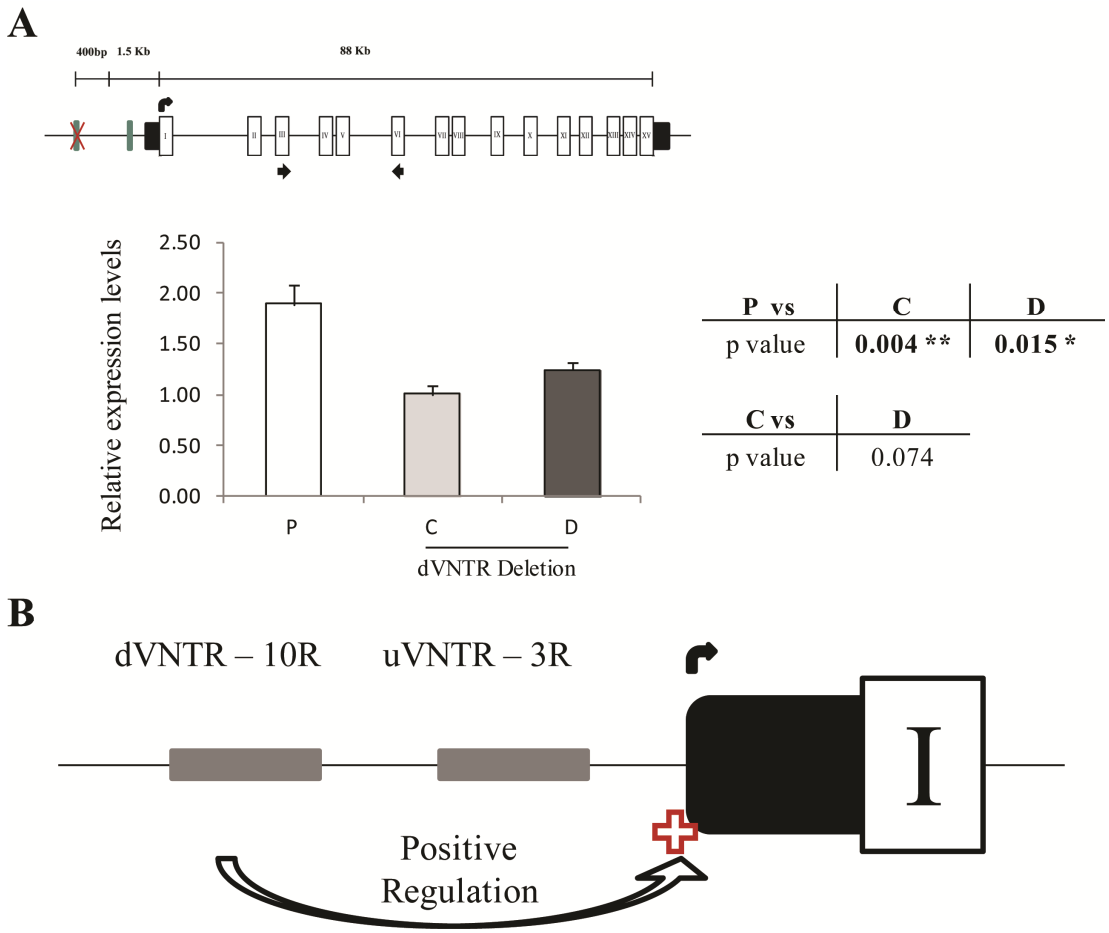


Figure 4.14 - Monoamine Oxidase A (*MAOA*) expression in HAP1 cell line – dVNTR deletion clones. Relative expression level of *MAOA* mRNA at basal condition. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19 (GENECODE v24 track). Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the conventional MAOA protein (long isoform). Black straight arrows show the forward and reveres primers respectively. P is the parental cell line, C and D are the dVNTR single deletion clones: 13_B5 and 13_B1 respectively. The table at the side provides the p-values obtained with the two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001. For each clone N=4. All values were normalised to β-Actin. **B.** Illustration of the *MAOA* promoter gene and the effect on the isoform expression of the dVNTR. Curved black line represents the TSS.

In all the four double KO cell lines that have been tested, there is a significant reduction ($p<0.01$) on the *MAOA* mRNA expression compared to parental cell line and the expression trend among them is equal as illustrated in **Figure 4.15 A**.

Figure 4.16 shows the expression patterns of the single KO clones B and C and the double KO clones that have been generated from them, compared to the parental cell line. In **Figure 4.16 A** it can be observed the effect of the uVNTR deletion compared to the double KO clones. As also reported in **Figure 4.19 A** the uVNTR deletion does not modify the *MAOA* expression, however when also the dVNTR is deleted from the promoter sequence, the *MAOA* expression obtained from the double KO cell lines is significantly reduced ($p<0.05$) compared to both the parental cell line and the uVNTR deletion clone according to the Student's T-Test.

In **Figure 4.16 B** the deletion of the dVNTR is sufficient to significantly reduce ($p<0.01$) the expression of *MAOA* which is comparable to the expression obtained from the double KO cell lines.

The *MAOA* uVNTR has always been considered a positive regulator of the *MAOA* gene since Sabol *et al.* (1998) characterised this polymorphism with expression constructs, although the 3R VNTR was considered low expression compared to the 4R (high expression). In their experiment they compared the ability of the different polymorphic variants of the *MAOA* uVNTR to drive the luciferase expression.

However in our cell line model, the uVNTR deletion does not modulate the expression of this particular *MAOA* isoform, it is in fact neutral and not a positive regulator.

These data altogether give more insight on the complexity of the *MAOA* promoter region where a single element like the uVNTR or the dVNTR alone are not sufficient to explain the expression of this gene. In fact, the deletion of both does not preclude

the expression of the gene, although the amount of mRNA is significantly reduced compared to parental cell line.

To support this fact has also to be reminded that rodents do not possess the human (or primate) specific VNTRs but they are able to express *MAOA*. Unfortunately, the major advantage we exploited of this cell line is also a major limitation on the experiments. The haploid genotype of the HAP1 cell line is perfect to assess the function of a genetic variant such as *MAOA* u- or dVNTR, but it also does not allow us to compare the expression of *MAOA* with different copy variant of u- or dVNTR.

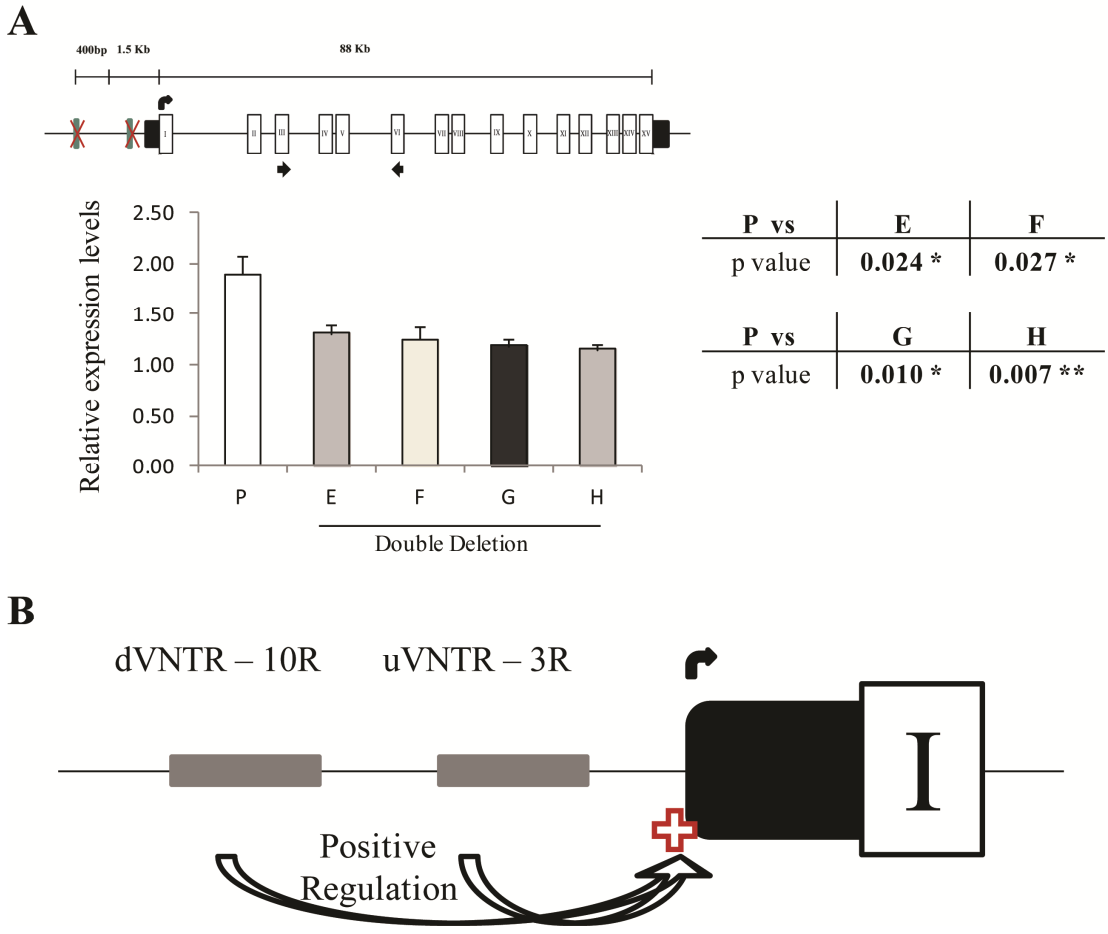


Figure 4.15 - Monoamine Oxidase A (*MAOA*) expression in HAP1 cell line – double deletion clones. Relative expression level of *MAOA* mRNA at basal condition. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19 (GENECODE v24 track). Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the conventional MAOA protein (long isoform). Black straight arrows show the forward and reveres primers respectively. 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 respectively. The table at the side provides the p-values obtained with the two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001. For each clone N=4. All values were normalised to β-Actin. **B.** Illustration of the *MAOA* promoter gene and the effect on the isoform expression of the dVNTR and uVNTR. Curved black line represents the TSS.

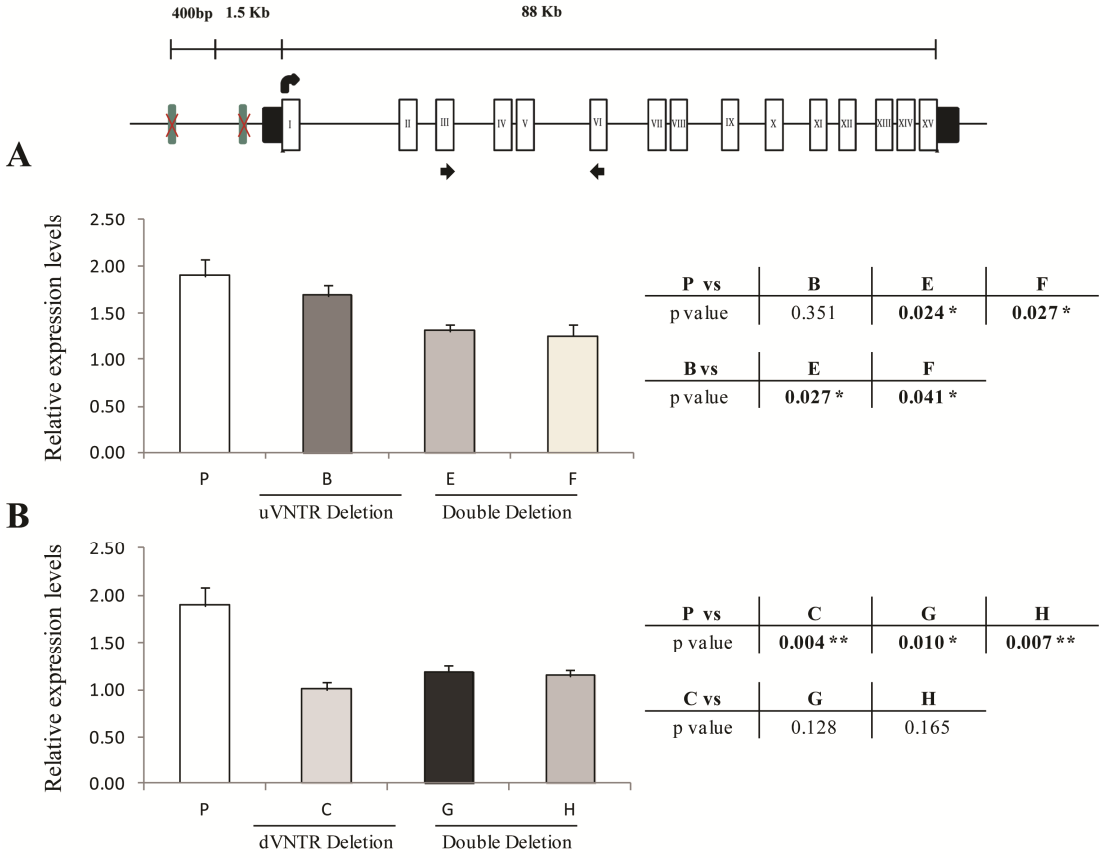


Figure 4.16 - Monoamine Oxidase A (*MAOA*) expression in HAP1 cell line – single deletion clones and related double KO clones. Relative expression level of *MAOA* mRNA at basal condition. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19. Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the conventional MAOA protein (long isoform). Black straight arrows show the forward and reveres primers respectively. **A.** P is the parental cell line, B single uVNTR deletion clone, 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 single dVNTR deletion clone, G and H are the *MAOA* VNTRs double KO clones: ud_D8 and ud_F3 respectively. The tables on the side provide the p-values obtained with the two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001. For each clone N=4. All values were normalised to β-Actin.

4.3.4 *MAOA* gene expression (Short isoform)

The element that allows us to differentiate between the long and short *MAOA* mRNA isoforms is the presence of the Exon IIA that contains a stop codon in the reading frame thus stopping the translation of the canonical MAOA protein generating a shorter protein.

The primer set that amplifies from Exon I to Exon IIA allow us to address only the expression of the short isoform since Exon IIA is present only in this isoform and not in the long one. This is illustrated in **Figure 4.17 A** for the uVNTR deletion clones. A complete opposite trend can be observed from the analysis of the short isoform expression where the parental cell lines express a low level of this mRNA (~8% of the total mRNA).

The expression obtained from the uVNTR single KO clones A and B is significantly higher compared to the parental cell line. No significant difference can be seen between the two clones. Similarly as observed for the uVNTR KO clones, the expression of both dVNTR KO clones, C and D, is significantly higher than the parental cell line as illustrated in **Figure 4.18 A**.

The deletion of either the uVNTR or the dVNTR seems sufficient to increase the expression of the *MAOA* short mRNA isoform highlighting the possibility of a positive regulatory element in both VNTRs specific for this mRNA isoform which in contrast can not be seen for the long isoform where only the dVNTR appears to be necessary to modulate the expression.

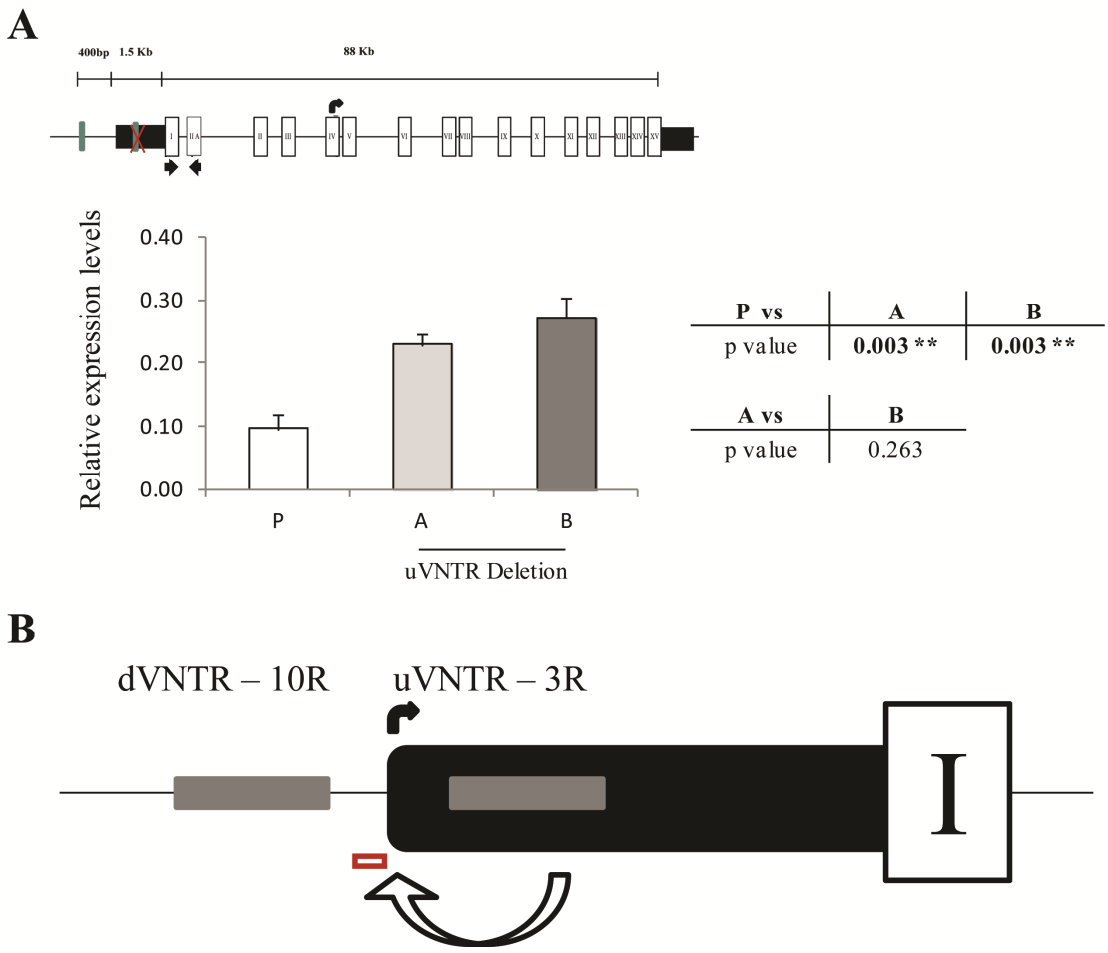


Figure 4.17 - Monoamine Oxidase A (*MAOA*) short isoform expression in HAP1 cell line – uVNTR deletion clones. Relative expression level of *MAOA* mRNA at basal condition. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19. Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the MAOA protein (short isoform). Black straight arrows show the forward and reveres primers respectively. P is the parental cell line, A and B are the uVNTR single deletion clones: 9_F4 and 9_E2 respectively. The table at the side provides the p-values obtained with the two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001. For each clone N=4. All values were normalised to β-Actin. **B.** Illustration of the *MAOA* promoter gene and the effect on the isoform expression of the uVNTR. Curved black line represents the TSS.

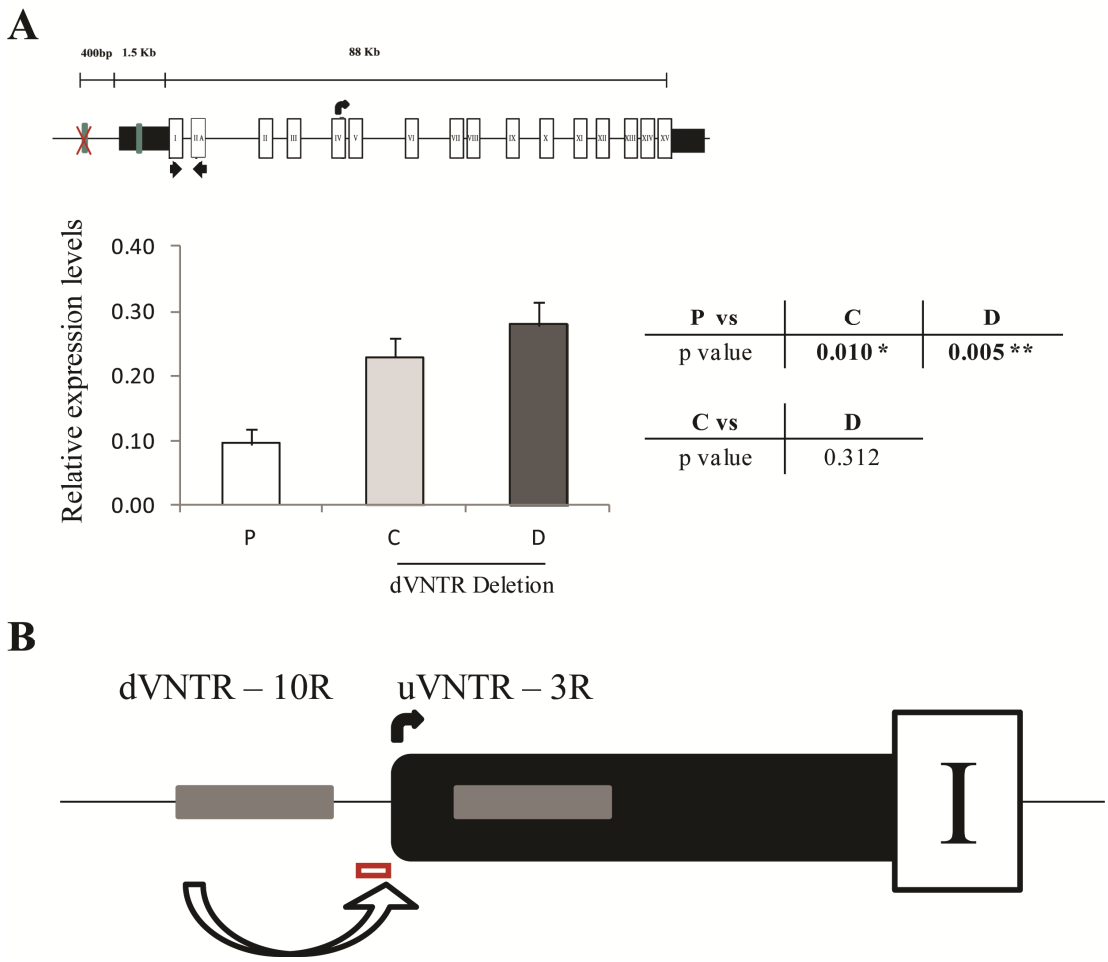


Figure 4.18 - Monoamine Oxidase A (*MAOA*) short isoform expression in HAP1 cell line – dVNTR deletion clones. Relative expression level of *MAOA* mRNA at basal condition. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19. Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the MAOA protein (short isoform). Black straight arrows show the forward and reveres primers respectively. P is the parental cell line, C and D are the dVNTR single deletion clones: 13_B5 and 13_B1 respectively. The table at the side provides the p-values obtained with the two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001. For each clone N=4. All values were normalised to β-Actin. **B.** Illustration of the *MAOA* promoter gene and the effect on the isoform expression of the dVNTR. Curved black line represents the TSS.

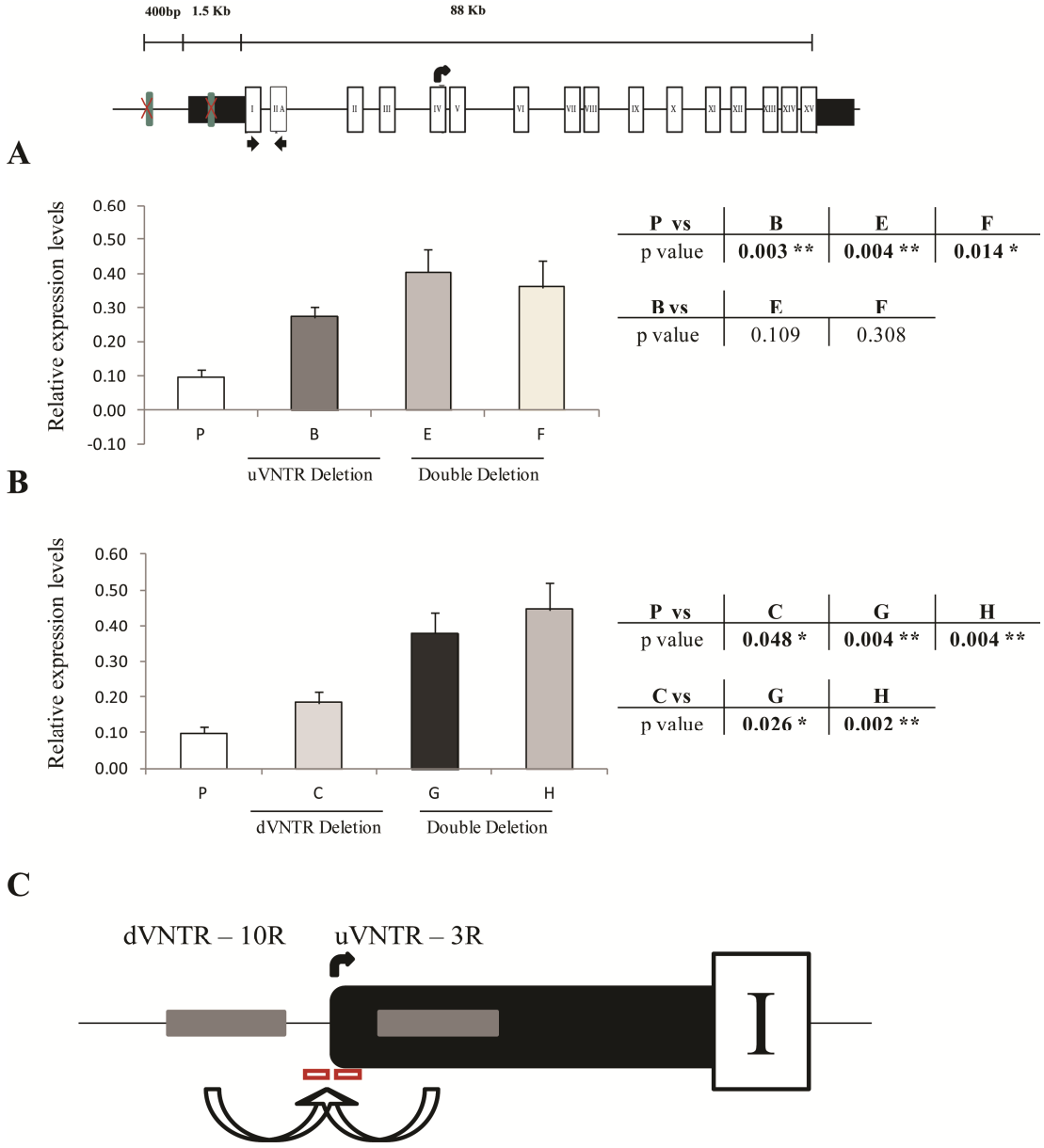


Figure 4.19 - Monoamine Oxidase A (*MAOA*) short isoform expression in HAP1 cell line – single deletion clones and related double KO clones. Relative expression level of *MAOA* mRNA at basal condition. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19. Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the conventional MAOA protein (long isoform). Black straight arrows show the forward and reveres primers respectively. **A.** P is the parental cell line, B single uVNTR deletion clone, E is the MAOA VNTRs double KO clone generated from Clone B: du_B5. **B.** P is the parental cell line, C single dVNTR deletion clone, G is the MAOA VNTRs double KO clone ud_D8. The tables at the side provide the p-values obtained with two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001. For each clone N=4. All values were normalised to β-Actin. **C.** Illustration of the *MAOA* promoter gene and the effect on the isoform expression of the dVNTR and uVNTR. Curved black line represents the TSS.

The difference between single and double KO, as illustrated in **Figure 4.19**, shows an additive effect of the two VNTRs in the short isoform expression when the double KO clones derived from clone C are assessed (**Figure 4.19 B**). However, although the same trend can be observed for the clones derived from clone B (**Figure 4.19 A**) they do not show a significant increase compared to clone B.

Surprisingly the expression data obtained using the uVNTR primer set (**Figure 4.20**) displays extremely similar values to the ones obtained with the primer set Exon I – Exon IIA.

These similarity in contrast with the expression data obtained by the expression analysis of the long and short isofom together (primer set Exon III – Exon VI) allow us to differentiate the 5′ UTR of the *MAOA* pre-mRNA, where currently, UCSC web browser excludes the shorter 5′ UTR from its database erasing this difference among isoforms.

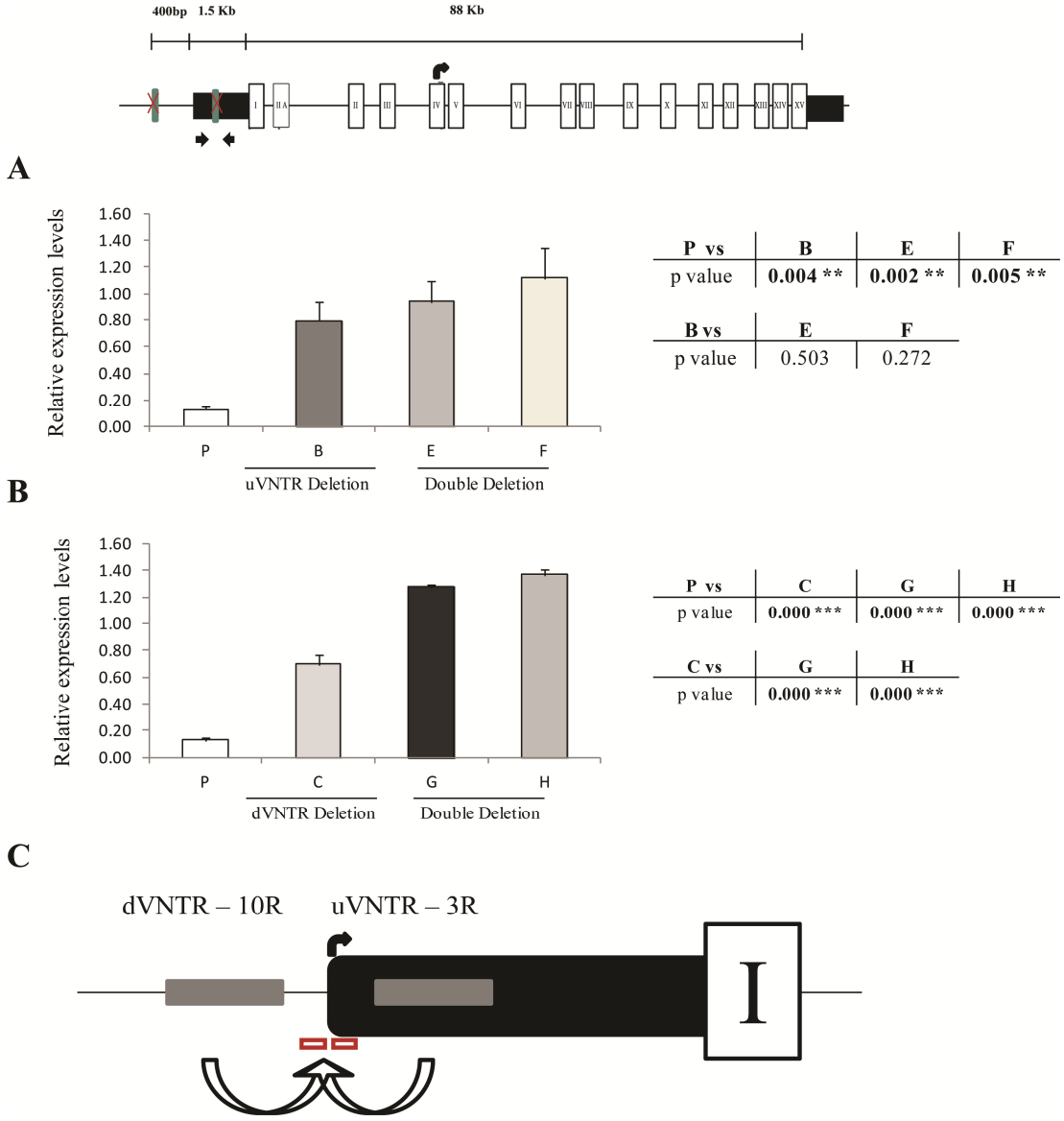


Figure 4.20 - Monoamine Oxidase A (*MAOA*) short isoform expression in HAP1 cell line – single deletion clones and related double KO clones. Relative expression level of *MAOA* mRNA at basal condition. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19. Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the MAOA protein (short isoform). Black straight arrows show the forward and reveres primers respectively. **A.** P is the parental cell line, B single uVNTR deletion clone, 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 single dVNTR deletion clone, 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 two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001. For each clone N=4. All values were normalised to β-Actin. **C.** Illustration of the *MAOA* promoter gene and the effect on the isoform expression of the dVNTR and uVNTR. Curved black line represents the TSS.

4.3.5 *MAOA* gene expression (Medium isoform)

For this *MAOA* isoform we were able to distinguish the expression of both the medium and long *MAOA* isoforms as Exon IV is completely missing from this transcript. This isoform splices out Exon IV from the coding sequence, hypothetically this encoded protein is unable to exert its normal function due to the fact that it is unable to bind the substrate because its binding site has been disrupted.

The expression data obtained from this assay were extremely low compared to the total *MAOA* expression. The amount of this isoform did not exceed 3% of the total *MAOA* mRNA. As illustrated in **Figure 4.21** the deletion of either the uVNTR or dVNTR nor the double deletion clones generated any alteration of the expression when compared to the parental cell line.

It is possible that this spliced isoform is an aberration of the normal *MAOA* Long mRNA and completely independent from the *MAOA* VNTRs.

Unfortunately the implication of this protein isoform and its mRNA remains unexplained at the moment but it would be interesting pursuing the research on this novel *MAOA* mRNA isoform.

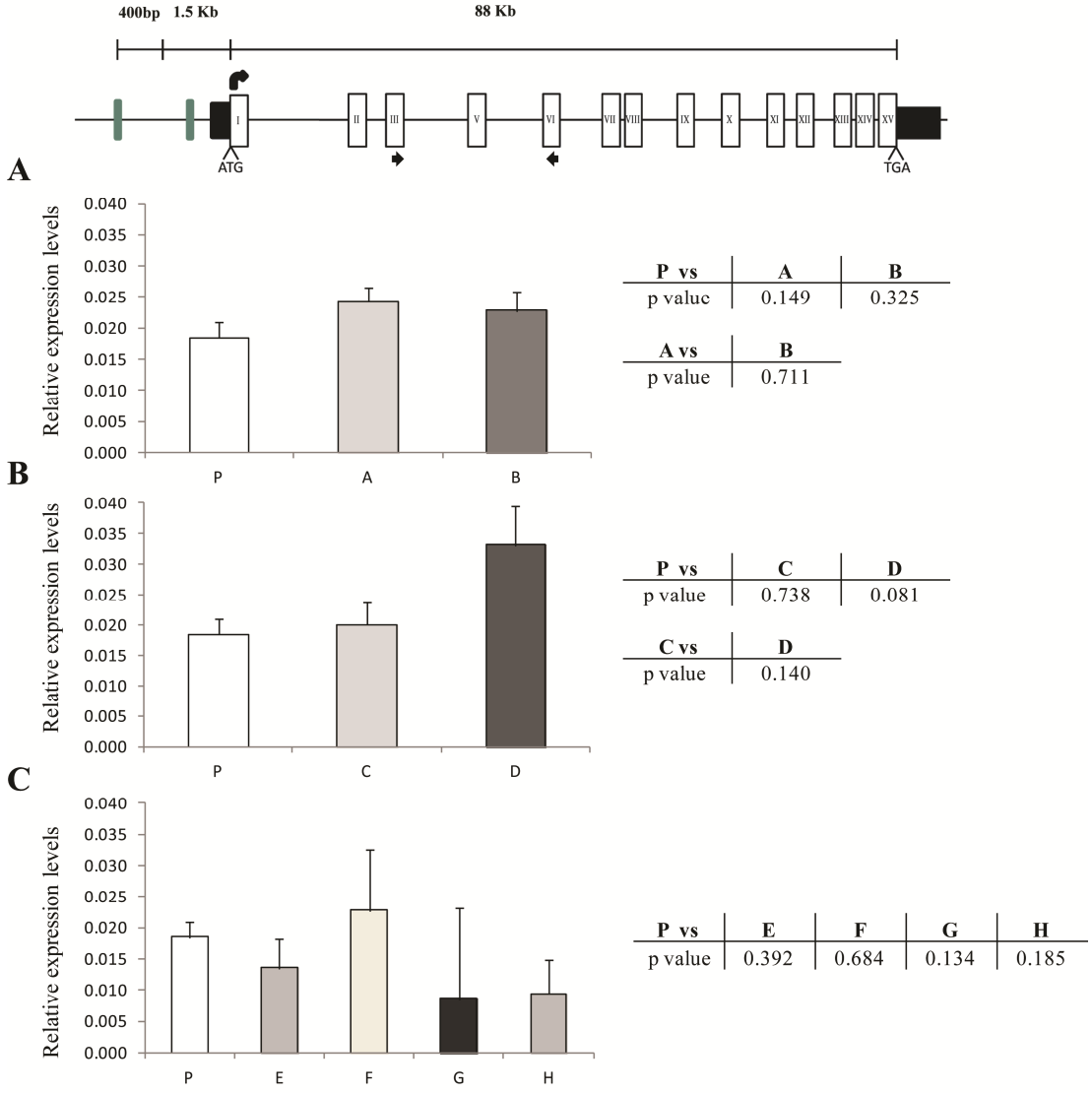


Figure 4.21 - Monoamine Oxidase A (*MAOA*) medium isoform expression in HAP1 cell line – single deletion clones and related double KO clones. Relative expression level of *MAOA* mRNA at basal condition. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19. Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the MAOA protein (medium isoform). Black straight arrows show the forward and reveres primers respectively. For each clone N=4. **A.** P is the parental cell line, B single uVNTR deletion clone, E is the *MAOA* VNTRs double KO clone generated from Clone B: du_B5. **B.** P is the parental cell line, C single dVNTR deletion clone, G is the *MAOA* VNTRs double KO clone ud_D8. **C.** 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 respectively. The tables at the side provide the p-values obtained with the two-tailed type 2 Student's T-Test. All values were normalised to β-Actin.

4.3.6 *MAOA* gene expression – valproate treatment

In chapter 3 we demonstrated that the *MAOA* 3R uVNTR was responsive to external stimuli in the neuroblastoma cell line SH-SY5Y, making this polymorphic region a target for a strong GxE interaction.

We therefore set to validate our previous results in the HAP1 cell line model, which possesses the *MAOA* 3R uVNTR, the most responsive VNTR in SH-SY5Y.

Figure 4.22 shows the expression level of the total *MAOA* mRNA that encodes for the canonical MAOA protein, representing more than 90% of the total *MAOA* mRNA, also including the mRNA that encodes for the short MAOA protein isoform.

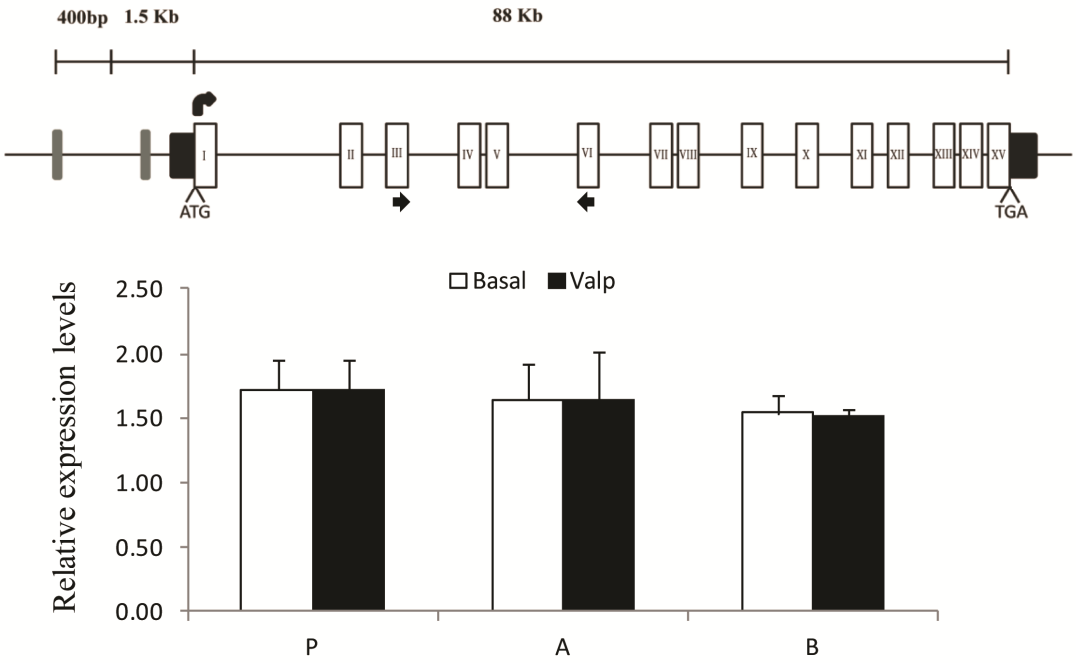
The PCR amplification uses the primer set from Exon III to Exon VI of the *MAOA* mRNA. After the Student's T test was performed, no significant difference among any of the analysed clones was found. The absolute expression levels, at basal condition, of this assay confirms the ones illustrated in **Figure 4.13** generated in a separate experiment for the parental cell line and uVNTR deletion clones A and B, where no differences between the parental cell line and the deletion clones were found. The addition of 5mM sodium valproate for 1 hour to the working media did not produce a visible effect on the *MAOA* expression.

On the contrary, the use of the *MAOA* uVNTR primer set, targeting the longer 5′ UTR containing the uVNTR in it, illustrated in **Figure 4.23**, shows in the parental cell line a 3-fold increase in the expression of this mRNA isoform after 1h treatment with 5mM sodium valproate. At basal condition the uVNTR KO clones showed the increase we observed in the previous experiment represented in **Figure 4.20** with a 3-fold increase in the expression compared to parental cell line.

Surprisingly, when the sodium valproate was administrated to the KO cell lines, we no longer observed the 3-fold increase displayed by the parental cell line after the same treatment.

It is possible that the deletion of the uVNTR, removing the repression exerted by the uVNTR itself on this isoform, brings the expression to its maximum level. However, it is also possible that the sodium valproate specifically acts in the uVNTR locus for this particular mRNA isoform, therefore the deletion of the uVNTR, makes the sodium valproate unable to exert its regulative mechanism.

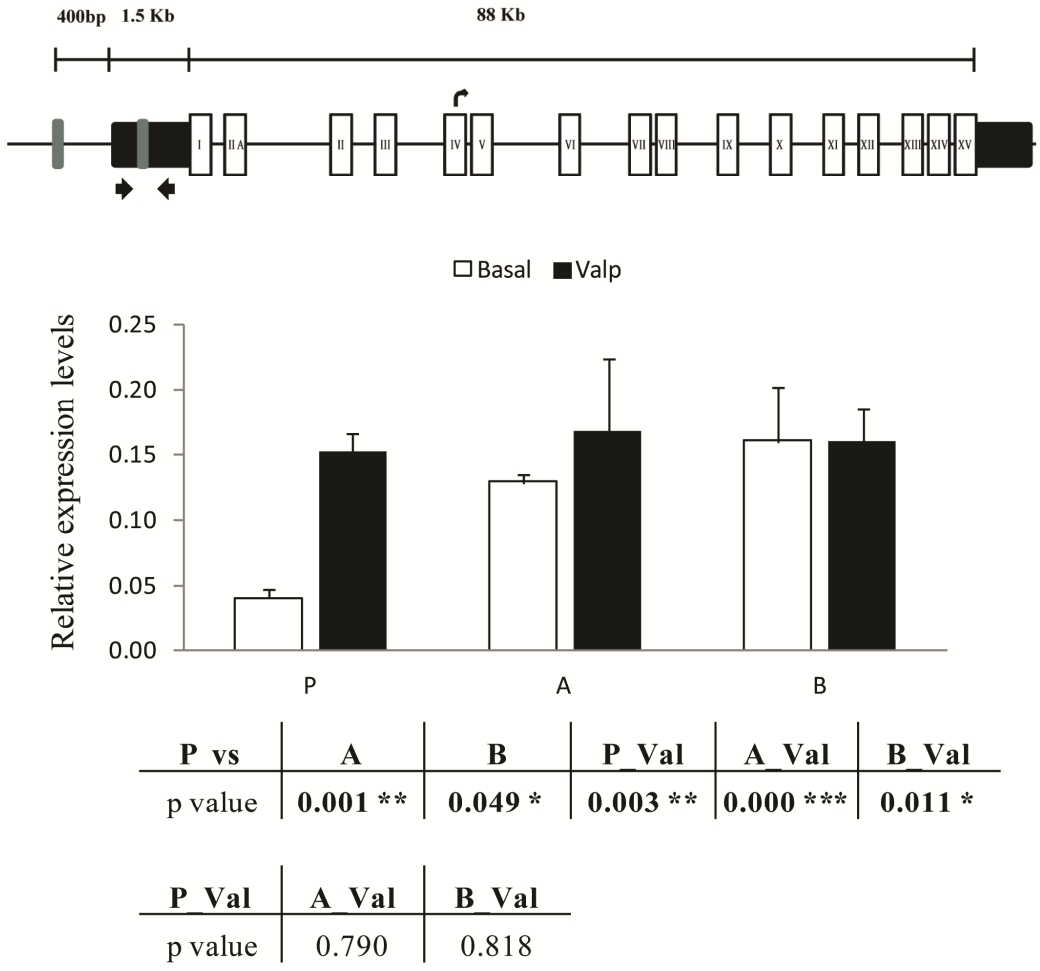
These results add another distinction on the two *MAOA* mRNA isoforms where the longer mRNA, encoding for the full length protein, is not affected by the mood stabiliser sodium valproate, while the isoform with the longer 5′ UTR is responsive to external stimuli, according to our previous results in chapter 3 (**Figure 3.3**).



P vs	A	B	P_Val	A_Val	B_Val
p value	0.836	0.528	0.989	0.860	0.438

P_Val	A_Val	B_Val
p value	0.868	0.446

Figure 4.22 - Monoamine Oxidase A (*MAOA*) expression in HAP1 cell line after 1h exposure to sodium valproate – uVNTR deletion clones. Relative expression level of *MAOA* mRNA at basal condition and after 1h exposure to 5mM sodium valporate. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19. Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the conventional MAOA protein (long isoform). Black straight arrows show the forward and reveres primers respectively. P is the parental cell line, A and B are the uVNTR single deletion clones: 9_F4 and 9_E2 respectively. The table at the bottom provides the p-values obtained with the two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001. For each clone N=3. All values were normalised to β-Actin.



P vs	A	B	P_Val	A_Val	B_Val
p value	**0.001 ****	**0.049 ***	**0.003 ****	**0.000 *****	**0.011 ***

P_Val	A_Val	B_Val
p value	0.790	0.818

Figure 4.23 - Monoamine Oxidase A (*MAOA*) expression in HAP1 cell line after 1h exposure to sodium valproate – uVNTR deletion clones. Relative expression level of *MAOA* mRNA at basal condition and after 1h exposure to 5mM sodium valporate. At the top illustration of *MAOA* gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19. Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Curved black arrow sets the translational start site (TSS) for the short MAOA protein. Black straight arrows show the forward and reveres primers respectively. P is the parental cell line, A and B are the uVNTR single deletion clones: 9_F4 and 9_E2 respectively. The table at the bottom provides the p-values obtained with the two-tailed type 2 Student's T-Test. *p<0.05, **p<0.01, ***p<0.001. For each clone N=3. All values were normalised to β-Actin.

Due to time limitation I was unable to assess the other HAP1 clones for the dVNTR deletion and the double deletions. An hypothesis can however be formulated. If the sodium valproate acts exclusively through the uVNTR we would expect at least a 3-fold increase in the expression for the dVNTR clones, as we can observe for the parental cell line (**Figure 4.19**). This expression increase will be added to the already increased expression level due to the dVNTR deletion which (**Figure 4.16)** is 3-4-fold compared to the parental cell line. Whenever the sodium valproate would have a regulative effect on the dVNTR as well we might expect the same expression as basal condition. At the same time the double KO clones would not be affected by sodium valproate since they do not possess the uVNTR nor the dVNTR. Their expression level should remain the same as basal condition.

4.3.7 MAOA protein

The relative expression of the MAOA protein has been performed in the lab by Miss Andrea McKavanagh, a master student currently under the supervision of Dr. Jill Bubb. The western blot analysis on the parental cell line (P), the uVNTR single KO clones (A and B) and the double KO cell line ud_B5 (E) is in line with the mRNA expression data: no differences can be observed in the protein expression levels between the parental cell line and the two uVNTR KO cell lines A and B (image not shown).

However, despite a significant reduction in the mRNA expression between the parental cell line and the double KO cell line (E), the protein level of the latter shows no significant difference compared to parental cell line. This partially unexpected result may be due to the long turnover of the MAO proteins.

In-vivo studies on baboon's brains on MAOB determined a half-life turnover of the protein to be 30 days (Arnett *et al.*, 1987). Similar studies in rat liver determined the half-life of the protein to be close to 3 days (Della Corte and Tipton, 1980).

More recently, an experiment conducted on mpkCCD cells, mouse derived renal cortical collecting duct cells, Sandoval *et al.* (2013) assessed the half-life of the expressed protein in this cell model and specifically set the MAOA protein half-life to 78.5 hours. In addition, a different MAOA mRNA/protein level has been previously reported in rat hippocampus (Morishima *et al.*, 2006), and is in line with our results in HAP1 cell line model, where a reduction in the mRNA does not correspond to a reduction in the protein. Furthermore, the monoclonal MAOA antibody which specifically targets the C-terminus of the MAOA protein, will be unable to detect short isoform of the MAOA protein as well as the medium one.

It has been observed that long 5′ UTRs can interfere with the translational processes through post transcriptional modification that can modify the translational efficiencies (Mignone *et al.*, 2002, van der Velden and Thomas, 1999) or weaken the start context (Rogozin *et al.*, 2001). This might explain the reason of the short MAO protein absence.

However, in a previous experiment, the protein analysis in the neuroblastoma cell line SH-SY5Y, showed two bands that are consistent with the expected size of the MAOA protein isoforms and the ratio RNA:protein. As illustrated in **Figure 4.24 C** the amount of the normalised values for the isoform M and S does not exceed 10% of the total MAOA protein.

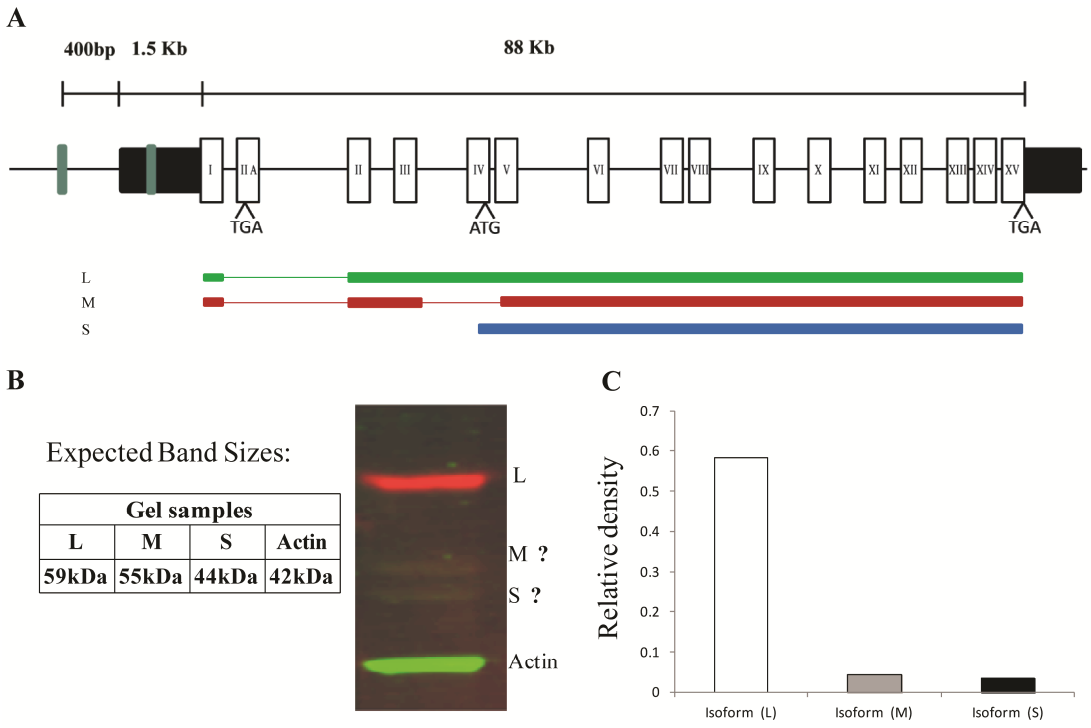


Figure 4.24 - Monoamine Oxidase A (MAOA) protein expression in SH-SY5Y cell line. A. Illustration of MAOA gene as reported in UCSC genome browser Hg38 and the most recent version of the Hg19. Black boxes represent 5′ and 3′ untranslated regions (UTRs), white boxes the exons. Colored lines represent the exons translated for each protein isoform: green for the Long, red for the Medium and blue for the short respectively. **B.** Expected sizes of the three MAOA protein isoforms based on amino acid sequence translation. Image if the western blot representing in red the MAOA protein, in green β-actin. **C.** Densitometric analysis of the three isoform after normalization to β-actin.

4.4 Discussion

All the major online web browsers such as NCBI or UCSC recognise and accept two distinct isoforms of the *MAOA* gene. These isoforms can be easily distinguished at several levels. The "canonical" *MAOA* isoform is composed of 15 exons with short 5′ and 3′ UTRs. The second isoform has 16 exons and long 5′ and 3′ UTRs. In addition, in the 5′ UTR of this latter isoform, the polymorphic variant, termed uVNTR (Sabol *et al.*, 1998), is part of the sequence.

The extra exon present in this isoform, which here has been termed as Exon IIA, contains a TGA stop codon in the sequence reading frame that would now allow the first ATG codon for the protein translation to be within Exon IV (**Figures 4.1 – 4.4**). The N-terminus of the resulting protein is missing in this isoform and the enzyme is 133AA shorter than the canonical one. As a consequence, the FAD binding domain is disrupted, as the N-terminus residues contribute to this domain; thus hypothetically this leaves this shorter isoform unable to break down monoamine neurotransmitters.

The evolutionary analysis of Exon IIA (**Figure 4.8**) and Exon IV (**Figure 4.9**) do not highlight particular differences between humans and non human primates for both exons, which is in agreement with previous studies on the *MAOA* gene (Andres *et al.*, 2004) and the X chromosome (Ebersberger *et al.*, 2002) conservation in humans and non human primates. However Exon IIA appears to be less conserved in rodents. In fact rodents do not seem to have the TGA stop codon in the reading frame. It has to be noted that UCSC web browser does not consider this sequence part of the transcript in any species other than human although the overall sequence appears to be highly conserved when aligned.

Extending our results on *MAOA* expression, reported in Chapter 3, we attempted to further analyse the expression pattern of the *MAOA* gene. The major aims were to assess differences of the three *MAOA* isoforms and the effect of the two polymorphic

VNTRs may have on the expression. In order to accomplish our goals we exploited the semi-haploid cell line HAP1 that has been engineered to remove either the uVNTR, the dVNTR or both from the *MAOA* promoter sequence.

We obtained a total of 9 cell lines. From the parental cell line (P) four single KO cell lines have been made after a CRISPR/Cas9 deletion; of these, two clones had the uVNTR deleted from their genome (clone A and B) and still maintain the dVNTR, and two clones had the dVNTR deleted (clones C and D) maintaining the uVNTR. The four double KO cell line (E, F, G and H) have been created with a second, *de-novo*, CRISPR/Cas9 deletion using a single KO cell line as a start point. Clones E and F derived from clone B and clones G and H from clone C (**Figure 4.10**).

To analyse the expression pattern of the *MAOA* mRNAs we designed different primer sets that allowed us to distinguish the expression of the short *MAOA* isoform from the total *MAOA* mRNA. For the latter the primers spanned from Exon III to Exon VI. These exons are present in both mRNA isoform sequences, thus this will measure the total amount of *MAOA* mRNA expressed.

To isolate the short MAOA isoform from the total *MAOA* mRNA, we used a primer set that amplifies from Exon I to Exon IIA. In addition the primer set that amplifies the uVNTR has be used for comparison and a possible feature to distinguish and confirm the length difference in the 5′ UTRs of the two *MAOA* mRNA isoforms.

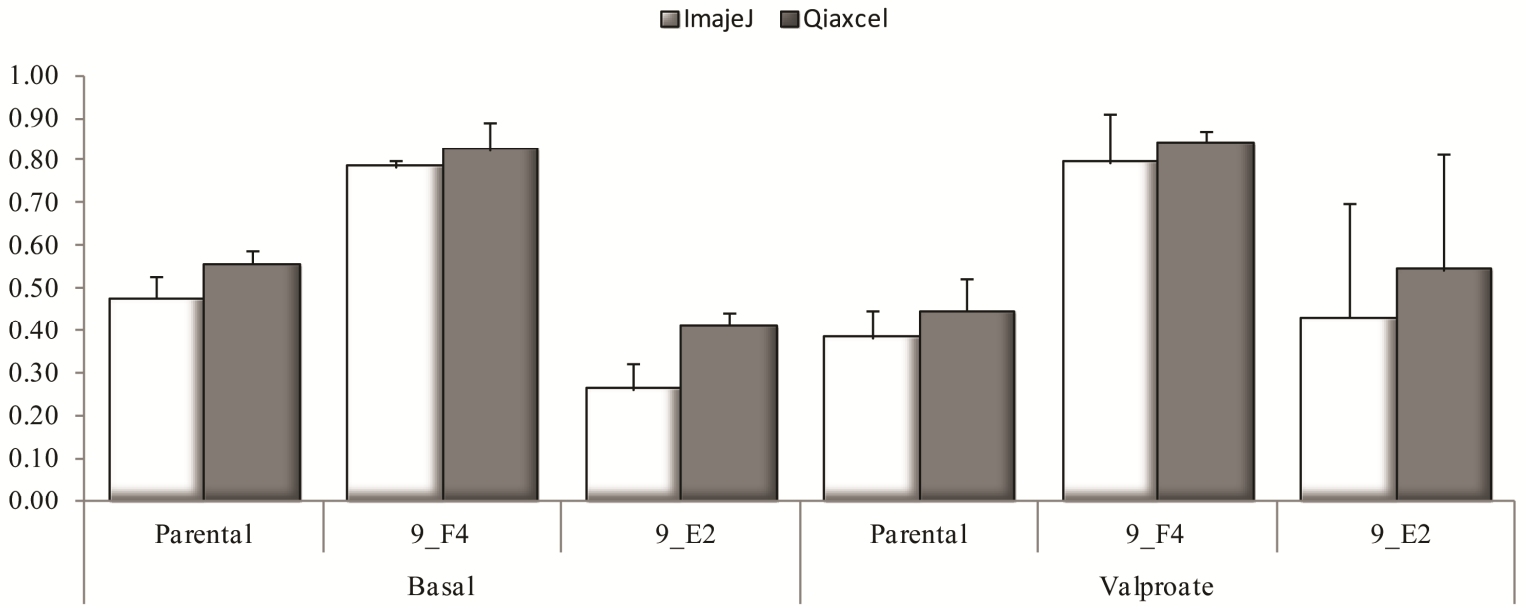
The first difference to be noted is the total amount of PCR product reported for the different isoforms. The Qiaxcel System (Qiagen), used for these experiments, at the end of its analysis provides the concentration (ng/ul) of peaks corresponding to the bands of an agarose gel. Since the RNA converted into cDNA and the amount of cDNA used for the analysis has been the same for all samples and every single

amplification, it can be assumed that the absolute expression levels of each assay can be compared.

A comparison of the Qiaxcel system and the intensity analysis obtained from the agarose gel is reported in the **Figure 4.25** and no difference between the two systems has been found.

Comparing the expression level of the total amount of *MAOA* mRNA (Exon III – Exon VI primers) and the second (short) isoform (Exon I – Exon IIA) in the parental cell line we go from a concentration of 1.90 ng/ul to 0.16 ng/ul respectively which sets the short isoform to 8% of the total *MAOA* mRNA (**Figures 4.13 – 4.17**). Similarly, using the uVNTR primer set we obtained a concentration of 0.13 ng/ul (7% of the total). This similarity strongly suggests, consistent with NCBI (http://www.ncbi.nlm.nih.gov/nuccore/AK293926.1), that the longer 5′ UTR, containing the uVNTR in it, is associated to the short *MAOA* isoform and its expression is less than 10% of the total *MAOA* mRNA at basal condition (**Figures 4.19 – 4.20**).

The expression of the *MAOA* isoforms are also affected differently by the two polymorphic VNTRs present in the *MAOA* promoter region. The uVNTR polymorphism is the most often cited example of a genotype by environment interaction (G×E) in which an individual's genotype moderates the effect of environmental experience to alter mental health outcomes and has always been considered a positive regulator of the *MAOA* gene but, as shown in **Figure 4.13**, its deletion from the promoter region does not affect the overall expression of the total amount of *MAOA* mRNA.



Basal			Valproate			
Parental	**9_F4**	**9_E2**	**Parental**	**9_F4**	**9_E2**	
0.134	0.291	0.081	0.287	0.363	0.396	One-Tailed, Type 2 T-test
0.269	0.581	0.162	0.573	0.727	0.791	Two-Tailed, Type 2 T-test

Figure 4.25 - Expression comparison. mRNA expression comparison between two different methods. In white is the densitometric analysis with the ImageJ software, in dark gray the results obtained with the Qiaxcel software. The same sample set has been used for both methods.

However, a significant difference is observed in the expression levels of the short isoform where the uVNTR appears to be a negative regulator (**Figures 4.17 – 4.20 A**). The expression of the short isoform in the absence of the uVNTR has a 3-4 fold increase in comparison to the parental cell line.

In addition when the cells are subjected to the mood stabiliser sodium valproate the uVNTR plays a major role, once again, only in the short isoform expression. As illustrated in **Figure 4.22**, sodium valproate does not affect the expression of the total *MAOA* mRNA in any HAP1 cell line while we can observe a 3-fold increase in the parental cell line after 1h treatment with sodium valproate (**Figure 4.23**) when only the short isoform is assessed with the uVNTR primer set.

The fact that the uVNTR deletion clones do not report the same increase as the parental cell line, but the expression level is the same as basal condition, suggests that the valproate exerts its effect through the uVNTR itself.

Unfortunately we were not able to clarify the mechanisms through which the uVNTR regulates the *MAOA* expression, nor the sodium valproate role in the *MAOA* gene expression regulation. It might be possible that post-translational mechanisms take place in presence of the uVNTR, thus its deletion increases the amount of this mRNA in the cells perhaps making it more resistant to degradation.

Conversely the dVNTR affects the expression of both *MAOA* isoforms, long and short, in the opposite direction. In fact its deletion significantly reduces the expression of the total amount of *MAOA* mRNA (**Figure 4.14**) giving it a positive regulator role with respect to the mRNA with the short 5′ UTR. On the contrary it acts as a negative regulator on the mRNA isoform with the long 5′ UTR. Its deletion, as shown in **Figures 4.18** and **4.20 B** significantly increases the expression.

As might be expected, the effect on the expression of the canonical *MAOA* isoform in the double KO cell lines is significantly lower than the parental cell line

(**Figure 4.15**) and at the same time significantly lower than the uVNTR deletion clones (**Figure 4.16 A**). However if we compare the expression of the double KO clones to the single dVNTR deletion clones we do not find any significant difference (**Figure 4.16 B**). This result strengthens the role for the dVNTR in driving the expression of the MAOA gene at the expense of the most studied uVNTR.

In the double KO clones G and H, in both **Figures 4.19** and **4.20 B**, an additive effect of the uVNTR and dVNTR over the expression of the short *MAOA* isoform can be observed but the same result does not appear for the clones E and F.

Table 4.1 – Summary of Monoamine Oxidase A (*MAOA*) isoforms expression

Basal Condition	Levels of expression vs Parental cell line		
	MAOA Isoform		
	L	M	S
uVNTR Deletion	Neutral	Neutral	Increased
dVNTR Deletion	Reduced	Neutral	Increased
Double Deletion	Reduced	Neutral	Major Increase

Valproate Treatment	Level of expression vs basal condition	
	MAOA Isoform	
	L	S
Parental	No Effect	Increased
uVNTR Deletion	No Effect	No Effect

Note: effect on the expression of the MAOA isoforms in the different HAP1 KO cell lines. "L" = full length MAOA; "M" = medium isoform; "S" = short isoform

This difference encountered between clones E – F and G – H, might be identified in the 43bp difference between the clones in the uVNTR locus (**Figure 4.11 A**). Further analysis will be required to identify the role of these specific 43 bp in the regulation of this *MAOA* isoform expression. However these data highlight further the role of the uVNTR in this short *MAOA* isoform expression and the primary role that it might have over people's mental health outcomes.

Table 4.1 summarises the expression patterns at basal condition and after sodium valproate treatment for the parental cell line, the uVNTR deletion cell lines and where possible the dVNTR and double KO cell lines.

These findings together identify the *MAOA* dVNTR as major positive regulator of the *MAOA* canonical isoform and at the same time set its role as a negative regulator on the short *MAOA* isoform. A model of the regulation exerted by the two polymorphic VNTRs in *MAOA* promoter region is summarised in **Figure 4.26**.

The MAOA uVNTR, a well established risk factor for several CNS disorders and conditions, appears to have a neutral effect on the expression of the *MAOA* canonical mRNA encoding for the full length MAOA protein, a role which instead is assigned to the less studied dVNTR, positive regulator of this isoform. Even when both VNTRs are deleted from the *MAOA* promoter, the expression is comparable to the dVNTR deletion alone. However, in the short *MAOA* isoform expression, both VNTRs display a negative regulatory activity which becomes additive when both are deleted from the promoter region.

The work in this chapter, including tissue culture, cell treatments, RNA extraction, cDNA conversion, PCRs and data analysis is all in collaboration and shared with Mrs Veridiana Pessoa.

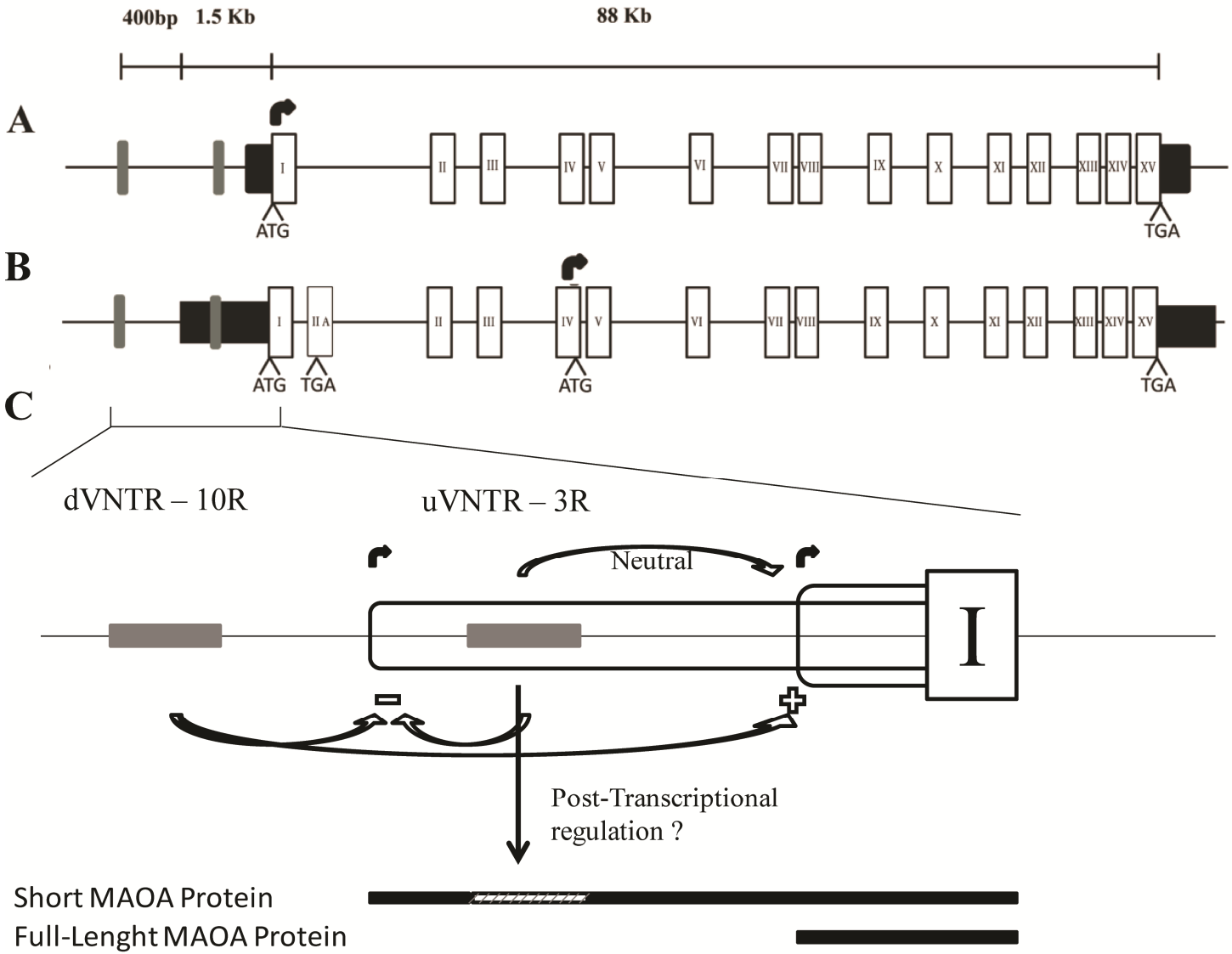


Figure 4.26 - Monoamine Oxidase A (*MAOA*) expression model chart. A. *MAOA* major isoform. **B.** *MAOA* secondary isoform. **C.** Graphic representation of the *MAOA* promoter region and the regulatory effect exerted by the u- and dVNTR on the two *MAOA* isoforms. White box Exon I, grey bars represent MAOA dVNTR and uVNTR from left to right respectively. Curved black lines represent the transcriptional start sites (TSSs) of the two *MAOA* isoforms. Underneath the curved lines are represented the 5′ untranslated regions (UTRs) of the two *MAOA* isoforms.

Chapter 5

The promoter of the schizophrenia GWAS gene *CACNA1C*

5.1 Introduction

Schizophrenia is a devastating neuropsychiatric disorder that affects approximately 1% of the worldwide general population (Guan *et al.*, 2014). Recent research strongly suggests that both environmental and genetic factors are causative in the predisposition to the disorder (Schizophrenia Psychiatric Genome-Wide Association Study, 2011) and term this interaction the G x E mechanism. The latter is particularly apparent in twin studies which establish a high heritability for SCZ (64%). The exact aetiology and genetic mechanism of SCZ is still unknown, however it is believed to be developmental in origin and involve multiple candidate genes, which deregulate and accumulate to cause nervous system dysfunction in response to environmental challenge (Karlsgodt *et al.*, 2010).

The *CACNA1C* gene has, in the last years, gained the attention of many scientists as it has been associated to SCZ by several genome wide association studies (GWAS). Along with other genes such as *ANK3*, *ITIH3-ITIH4* and MIR-137, *CACNA1C* association has been the most widely reproduced (Schizophrenia Psychiatric Genome-Wide Association Study, 2011). The consistency of its association with SCZ has led to it being one of the most established and studied SCZ risk loci (Obermair *et al.*, 2004).

It has recently shown that the SNP (rs1006737) located in the third intron of *CANCA1C* gene and strongly associated with SCZ (Nyegaard *et al.*, 2010), is able to modulate *CANCA1C* expression. Specifically, the risk allele for this SNP, with AA genotype, has been associated with a reduction in the mRNA expression of the gene (Eckart *et al.*, 2016).

CACNA1C promoter region contains, within 1Kb of its transcriptional start site (TSS), two tandem repeats, making the overall structure of the *CACNA1C* gene promoter region similar to the one in the *MAOA* gene.

It has been established that the *MAOA* uVNTR modulates gene expression in an allele-dependent manner (also see Chapters 3 and 4) and a preliminary bioinformatic analysis of the human *CACNA1C* gene demonstrated two possible repetitive DNA domains, within 1kb upstream of the proximal promoter (**Figure 5.1**).

We therefore set out to address the hypothesis that the *CACNA1C* TRs may too be polymorphic and may differentially support gene expression.

5.2 Aims

- Determine the polymorphic nature of the TRs in the *CACNA1C* locus in proximity to its promoter region
- Validate the effect of these TRs on gene expression through gene expression constructs
- Assess the effect of external stimuli and their ability to modulate the expression through the *CACNA1C* TRs

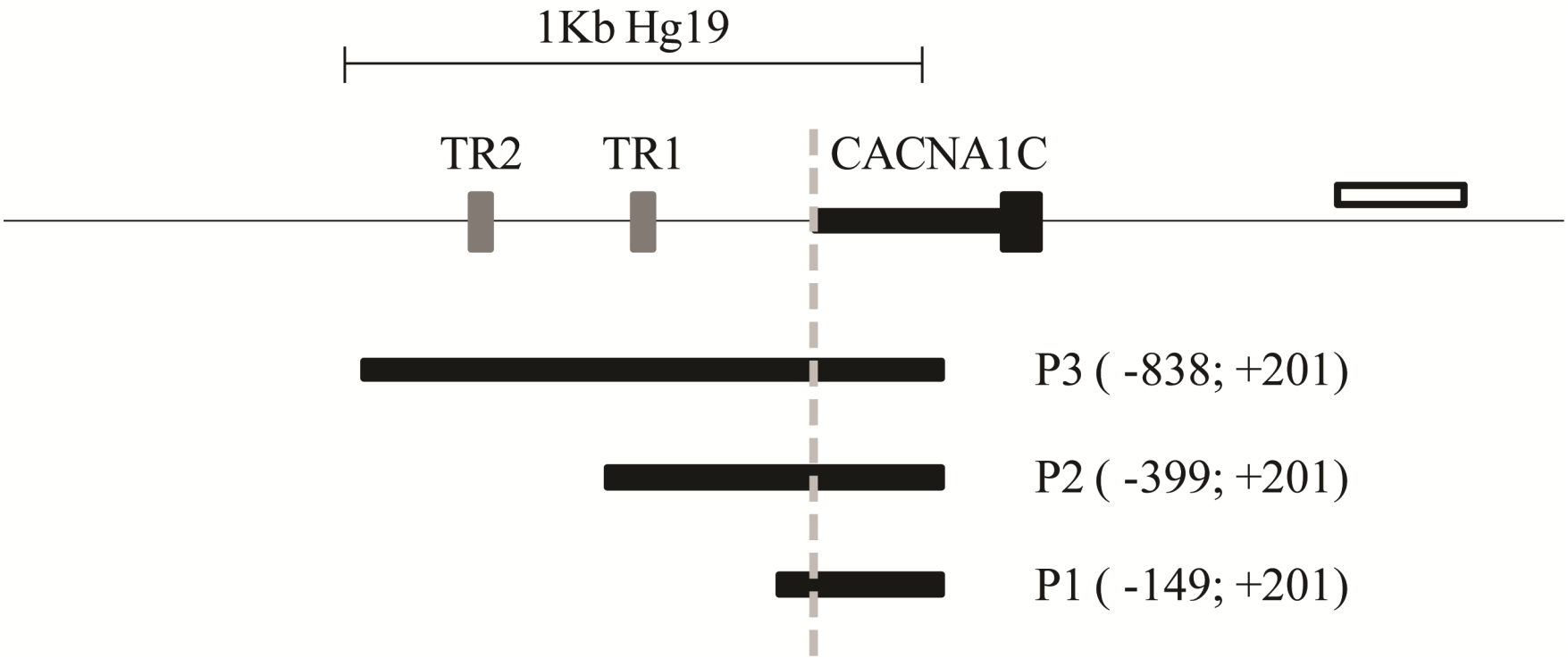


Figure 5.1 - Calcium Voltage-Gated Channel Subunit Alpha1 C *(CACNA1C)* promoter region. Schematic showing the *CACNA1C* gene described in UCSC Genome Browser, Assembly GRCh37/hg19**.** Initial Bioinformatic analysis of the CACNA1C proximal promoter identified two tandem repeats approximately 1kb upstream of the TSS (marked by grey vertical lines). Primers were designed using *in silico* PCR analyses ran by UCSC Genome browser. The white box indicates CpG island. The black box indicates the start of the *CACNA1C* coding gene. The coverage of each reporter gene construct is illustrated by the black boxes labelled: P1 (chr12:2,162,269-2,162,618) P2 (chr12:2,162,019-2,162,618) and P3 (chr12:2,161,580-2,162,618).

5.3 Results

5.3.1 Bioinformatic analysis

Due to its similar genetic architecture compared to *MAOA* we hypothesised that the proximal STR within the *CACNA1C* proximal promoter may be polymorphic and therefore a VNTR. Yet initial bioinformatic analysis using both UCSC genome browser and STR web browser (http://strcat.teamerlich.org) failed to determine polymorphism in the tandem repeat. UCSC genome browser did not identify any SNPs within the proximal promoter region and when the STR was inputted into the STR viewer (a catalog of human STR variation), it failed to identify the TR as polymorphic.

However more recent data has now supported our hypothesis and determined that the proximal STR is in fact variable in copy number and therefore a VNTR. Recently updated UCSC data has determined that there are two in-dels (Insertion-Deletion) within the proximal TR: rs530020760 (chr12:2162060-2162067 Hg19) which is a deletion of the "CGGG,CGGG" repeat motif seen in 2.08% of the population and rs548465087 (chr12:2162068-21620670 Hg19) which is an insertion of the "CGGG" repeat motif seen in 97.92% of the normal population. UCSC calls the STR as a 4bp repeat motif of "CGGG" therefore as recently updated SNP data reveals that an individual can have this motif either inserted or deleted the copy number is variable and the STR is in fact, like initially hypothesised a VNTR.

5.3.2PCR amplification of *CACNA1C* TRs

The primer design and PCR amplification of both the TRs genetic region within the *CACNA1C* proximal promoter was highly difficult due to the high GC content of the regions, just below 80% for both TR1 and TR2.

After a long optimization process with different enzymes and PCR protocols we were able to successfully amplify TR2 out of human genomic DNA. The genotype of TR2, a sample is illustrated in **Figure 5.2**, shows no polymorphic nature in any of the analysed samples.

Even more difficult was the optimization of *CACNA1C* TR1. Since bioinformatic analysis of the *CACNA1C* promoter region with UCSC web browser reported the polymorphic nature of the region, present only in the 2% of the population, we therefore decided to drop the optimization process and genotyping of *CANCA1C* TR1 genetic region.

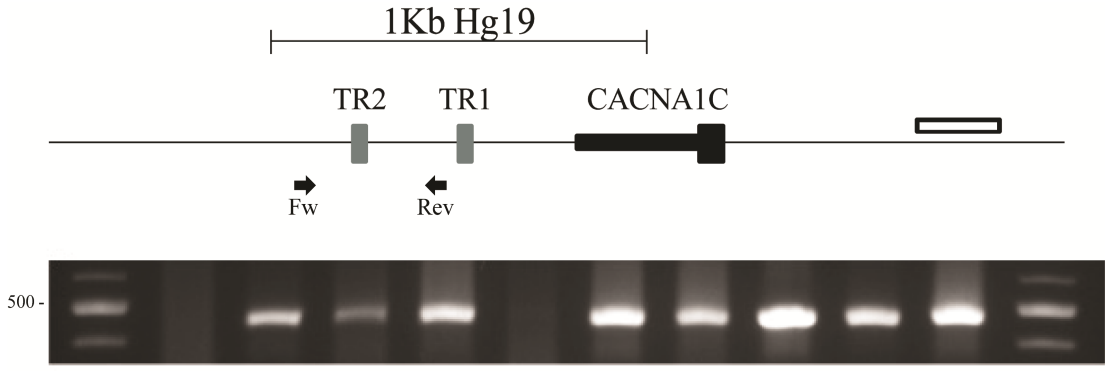
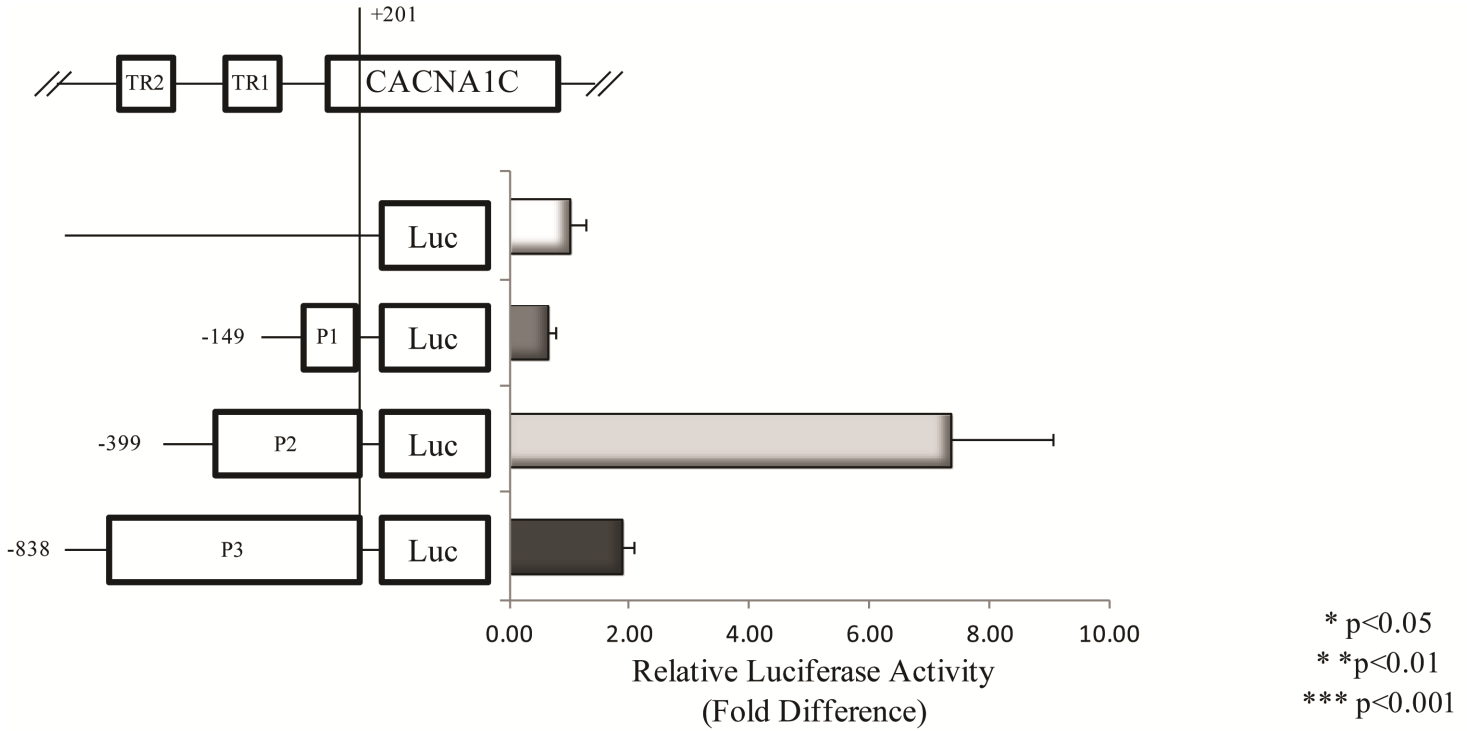


Figure 5.2 - Amplification of Calcium Voltage-Gated Channel Subunit Alpha1 C *(CACNA1C)* TR2. Schematic representation of the *CACNA1C* gene described in UCSC Genome Browser, Assembly GRCh37/hg19. The white box indicates CpG island. The black box indicates the start of the CACNA1C coding gene. Gray boxes represent the TRs. Black arrows are the forward and reverse primers used for the amplification from left to right respectively. The gel shows PCR amplification of 10 human samples for the TR2 genetic region enclosed within the forward and reverse primers highlighted in the graph at the top.

5.3.3 Luciferase Reporter Gene Expression

To characterise the *CACNA1C* proximal promoter activity we generated three reporter gene constructs; P1 from -149bp to +201bp, P2 from -399bp to +201bp and P3 from -838bp to +201bp where +1 is considered the first nucleotide of the *CACNA1C* mRNA transcript (**Figure 5.1**). Initial bioinformatic analysis identified two STR's 1kb upstream of the CACNA1C proximal promoter. Therefore the constructs were designed accordingly to include/exclude the STR's to observe any possible regulatory properties. As illustrated in **Figure 5.3** the minimal *CACNA1C* promoter P1 shows a similar expression to the empty vector pGL3B. The second construct, P2, containing the "CGGG" in-del previously described, supported a large increase in reporter activity (increase close to 11 fold), which resulted in a significant difference in activity compared to the pGL3B backbone (*** $p<0.001$) and P3 supported a slight, although non-significant increase in the expression of the luciferase gene over the pGL3B plasmid.

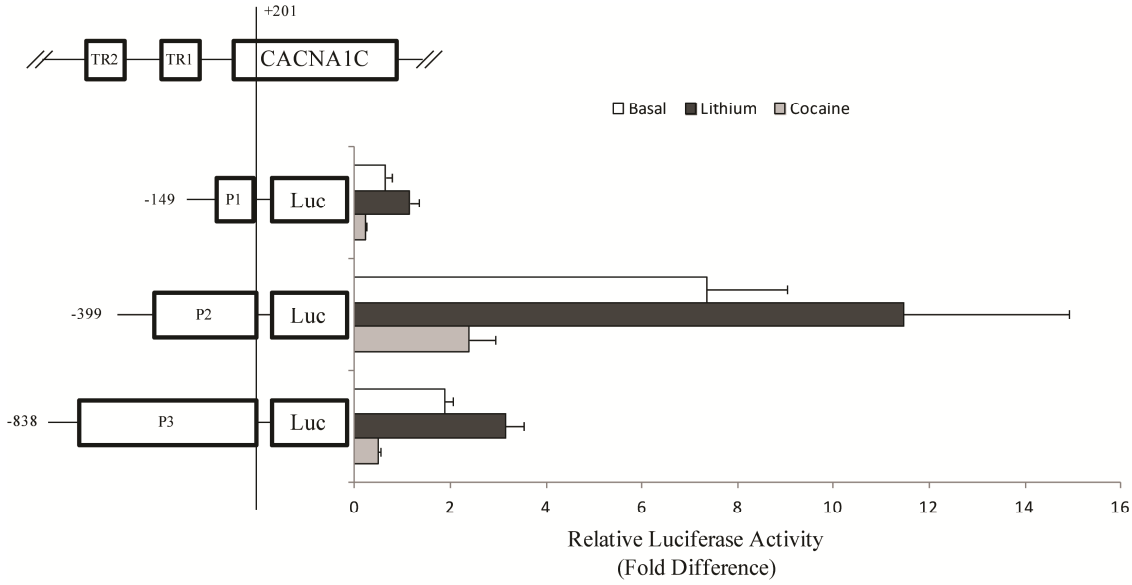
These data all together strongly suggest that at least in the *in-vitro* model neuroblastoma cell line SH-SY5Y, *CACNA1C* proximal promoter contains a minimum of two regulatory domains -399 to -250 and -838 to -399, which can function as positive and negative regulator of the expression respectively.



pGL3B vs	P1	P2	P3	P2 vs P1	P1 vs P3	P2 vs P3
	0.06738	**0.00147 ****	0.17856	**0.00057 *****	**0.00002 *****	**0.00200 ****
	0.13477	**0.00294 ****	0.35712	**0.00114 ****	**0.00004 *****	**0.00399 ****

Figure 5.3 - Validation of *CACNA1C* promoter region in SH-SY5Y cells. Schematic representation of *CACNA1C* promoter region constructs cloned in the pGL3B (Basic) reporter vectors in forwards orientation aligned to the *CACNA1C* gene. Average fold change in luciferase activity supported by the *CACNA1C* constructs over vector controls in SH-SY5Y cells. *N=3.* *Significant changes in luciferase activity over backbone control and between experimental conditions. *P<0.05, **P<0.01, ***P<0.001.

To determine if regulation behaved in a stimulus-induced manner, the SH-SY5Y cells were treated with a mood stimulant (cocaine) or a mood stabilizer (lithium). Both drugs have been shown to modulate signaling pathways in SH-SY5Y cells previously at the transcriptional and/or post-transcriptional level (Fernandez-Castillo *et al.*, 2015, Nciri *et al.*, 2015). Our analysis of reporter gene assays determined that gene expression is directed by the minimal promoter region (P1) and responds either positively or negatively in response to drug challenge, i.e. in a stimulus-inducible manner in all reporter gene constructs. Exposure to 1mM of lithium resulted in an approximate 1.5-fold increase in reporter gene activity in P2 compared to basal, whereas exposure to 10µM cocaine resulted in an approximate 3-fold decrease in reporter gene activity in P2. Overall the data also validates the previous finding that the P2 region contains a positive regulator and the P3 region contains a negative regulator**.** However the drug challenge data suggests that P1, the minimal *CACNA1C* promoter region directs gene expression, as it follows the same pattern of expression when exposed to drug challenge (**Figure 5.4**).



	Basal			Lithium			Cocaine		
pGL3B vs	**P1**	**P2**	**P3**	**P1**	**P2**	**P3**	**P1**	**P2**	**P3**
	0.067	**0.001** **	0.179	0.303	**0.004** **	**0.000** ***	0.057	**0.035** *	0.143
	0.135	**0.003** **	0.357	0.605	**0.009** **	**0.000** ***	0.114	0.071	0.286

	Basal		Lithium			Cocaine		
P1 Basal Vs	**P2**	**P3**	**P1**	**P2**	**P3**	**P1**	**P2**	**P3**
	0.002 **	**0.000** ***	0.496	**0.005** **	**0.000** ***	0.307	**0.001** **	0.122
	0.004 **	**0.000** ***	0.992	**0.010** *	**0.001** **	0.613	**0.001** **	0.244

	Basal	Lithium			Cocaine		
P2 Basal Vs	**P3**	**P1**	**P2**	**P3**	**P1**	**P2**	**P3**
	0.002 **	**0.001** **	0.213	**0.002** **	**0.001** **	0.402	**0.002** **
	0.004 **	**0.001** **	0.425	**0.004** **	**0.001** **	0.804	**0.003** **

	Lithium			Cocaine		
P3 Basal Vs	**P1**	**P2**	**P3**	**P1**	**P2**	**P3**
	0.000 ***	**0.017** *	0.339	**0.000** ***	**0.002** **	0.142
	0.000 ***	**0.034** *	0.678	**0.000** ***	**0.004** **	0.285

* p<0.05
* *p<0.01
*** p<0.001

Figure 5.4 - Calcium Voltage-Gated Channel Subunit Alpha1 C *(CACNA1C)* promoter region constructs in SH-SY5Y cells in a GxE environment contest. Schematic representation of *CACNA1C* promoter region constructs cloned in the pGL3B (Basic) reporter vectors in forwards orientation aligned to the CACNA1C gene. Average fold change in luciferase activity supported by the *CACNA1C* constructs over vector controls in SH-SY5Y cells at basal condition and after 1h treatment with 10uM cocaine and 1mM lithium. *N=3.* *Significant changes in luciferase activity over backbone control and between experimental conditions. *P<0.05, **P<0.01, ***P<0.001.

5.4 Discussion

We set out to address differential *CACNA1C* expression as it has been proposed that in SCZ modulation of Ca^{2+}-mediated signalling caused by differential *CACNA1C* and glutamatergic transmission could lead to the inefficient neuronal wiring observed in the disorder (Fromer *et al.*, 2014, Purcell *et al.*, 2014). Furthermore it has been proved that a SNP within intron 3 of the *CACNA1C* gene is able to modulate the expression of this gene (Eckart *et al.*, 2016). Thus, addressing further differential *CACNA1C* regulation could underpin another pathway modulating SCZ. Initial bioinformatic analysis identified tandem repeats within 1kb upstream of the CACNA1C proximal promoter according to UCSC web browser. Using the MAOA gene as a model, we hypothesised that the TRs in the *CACNA1C* proximal promoter region may be polymorphic. The primer design, optimization process and amplification of the two TR regions proved difficult due to the very high GC content. Within the designed primers, 80% of the entire sequence is GC rich. Amplification of the TR2 region in human genomic DNA proved this region not to be polymorphic (**Figure 5.2**).

Recently updated SNP data in the last UCSC web browser version (Hg19) has now validated our hypothesis and MAOA model, determining that the proximal TR (TR1) can be variable and therefore is a VNTR. However, the reported deletion of the “CGGG,CGGG” repeat motif can only be observed in 2% of the screened samples. Therefore, due to the difficulties encountered in the optimization and amplification processes and the very low abundance of this in-del in the population we decided to abandon the screening of this genetic region.

We characterised the *CACNA1C* proximal promoter, identifying an enhancer in the region of -399 to -250 (P2 construct) and a repressor in the region of -838 to -399 (P3 construct), interestingly the former contains *CACNA1C* VNTR (**Figure 5.3**).
We further determined that gene expression is directed by the minimal promoter region (P1 construct: -149 to +201) and have also supported the GxE mechanism proposed in SCZ by demonstrating in SH-SY5Y neuroblastoma cell line, *CACNA1C* expression can be differentially modulated by environmental stimuli i.e. positively in response to lithium and negatively in response to cocaine.

Bioinformatic analysis of the *CACNA1C* promoter region is of particular interest as ENCODE (Encyclopedia of DNA Elements) ChIP-Seq data set predicts transcriptional binding sites for several transcription factors (TFs). Among the TFs that bind to the region, PAX5 overlap with TR1 inside the P2 construct sequence.

PAX5 is member of the paired box (PAX) family. These proteins are believed to be important regulators in early development. *PAX5* has been implicated in SCZ as it is influential in neuronal development. Furthermore, recent postmortem studies have demonstrated that *PAX5* is downregulated in GABAergic neurons of patients that displayed neuropsychiatric and/or neuro-developmental disorders (Ohtsuka *et al.*, 2013).

The transcription factor Engrailed 1 (EN1) binds within the P1 construct, *CACNA1C* minimal promoter. EN-1is a homeobox gene that primarily helps to regulate developmental processes in the dorsal midbrain and anterior hindbrain (cerebellum and colliculi) of humans. Additionally EN-1 has been strongly associated with SCZ and with antipsychotic response (Webb *et al.*, 2008). EN-1 is fundamental in brain development which has been demonstrated by a team that generated a knockout mouse model with the *En1* homeobox, which resulted in death

less than 24 hours after birth as the mice refused to eat, despite having the physical ability to. Post-mortem examination of the mice discovered that due to the absence of the En1 TF the majority of the cerebellum, colliculi and cranial nerves were missing (Wurst *et al.*, 1994).

Furthermore, the Enhancer of zeste homolog 2 (*EZH2*) TF binds throughout the *CACNA1C* promoter region.

EZH2 is an important regulator of several developmental processes, including neurodevelopment, and it has been found dysregulated in several cancer types (Ronan *et al.*, 2013). It is considered a transcriptional repressor as its activity increases DNA methylation of its target genes acting in concert with DNA methyltransferases (Vire *et al.*, 2006). EZH2 has also been shown to interact with the REST-signalling complex and function as a co-repressor of neuronal gene expression (Dietrich *et al.*, 2012).

Taken together our data suggests that the *CACNA1C* minimal promoter is a stimulus inducible regulatory domain. This region could modulate *CACNA1C* expression in response to environmental factors, such as stress, in the medium to long term in addition to the immediate changes observed in our SH-SY5Y cell model, which would support the proposed GxE mechanism of SCZ. Bioinformatic analysis allowed us to identify the *CACNA1C* proximal TR (TR1) as variable in copy number and thereby a VNTR. Future analysis of this VNTR and the interaction that the two variants may have with TFs such as EZH2, EN1 and PAX5 over the expression of the gene may be important to a better understanding of SCZ aetiology.

Chapter 6

Characterisation of a VNTR polymorphism in the *REST* promoter; a biomarker for Alzheimer's disease?

6.1 Introduction

Alzheimer's disease (AD) is today one of the major cause of dementia in late adulthood affecting 24 million people worldwide (Hurd *et al.*, 2013). AD is a progressive degenerative disorder which involves anatomical, molecular and psychological changes in patients' brains (Sperling *et al.*, 2011). Memory loss, disorientation, depression and confusion are some of the symptoms presented to AD patients (Alzheimer's Association, 2015).

Genome-Wide Association Study (GWAS) studies have identified several genetic factors that have been associated to the early and late onset of this degenerative disorder, among these the Apolipoprotein E (APOE) which was the first genetic factor recognised as susceptibility gene for AD (Reitz, 2015b).

Recently, loss of transcription factor REST/NRSF has been described in patients with AD whilst healthy aged controls showed an increase of this protein compared to young control and AD patients, highlighting among other things a possible neuroprotective effect exerted by this transcription factor (Goetzl *et al.*, 2015, Lu *et al.*, 2014). Neuronal loss, which is thought to be a contributing cause of neurodegenerative disorders, including AD, might be contrasted by REST activity in normal brains since has been shown to repress pro-apoptotic genes and promote expression of genes that mediate oxidative stress resistance (Tothova *et al.*, 2007).

Despite the great advance of science and technology in recent years and all the genetic association studies that have been made, the processes that underlie most of the biological changes in AD still remain unexplained. Nevertheless they are postulated to be a combination of genetic and environmental factors (Reitz, 2015a).

Other well established genes, such as *MAOA* (chapters 3 and 4), that contain polymorphic genetic sequences in their promoter regions, have been extensively

proven to be able to modify gene expression and to be biomarkers for other CNS dysfunctions and/or behavioural traits. On these premises, we will try to identify a novel polymorphic biomarker that might regulate *REST* expression and be involved in the onset of AD, FTD or other CNS conditions such as schizophrenia, through the construction of expression vectors and genotyping *REST* VNTR in case vs control cohorts.

6.2 Aims

- Validate if the polymorphic region in *REST* promoter might regulate the expression of the gene through both in silico and experimental analysis
- Validate the potential to modify expression of this polymorphic VNTR in *REST* promoter through expression assays
- Genotype the *REST* VNTR to determine whether it can be used as a biomarker for predisposition to AD or other conditions such as schizophrenia, in which REST has been implicated

6.3 Results

6.3.1 Bioinformatic analysis of REST promoter locus

The undoubted importance of REST/NRSF and its involvement in a vast variety of regulatory pathways from childhood until late age, in both physiological and pathological environments, encouraged us to investigate possible biomarkers in proximity of its promoter region that might affect its expression. Following the study performed on the *MAOA* promoter region where two polymorphic VNTRs have been proven able to affect expression of the gene (Chapters 3 and 4), we performed a similar bioinformatic analysis on the *REST* gene.

A first look on UCSC genome browser gave us insight of a tandem repeat (TR) mapped right before the first of the three possible *REST* Transcriptional Start Sites. UCSC reports this TR as 2 copies of a 35bp sequence.

According to UCSC web browser several transcription factors bind in REST promoter region (**Figure 6.1**).

The ENCODE ChIP transcription factors database in UCSC shows that POLR2A strongly binds to the region encompassing all the three possible TSSs and that REST might self regulate the expression from the second and third TSSs. The histone activation marks also provide information on the activation of these TSSs.

The first one, starting on the 5′ end of the region, does not seem particularly active although the ENCODE data reports activation in embryonic cell lines. In addition, the overlap of this TSS with a CpG island suggests the presence of a promoter in this region (**Figure 6.1 A**).

Figure 6.1 - Characterisation of *REST* promoter region. A. Schematic representation of *REST* gene as described in the UCSC Genome Browser, assembly GRCh37/hg19, Feb 2009. Exons represented as white boxes; introns as connecting lines. Scale bar is represented at the top. Three different transcriptional start sites are identified by a different first exon. Vertical grey lines highlights putative variable regions according to UCSC Genome Browser. The green bar identifies a CpG island. Transcription factor binding and histone marks over the *REST* promoter from ENCODE V2 data **B.** Enlargement of the transcriptional start site containing the polymorphic microsatellite. From the top: scale bar, black arrows highlighting the relative position of the forward and reverse primers used for the genotype, green bar representing the CpG island. The white box is the first exon, each grey box represent a repetition of the microsatellite, grey bar show the putative VNTR according to UCSC Genome Browser. Number of repetitions and length in base pairs are shown for each variant. Underneath is shown the alignment of the sequences for each variant.

Meanwhile, a distinct bioinformatic resource, Short Tandem Repeat (STR) web browser (http://strcat.teamerlich.org/), provides evidence of a polymorphic microsatellite. This variant has been identified after sequencing a total of 24 human samples. Information from STR web browser are highlighted in **Figure 6.2**.

The most common allele, used as reference allele in STR web browser, is composed of 22bp, although being a repetition of 3bp the motif should be 21bp (7 copies of the CGG motif). The same can be said for a minor variant of 19bp (18bp – 6 copies) and a third variant of 28bp (27bp – 9 copies).

Interestingly this VNTR has been reported to be 2474bp upstream from the TSS. From our bioinformatic analysis, and according to UCSC web browser, this VNTR overlaps with the 5′ Untranslated Region (5′ UTR) of the first possible TSS (**Figure 6.1 A)**. AceView web browser directly includes this polymorphism in a putative protein translation, marking this region as part of a fully functioning protein.

This last statement has to be handled carefully as AceView browser does not report the first ATG codon as it should be for all the proteins. The first ATG in the reading frame is around 300bp upstream of this VNTR according to AceView web browser. However, the first available ATG codon would be exactly 36bp downstream of this polymorphism, excluding it from the translated protein. According to STR web browser, out of the 24 analysed samples, 14 American samples showed a single 6-copy variant of this polymorphism, 8 samples a 7-copy variant and 5 samples a 9-copy variant. In the European population only 8 samples have been sequenced for this locus and only the 7-copy and 9-copy variants have been identified with a 75-25% ratio (6 samples and 2 samples respectively) and demonstrating 35% of heterozygosity. A single African and a single Asian sample are homozygous for the 7-copy variant.

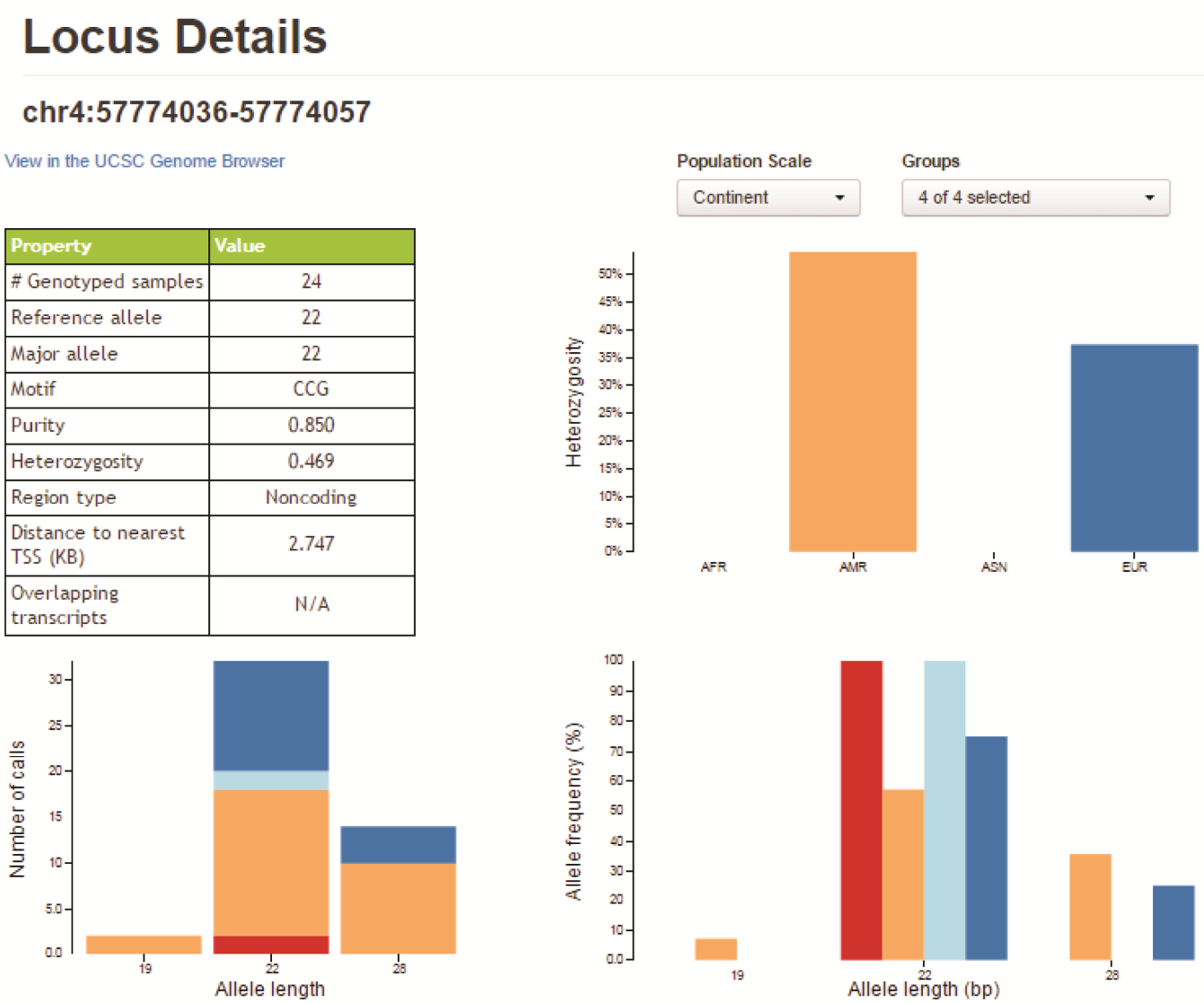


Figure 6.2- Snapshot of STR web browser for *REST* VNTR. Top left is a table summarizing the VNTR characteristics; on its left a bar chart showing the heterozygosity in the sequenced samples. Bottom left a chart with the relative genotype of the VNTR. On its right the frequency of each allele in all the analysed populations. Red bar is African population, orange bar American population, light blue bar Asian population, dark blue bar European population.

We decided to pursue the analysis of this already identified polymorphic region in addition to the NTR sequence that UCSC highlighted as possible polymorphism (**Figure 6.1 B**).

The PCR amplification of the region of interest in human neuroblastoma cell lines and human genomic DNA (Promega) is illustrated in **Figure 6.3** and it clearly appear polymorphic, although the 35bp difference of the NTR suggested by UCSC web browser appears not to be involved due to the very small bp difference in the agarose gel. This region has been subsequently purified, cloned and sequenced to confirm the nucleotide composition. The analysis of the human DNA samples performed in our laboratory also showed the presence of a previously unreported variant of 12 copy as illustrated in **Figure 6.4**.

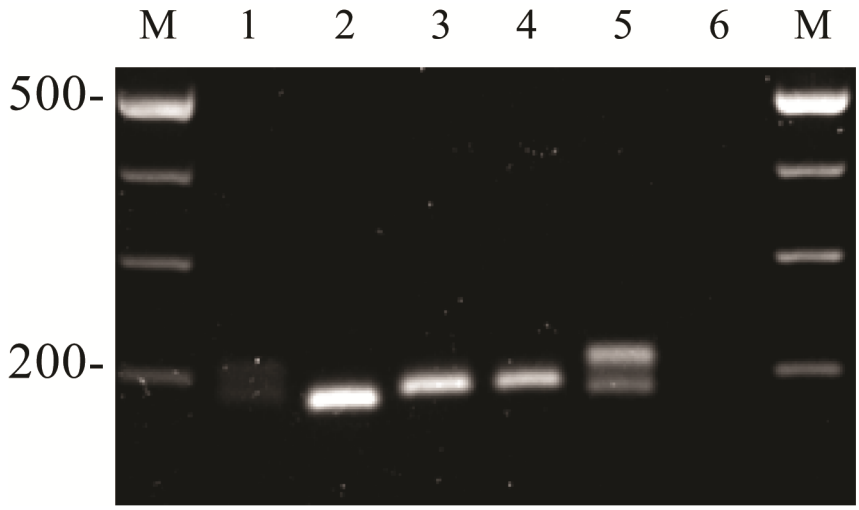


Figure 6.3 - Cell line genotype for *REST* VNTR. M: size marker 100bp (promega), lane 1: SK-n-AS, lane2: Be2C, lane 3: Kelly, lane 4:SH-SY5Y, lane 5 human gDNA, lane 6: PCR negative control.

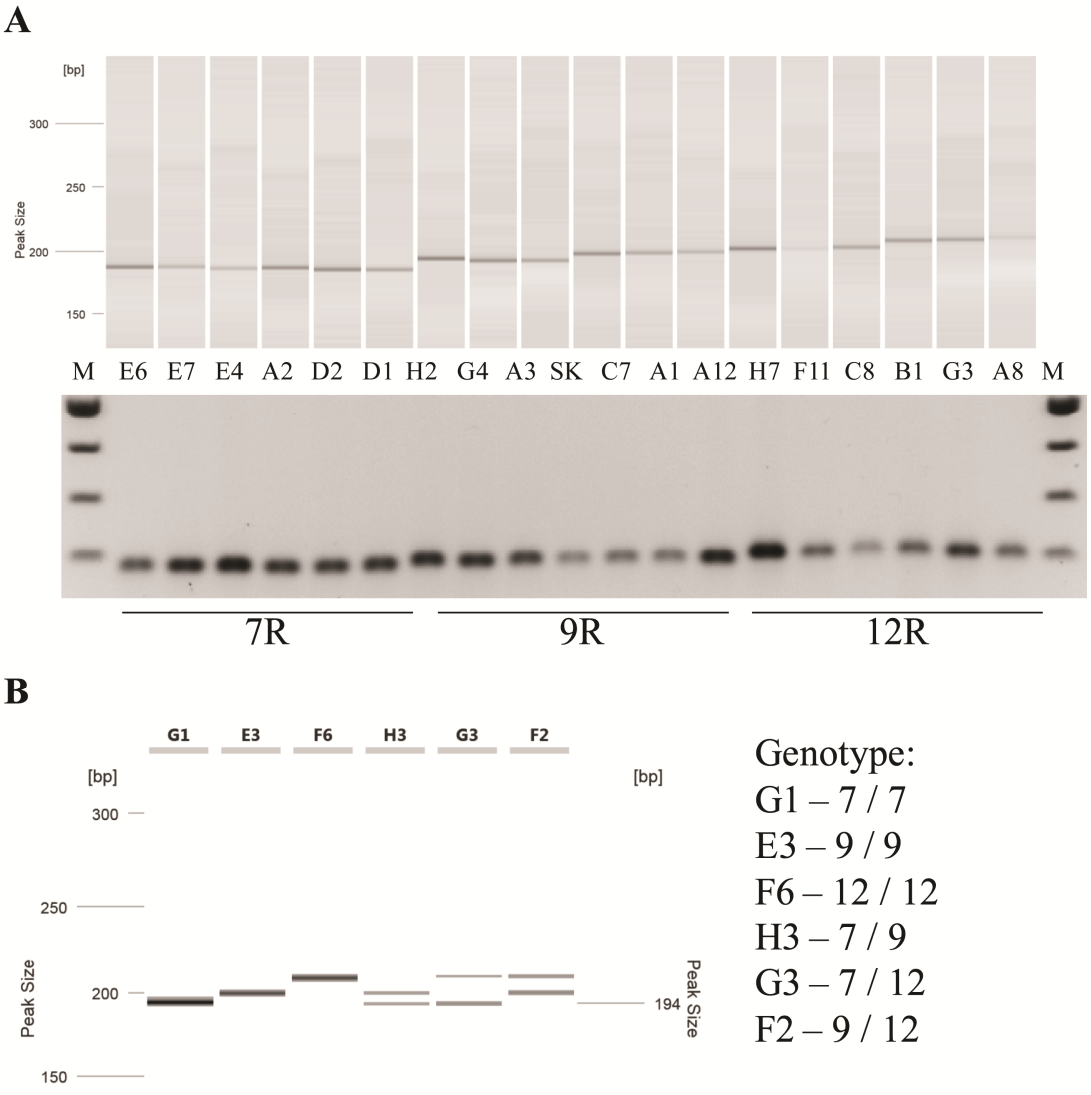


Figure 6.4 - *REST* VNTR polymorphism in human samples. PCR amplification of *REST* VNTR locus. **A.** Virtual gel generated from Qiaxcel of 6 homozygous samples for the 7 copy variants, 6 homozygous samples for the 9 copy variants and 6 homozygous samples for the 12 copy variants. Underneath 2% agarose gel of the same PCR reactions. **B**. Virtual gel generated from Qiaxcel highlighting the allelic genotype combinations. On the right side the table summarises the genotype relative to each sample.

6.3.2 *REST* VNTR supports reporter gene expression in the SH-SY5Y neuroblastoma cell line

To assess a possible regulatory function of *REST* VNTR, the 7-copies, 9-copies and 12-copies variants were cloned into the pGL3-Basic (pGL3B) luciferase reporter gene vector which lacks promoter and enhancer sequences.

The constructs included -119 to +60 bp, of the reference genome (7 copies), of the REST first TSS promoter using the first base of the pre-mRNA sequence as +1, which incorporated the entire VNTR sequence, the 9 copies and 12 copies constructs were -119 to +66 and -119 to +75 respectively.

The assay demonstrated the ability to modify the expression levels of the different *REST* VNTR constructs compared to the empty vector, showing a 6.21, 3.75 and 2.9 fold increase for the 7-copies, 9-copies and 12-copies respectively (**Figure 6.5**; ***$p<0.001$ for the average fold difference of luciferase activity supported by all the alleles of the *REST* VNTR over the pGL3B control).

A significant difference (***$p<0.001$) has been found in the expression driven by these polymorphisms where the most common variant (7-copies) appears to be more capable to drive the expression of the luciferase gene followed by the 9-copies and the 12-copies.

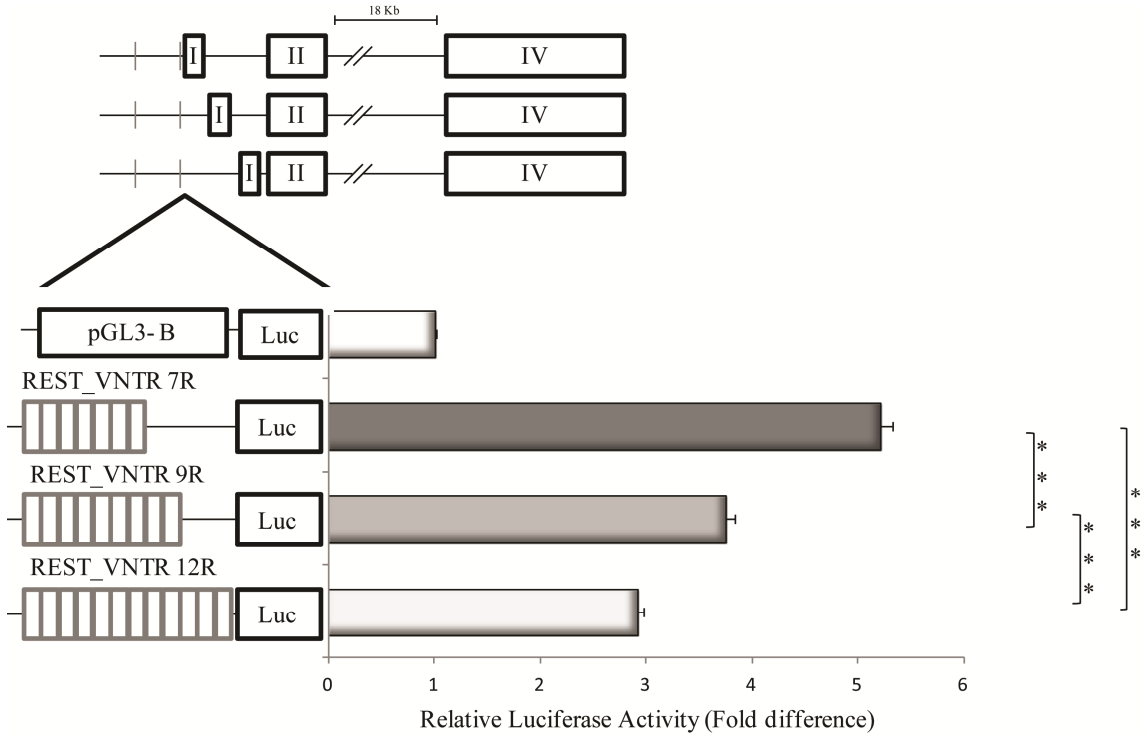


Figure 6.5 - Validation of *REST* VNTR potential to regulate gene expression in SH-SY5Y cells. Schematic representation of *REST* VNTR constructs aligned to the REST gene showing the 7-copy, *REST* VNTR, 9-copy *REST* VNTR and 12-copy REST VNTR, variants of the *REST* VNTR ± the proximal flank region in the pGL3B (Basic) reporter vectors in forwards orientation. Average fold change in luciferase activity supported by the *REST* VNTR constructs over vector controls in SH-SY5Y cells. *N=3.* *Significant changes in luciferase activity over backbone control and between experimental conditions. *P<0.05, **P<0.01, ***P<0.001.

6.3.3 Genotype Variation of the *REST* VNTR in Alzheimer's disease

The *REST* VNTR has been demonstrated to support differential reporter gene expression in the SH-SY5Y cell line model. We therefore hypothesised that this polymorphism might be a risk factor in the development of CNS diseases such as Alzheimer's disease, especially because the dramatic change in expression of this protein found in AD vs Control brains (Lu *et al.*, 2014). The genotype of this repetitive element was carried out in 225 AD patients, 71 age matched controls and 340 healthy controls representing the ageing population. Three different alleles of the 3bp *REST* VNTR were identified in our analysis including a 7-copies, 9-copies and 12-copies polymorphism.

	Alleles						N	Heterozygosity %
	7R	%	9R	%	12R	%		
STR browser	12	75.0	4	25.0			8	35
AD cohort	211	46.9	154	34.2	85	18.9	225	55.2
Aged CTRLs	61	43.0	54	38.0	27	19.0	71	63.4
Ageing CTRLs	298	43.8	246	36.2	136	20.0	340	66.5

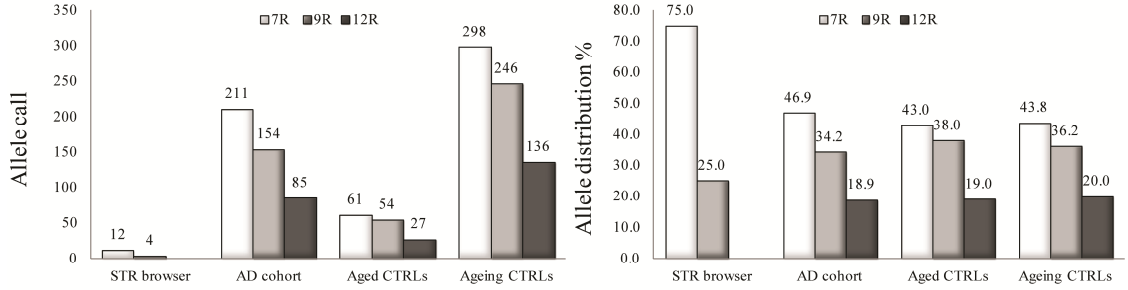


Figure 6.6 - Allele distribution in comparison with STR web browser. Analysis of *REST* VNTR comparing AD, Aged Controls, Ageing Controls cohorts and STR web browser European sequenced samples.

The allele distribution for our analysed cohorts and STR web browser is summarised in **Figure 6.6**. Consistent with STR web browser the most common alleles in our cohorts are the 7-copies and 9-copies with an average of 45% and 35% respectively. The 12-copies is represented in 19% of our population.

Interestingly STR web browser does not report the 12 copies polymorphism whilst we were unable to find any 6-copy variant in our cohorts highlighting the possibility of the 6 copies *REST* VNTR to be an extremely rare variant.

This demonstrate that the allele frequency is comparable within the cohorts but the difference in the heterozygosity of the AD cohort and the two control cohorts might represent the real difference in the representation of a risk factor.

Hardy–Weinberg (H-W) test was performed on each of the cohorts to evaluate whether the genotypes were in equilibrium. **Figure 6.7** illustrates the expected genotypes beside the observed ones for each of the tested groups.

In the control groups, whether analysed separately or together, no significant differences were found. In the AD cohort, however, the $\chi 2 = 11.48$ and the p-value with 3 df has been found <0.01 meaning less than a 1% chance that the discrepancy is due to chance or sampling error. Therefore, the assumption that the population is in H-W equilibrium (HWE) can be rejected.

The results given by the HWE analysis strengthen our hypothesis of an association between the *REST* VNTR and AD. This association may lie in the 7_7 or 12_12 genotype, both overrepresented in the AD cohort in comparison to the control groups, especially in the age matched controls where the 12_12 genotype is more present than in the ageing controls.

To validate our hypothesis, however, we performed the clump statistical analysis (sham and Curtis 1995). The *Clump 24* analysis software, which has been used for

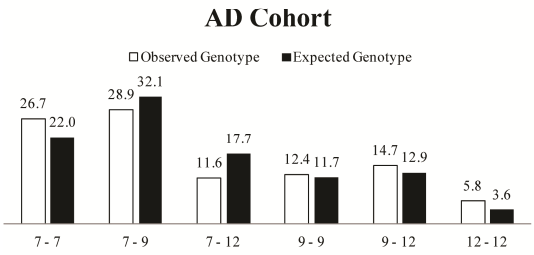
our study, assesses the significance of the observed values from the expected values using a Monte Carlo-based approach, as also explained in the Material and Method chapter.

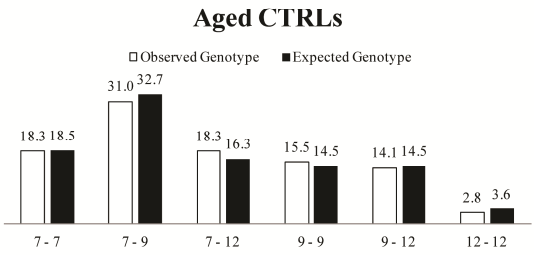
AD cohort

Observed Genotype	Count	%	Expected Genotype	Count	%
7 - 7	60	26.7	7 - 7	49	22.0
7 - 9	65	28.9	7 - 9	72	32.1
7 - 12	26	11.6	7 - 12	40	17.7
9 - 9	28	12.4	9 - 9	26	11.7
9 - 12	33	14.7	9 - 12	29	12.9
12 - 12	13	5.8	12 - 12	8	3.6
	225	100.0		225	100

Aged CTRLs

Observed Genotype	Count	%	Expected Genotype	Count	%
7 - 7	13	18.3	7 - 7	13	18.5
7 - 9	22	31.0	7 - 9	23	32.7
7 - 12	13	18.3	7 - 12	12	16.3
9 - 9	11	15.5	9 - 9	10	14.5
9 - 12	10	14.1	9 - 12	10	14.5
12 - 12	2	2.8	12 - 12	3	3.6
	71	100.0		71	100



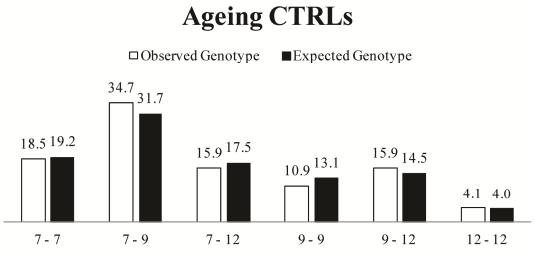


Ageing CTRLs

Observed Genotype	Count	%	Expected Genotype	Count	%
7 - 7	63	18.5	7 - 7	65	19.2
7 - 9	118	34.7	7 - 9	108	31.7
7 - 12	54	15.9	7 - 12	60	17.5
9 - 9	37	10.9	9 - 9	44	13.1
9 - 12	54	15.9	9 - 12	49	14.5
12 - 12	14	4.1	12 - 12	14	4.0
	340	100.0		340	100

Pooled CTRLs

Observed Genotype	Count	%	Expected Genotype	Count	%
7 - 7	76	18.5	7 - 7	78	19.1
7 - 9	140	34.1	7 - 9	131	31.9
7 - 12	67	16.3	7 - 12	71	17.3
9 - 9	48	11.7	9 - 9	55	13.3
9 - 12	64	15.6	9 - 12	59	14.5
12 - 12	16	3.9	12 - 12	16	3.9
	411	100.0		411	100



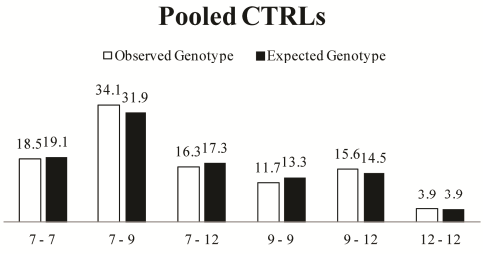


Figure 6.7 - Hardy–Weinberg prediction for expected genotypes. Comparison of the expected and observed *REST* VNTR genotypes in AD, aged control, ageing control and pooled control cohorts. Each table on the left side represent the observed genotypes for each cohort, on the right side the expected genotypes. The graphs underneath report the expected and observed genotypes.

The comparison of the genotype and allele frequencies between AD samples and aged controls are outlined in **Table 6.1**. Despite the significant association given by the HWE for the AD cohort, the clump analysis test, whose results are highlighted in the table on the left side of **Table 6.1** does not report any significant difference between the AD and aged control groups.

Table 6.1 - Genotype and allele frequencies of *REST* VNTR in AD cohort and matched aged controls.

AD

Genotype	Count	%
7 - 7	60	26.7
7 - 9	65	28.9
7 - 12	26	11.6
9 - 9	28	12.4
9 - 12	33	14.7
12 - 12	13	5.8
	225	100

Aged CTRL

Genotype	Count	%
7 - 7	13	18.3
7 - 9	22	31.0
7 - 12	13	18.3
9 - 9	11	15.5
9 - 12	10	14.1
12 - 12	2	2.8
	71	100

Clump Analysis

	AD vs CTRL $\chi 2$ (p value)
T1	4.80 (0.440)
T4	3.30 (0.484)

Allele	Count	%
7	211	46.9
9	154	34.2
12	85	18.9
	450	100

Allele	Count	%
7	61	43.0
9	54	38.0
12	27	19.0
	142	100

Note: AD = Alzheimer's disease. Tables at the top report the genotypes in the two different cohorts with count and percentage to the total. Tables below report the allele frequencies for REST VNTR. On the right, the clump analysis table with the χ^2 and p value. Refer to section 2.2.12 for details on clump analysis.

Although the clump analysis was not able to detect any significant difference between the groups, it is notable that a difference of 8.4% increase in the 7_7 genotype, a 3% increase in the 12_12 genotype and a reduction of 6.7% in the 7_12 genotype was observed for AD compared to controls.

The small sample size, especially for the control group, might represent a reason for the lack of power in the test. Moreover, increasing the sample size as in **Table 6.4**, where the aged and ageing controls were pooled together, resulted in the p-value by the clump analysis dramatically falling in all the 4 available tests.

We therefore decided to compare the two control cohorts in order to evaluate any possible difference between the groups. These two groups represent two different kind of healthy populations: the first, aged population, which are less likely to develop AD and the second, a population that due to the wider age range still has some possibility to develop AD in the future.

Our hypothesis in this case would be to find a protective genotype that might be present in the aged population that has not developed AD. The results of this analysis are summarised in **Table 6.2**.

The aged control group on the left and the ageing control group on the right respectively, although having different age range, are very similar between each other. The allelotype between the groups does not show any difference, however some slight, non significant, differences can be found in the genotypes.

The most notable difference, once again, is in the 12_12 genotype with 2.8% and 4.1% for aged and ageing cohort respectively and in the 9_9 genotype appearing more represented in the aged controls (16.5% vs 10.9%).

Table 6.2 - Genotype and allele frequencies of *REST* VNTR in aged and ageing controls.

Aged CTRLs

Genotype	Count	%
7 - 7	13	18.3
7 - 9	22	31.0
7 - 12	13	18.3
9 - 9	11	15.5
9 - 12	10	14.1
12 - 12	2	2.8
	71	100

Ageing CTRLs

Genotype	Count	%
7 - 7	63	18.5
7 - 9	118	34.7
7 - 12	54	15.9
9 - 9	37	10.9
9 - 12	54	15.9
12 - 12	14	4.1
	340	100.0

Clump Analysis

	CTRLs
	AD Vs SCZ χ^2 (p value)
T1	1.89 (0.869)
T4	1.44 (0.859)

Allele	Count	%
7	61	43.0
9	54	38.0
12	27	19.0
	142	100

Allele	Count	%
7	298	43.8
9	246	36.2
12	136	20.0
	680	100.0

Note: Tables at the top report the genotypes in the two different control cohorts with count and percentage to the total. Tables below report the allele frequencies for REST VNTR. On the right, the clump analysis table with the χ^2 and p value. Refer to section 2.2.12 for details on clump analysis.

As there are only small differences between the control groups, we decided that it was appropriate to pool them together in attempt to give more strength to the clump analysis. The results are summarised in **Table 6.3**.

Despite this, the clump analysis was not able to detect any significant difference between the groups, but now a trend can be delineated. It can still be observed that there was a 8.2% difference in the 7_7 genotype (higher in AD) which remained the same as in the previous analysis. The 7_12 genotype is still lower in the AD cohort although only by 4.7% and the 12_12 genotype difference lowered to 1.9%. The 7_9 genotype is also lower in AD compared to controls.

In every analysis performed the allelotype does not show any notable difference between the groups.

Table 6.3 - Genotype and allele frequencies of *REST* VNTR in AD cohort and pooled controls.

AD			Pooled CTRLs		
Genotype	**Count**	**%**	**Genotype**	**Count**	**%**
7 - 7	60	26.7	7 - 7	76	18.5
7 - 9	65	28.9	7 - 9	140	34.1
7 - 12	26	11.6	7 - 12	67	16.3
9 - 9	28	12.4	9 - 9	48	11.7
9 - 12	33	14.7	9 - 12	64	15.6
12 - 12	13	5.8	12 - 12	16	3.9
	225	100.0		411	100.0

Allele	**Count**	**%**	**Allele**	**Count**	**%**
7	211	46.9	7	359	43.7
9	154	34.2	9	300	36.5
12	85	18.9	12	163	19.8
	450	100		822	100.0

Clump Analysis

	AD vs CTRL $\chi 2$ (p value)
T1	9.27 (0.098)
T4	7.65 (0.082)

Note: AD = Alzheimer's disease. Tables at the top report the genotypes in the two different cohorts with count and percentage to the total. Tables below report the allele frequencies for REST VNTR. On the right, the clump analysis table with the χ^2 and p value. Refer to section 2.2.12 for details on clump analysis.

6.3.4 Genotype Variation of the *REST* VNTR in FTD and schizophrenia

In addition to the AD cohort, a small FTD cohort and a schizophrenia cohort, made available for the study, have been genotyped for the *REST* VNTR. In FTD there is a depletion of REST in neuronal nuclei (Lu *et al.*, 2014). As illustrated in **Table 6.4**, the genotype for the FTD cohort does not show any particular difference compared to the aged controls except for a 2.1% increase in the 12_12 genotype and a 2.4% reduction in the 9_9 genotype. The p value, close to 1, obtained from the Clump analysis confirms that there is no difference between the 2 cohorts. For FTD

as well as for AD the allelotype does not show any difference, giving more strength to our hypothesis that a particular genotype will be more influential than the presence of a particular allele alone.

Table 6.4 - Genotype and allele frequencies of *REST* VNTR in FTD cohort and age matched controls

	FTD			CTRL	
Genotype	**Count**	**%**	**Genotype**	**Count**	**%**
7 - 7	32	17.6	7 - 7	13	18.3
7 - 9	61	33.5	7 - 9	22	31.0
7 - 12	31	17.0	7 - 12	13	18.3
9 - 9	22	12.1	9 - 9	11	15.5
9 - 12	27	14.8	9 - 12	10	14.1
12 - 12	9	4.9	12 - 12	2	2.8
	182	100		71	100

	FTD vs CTRL
	$\chi 2$ (p value)
T1	1.16 (0.946)
T4	0.59 (0.972)

Allele	**Count**	**%**	**Allele**	**Count**	**%**
7	156	42.9	7	61	43.0
9	132	36.3	9	54	38.0
12	76	20.9	12	27	19.0
	364	100		142	100

Note: FTD = Fronto temporal dementia. Tables at the top report the genotypes in the two different cohorts with count and percentage to the total. Tables below report the allele frequencies for REST VNTR. On the right, the clump analysis table with the χ^2 and p value. Refer to section 2.2.12 for details on clump analysis.

In the schizophrenia cohort we can observe an inversion in the trend for the 7_7 genotype with a slight reduction, an increase of 6.7% and 2.3% in the 9_9 and 12_12 genotype respectively in comparison to the healthy controls. The heterozygosity of this group (59%) is between the AD and control groups.

Finally in **Table 6.6** are represented the three cases cohorts that we have genotyped for a direct comparison. Once again, the clump analysis does not show any significant differences among any of the cohorts.

Table 6.5 - Genotype and allele frequencies of *REST* VNTR in schizophrenia cohort and age matched controls

	SCZ			CTRL		
Genotype	**Count**	**%**	**Genotype**	**Count**	**%**
7 - 7	53	17.9	7 - 7	63	18.5
7 - 9	78	26.4	7 - 9	118	34.7
7 - 12	51	17.2	7 - 12	54	15.9
9 - 9	49	16.6	9 - 9	37	10.9
9 - 12	46	15.5	9 - 12	54	15.9
12 - 12	19	6.4	12 - 12	14	4.1
	296	100.0		340	100.0

	SCZ vs CTRL $\chi 2$ (p value)
T1	9.18 (0.106)
T4	6.61 (0.133)

Allele	**Count**	**%**	**Allele**	**Count**	**%**
7	235	39.7	7	298	43.8
9	222	37.5	9	246	36.2
12	135	22.8	12	136	20.0
	592	100.0		680	100.0

Note: SCZ = schizophrenia. Tables at the top report the genotypes in the two different cohorts with count and percentage to the total. Tables below report the allele frequencies for REST VNTR. On the right, the clump analysis table with the χ^2 and p value. Refer to section 2.2.12 for details on clump analysis.

However, the 7_7 genotype is overrepresented in the AD group, with an increase of 9.1% and 8.8% compared to FTD and schizophrenia respectively.

The heterozygous 7_9 genotype is increased in the FTD cohort by 5% compared to the other cohorts. In the 7_12 genotype it can be observed a reduction of 6% in the AD cohort, whilst in FTD and SCZ cohort there is no difference.

The 9_9 genotype is 4% higher in the schizophrenia cohort than in AD and FTD. Finally no difference is observed in the 9_12 genotype and a very small difference is present in the 12_12 genotype among the groups. In conclusion the major difference (8.4%) in the 7_7 genotype appear to be specific for the AD.

The 12_12 genotype is increased in the AD and schizophrenia cohorts compared to the relative controls.

Table 6.6 - Comparison for the genotype of *REST* VNTR in AD, FTD and schizophrenia cohorts.

AD			FTD			SCZ		
Genotype	**Count**	**%**	**Genotype**	**Count**	**%**	**Genotype**	**Count**	**%**
7 - 7	60	26.7	7 - 7	32	17.6	7 - 7	53	17.9
7 - 9	65	28.9	7 - 9	61	33.5	7 - 9	78	26.4
7 - 12	26	11.6	7 - 12	31	17.0	7 - 12	51	17.2
9 - 9	28	12.4	9 - 9	22	12.1	9 - 9	49	16.6
9 - 12	33	14.7	9 - 12	27	14.8	9 - 12	46	15.5
12 - 12	13	5.8	12 - 12	9	4.9	12 - 12	19	6.4
	225	100		182	100		296	100.0

	AD vs FTD $\chi 2$ (p value)	AD vs SCZ $\chi 2$ (p value)	FTD vs SCZ $\chi 2$ (p value)
T1	6.66 (0.248)	9.21 (0.104)	3.96 (0.557)
T4	4.90 (0.267)	6.52 (0.145)	2.80 (0.603)

Note: AD = Alzheimer's disease; FTD = Fronto temporal dementia; SCZ = schizophrenia. Tables at the top report the genotypes in the three different cohorts with count and percentage to the total. Tables below report the clump analysis with the χ^2 and p value. Refer to section 2.2.12 for details on clump analysis.

6.4 Discussion

The *REST* gene has been recently associated with AD, in particular REST protein levels have been found dramatically reduced in AD CNS (Goetzl *et al.*, 2015, Lu *et al.*, 2014). Bioinformatic analysis on the REST locus highlighted the presence of three possible TSSs. Two of these TSSs overlap with a CpG island and ENCODE data suggest that POLR2A binds across them, suggesting the presence of a promoter. The ENCODE database also gives information about the binding of REST itself to the second and the third TSSs.

Several other genes such as *MAOA*, *5HTT*-LPR, *STIN2* (Hill *et al.*, 2013, Guindalini *et al.*, 2006, Haddley *et al.*, 2008, Ali *et al.*, 2010, Vasiliou *et al.*, 2012, Galindo *et al.*, 2011, Roberts *et al.*, 2007) have been identified to possess polymorphic copy variants comprised of repetitive elements within their promoters which have been linked to CNS conditions, diseases or behavioural traits, both by themselves or in the Gene x Environment interaction.

The data generated from the reporter gene constructs indicate that *REST* VNTR can significantly modulate reporter gene expression. In addition the length of the repetitive element differentially modulates the expression, the 7-copy *REST* VNTR having the highest power of driving the expression, while the 12-copy has the lowest.

To test our hypothesis of *REST* VNTR as a potential biomarker for AD, we genotyped an AD cohort with age matched and non-matched control cohorts. In the analysed samples 3 different variants have been found consisting in 7 copies, 9 copies and 12 copies of a CGG triplet.

The first 2 variants have been previously reported by the STR web browser while the third, unreported previously, has been experimentally found in our study. Another variant, consisting in 6 copies, has been reported by STR web browser in

just one sample but we were unable to find it in our cohorts suggesting to be a very uncommon variant.

Our cases and control groups differ is heterozygosity; the AD samples show an average reduction of 10% compared to controls either aged matched or not (**Figure 6.5**). Furthermore, the allele distribution does not differ among the groups suggesting that the specific individual genotype would be more important than the presence of a particular allele.

Comparing the AD cohort with its age matched controls we find a reduction (8.8%) in heterozygosyty especially in the 7_9 and 7_12 alleles which is compensated by an increase (8.4%) of the homozygous 7_7 genotype in the AD group (**Table 6.1**). At the same time they also show an increase of 3% in the 12_12 genotype and a reduction of 3% in the 9_9 genotype.

Considering both the genotype and the reporter gene data we could draw some predictions.

The 7_7 genotype, which provides the highest reporter gene expression in our assays might actually result in an overall down-regulation of REST, since REST might inhibit its own expression, whilst the 12_12 genotype, with the lowest reporter gene expression, might bring an overabundance of it.

Therefore, the thousands of genes regulated by REST will be dramatically affected (Bruce *et al.*, 2009, Johnson *et al.*, 2007).

In addition these genotypes might be linked to the different cognitive impairment seen by Lu *et al.* (2014) where patients with severe cognitive impairments have been found with an even lower REST nuclear expression compare to those with mild or no cognitive impairment. Unfortunately we were unable to prove this hypothesis in our cohorts because we have been unable to access the relevant clinical data.

The 9_9 genotype is 5% more abundant in the aged control group, this might be a sign of an optimal expression rate that might reveal itself to be a protective trait at least for AD, although this same genotype has been found particularly high in the schizophrenia cohort compared to the healthy controls (16.9% vs 10.9%). The ageing population might still develop this condition while the aged controls have less chances to have a late onset for AD.

Finally, consistent with our initial hypothesis of an association between AD and *REST* VNTR, the Hardy–Weinberg prediction highlights unexpected genotype combinations in our AD cohort ($p<0.01$ $df=3$). In particular the 7_7 and the 12_12 genotypes are higher and at the same time the 7_9 and 7_12 genotypes are lower than what we would expect to find in the general population.

On the contrary, in our aged cohort the only difference between the observed and the expected values, is the 12_12 genotype with a slight reduction in the observed samples. This difference is not present in the ageing cohort and that is concealed in the pooled control cohort.

Unfortunately due to the relatively small sample size we have not been able to show enough statistical power to validate our genotype in AD.

However a clear trend can be observed especially when we increased the control samples size pooling together the aged and ageing control groups, although no statistical difference can be observed (**Table 6.2**).

REST has been also recently indirectly linked to schizophrenia, either as a regulator of a Genome-Wide Associated Gene MIR137 (Warburton *et al.*, 2015) or negatively regulating the normal developmental switch of the 2B subunit in the N-methyl-D-aspartate (NMDA) receptor which is thought to be responsible for a

greater hippocampal excitement and consequently source of the psychosis symptoms in schizophrenia (Tamminga and Zukin, 2015).

In the first case a VNTR in mir137 promoter region has been proven to be differentially regulated by REST where overexpression of REST reduces mir137 expression, which is positively linked to schizophrenia.

In the second case, REST will be necessary for the REST-mediated subunit switch in the NMDA receptors. Failure of this switch will result in a greater expression and presence of GluN2B in the NMDA receptors with an immature phenotype which is more observed in the CA3 of humans with schizophrenia.

Once again the relatively small sample size, and the lack of clinical information, does not allow us to clarify REST involvement in this condition, although homozygosity seems to play an important role in this condition as well as for AD.

Given the importance of this protein from the early development to the late adulthood an imbalance of its expression, either too high or too low, could represent a tangible risk and a really important biomarker for detecting this condition way before its onset, thus could provide the basis for a tailored genetic treatment in patients.

Further analysis will be required and these data will have to be repeated in a much larger cohort to prove statistical power, nevertheless this finding might give great insight in the future of this pathology.

Chapter 7

Final Conclusions

7.1 Project overview

In this investigation we have given new insights into the role of novel genetic polymorphisms and we expanded the knowledge of other already known and well established GxE susceptible loci.

Their localisation in highly CNS disease-sensitive gene promoter regions further increases their importance in their role as gene regulators in response to environmental stimuli.

In relation to our initial project aims and objectives the following observations can be made:

- In the neuroblastoma cell line SH-SY5Y the *MAOA* promoter region is susceptible to chromatin modifications that allow single stranded binding factor to bind DNA after exposure to sodium valproate (**Figure 3.6**) and at the same time displayed a variable methylation pattern in the same genetic region (**Figure 3.7**)
- We have been able to differentiate the expression of the two *MAOA* mRNA isoforms by basing our measurements on an alternative splicing exon within the second isoform and comparing it to the 5′ UTR where the uVNTR is located (**Figures 4.13 – 4.14**)
- The *MAOA* VNTRs KO cell line model HAP1 proved, through our experiments, that the uVNTR is a negative regulator of the second *MAOA* isoform and it doesn't have any activity on the primary *MAOA* isoform, despite the fact that it has always been defined as positive regulator (**Figures 4.13 – 4.17**)

- The major role in *MAOA* gene expression is attributed to the less studied dVNTR that proved itself to be a positive regulator of the primary *MAOA* isoform and a negative regulator of the second isoform (**Figure 4.24**).
- Our data strongly suggest that the mood stabiliser, sodium valproate, exerts its function through the uVNTR (**Figure 4.23**)
- The GWAS candidate gene for schizophrenia *CACNA1C* contains a VNTR in its promoter region that, through bioinformatic analysis, has been shown to be present in less than 2% in the population
- The VNTR (TR1) in the *CACNA1C* promoter has been proven to contain a positive regulator while the TR2 contains a negative regulator (**Figure 5.3**)
- *CACNA1C* minimal promoter directs gene expression in our reporter gene constructs and it responds in a stimulus-inducible manner when subjected to either cocaine or lithium chloride (**Figure 5.4**)
- We identified a novel polymorphic microsatellite within the *REST* gene promoter region. Each variant, compared to each other, has been proven to drive expression gene constructs significantly (**Figure 6.5**)
- No significant difference can be found among the genotyped FTD, SCZ or AD cohorts. However the Hardy–Weinberg prediction show a significant difference within the AD cohort (**Figure 6.7**) strongly suggesting a correlation within this condition and the *REST* VNTR genotype.

  However, because of the relatively small sample size we analysed, we were not able to prove statistical significance.

7.2 The role of VNTR polymorphisms

It is now widely accepted that both genetic and environmental factors shape our own individuality and our behaviours and are able to create the basis for CNS diseases and conditions. It was already common knowledge that genetic polymorphic variants within the population may represent either protective or risk factors for several conditions. However the deep relation between genes and environment is still not completely understood.

Microsatellites and VNTRs constitute 3% of the entire human genome (Lander *et al.*, 2001) and they have been widely involved in gene regulation processes (Gymrek *et al.*, 2016, Li *et al.*, 2007). Furthermore, a wide range of microsatellites and VNTR have been associated to just as many diseases (**Table 1.2**), CNS conditions and behavioural traits (**Table 1.1**).

However, despite their importance, VNTRs and microsatellites are often poorly studied due to the peculiar difficulties that these genetic regions possess. The high CG content that usually constitute them and the repetitive nature *per-se* of these regions, are among the factors that persuade scientists to go for easier ways.

SNPs are often better candidates in terms of efficiencies and effectiveness in scientific studies. Several equipments have been developed and a wide range of bioinformatic tools have been designed in order to facilitate SNP studies.

In the last decade, however, the advancement in DNA-sequencing technologies and better bioinformatic tools, made possible a deeper understanding of these repetitive and polymorphic genetic regions (Treangen and Salzberg, 2012).

7.3 The *MAOA* VNTRs

The *MAOA* uVNTR has been implicated, as a risk factor, in a wide range of behavioural processes and CNS conditions (**Table 1.4**). However, our data in Chapters 3 and 4 strongly suggest a primary role of the less studied dVNTR in the MAOA canonical protein expression, while the uVNTR appears to not be involved in regulating this isoform expression.

Both VNTRs negatively regulate the expression of a secondary *MAOA* isoform, containing the uVNTR within the longer 5′UTR (**Figure 4.26**), and generating a shorter protein that hypothetically loses the ability to break down monoamine neurotransmitters, as the FAD binding domain is disrupted.

Furthermore, as reported in Chapter 3, the mRNA containing the uVNTR in the longer 5′UTR, appears to be more responsive to external stimuli, in particular the 3R variant of the *MAOA* uVNTR in comparison to the 4R variant. In Chapter 4, I validated the responsiveness to external stimuli of this isoform and the 3R uVNTR in the haploid cell line HAP1 (**Figure 4.23**).

Overall, these results reveal a dual role of the dVNTR, as it can be an activator or a repressor depending on the isoform that is going to be transcribed. In addition, these data point to a previously ignored *MAOA* mRNA isoform that could be responsible for the association that has been made between the *MAOA* gene and CNS conditions from several association studies.

Data generated from our group strongly suggest linkage disequilibrium between these VNTRs, with a further layer of stratification in the behavioural traits that specific genotypes generate in response to environmental stimuli (data not published). If taken into consideration, the expression data on the longer 5′UTR *MAOA* isoform, and the behavioural traits associated with a particular genotype,

might represent the explanation of the strong GxE interaction that the *MAOA* gene displays.

7.4 The *CACNA1C* gene

The association between the *CACNA1C* gene and SCZ is one of the most reproduced by GWAS studies (Schizophrenia Psychiatric Genome-Wide Association Study, 2011). Several SNPs within *CACNA1C* introns have been repeatedly linked to SCZ (Eckart *et al.*, 2016, Nyegaard *et al.*, 2010). However, linkage disequilibrium analysis on these SNPs and the two *CACNA1C* promoter region's TRs, on which we focused our attention, failed to produce any significant association (data not shown).

The analysis on the TR located on the far 5′ side of the *CACNA1C* promoter region (TR2 – **Figure 5.2**) revealed no polymorphic nature. Bioinformatic analysis on TR1, however, showed a polymorphic in-del: rs530020760 (chr12:2162060-2162067 Hg19) which is a deletion of the "CGGG,CGGG" repeat motif seen in 2.08% of the population and rs548465087 (chr12:2162068-21620670 Hg19), which is an insertion of the "CGGG" repeat motif seen in 97.92% of the normal population.

Reporter gene construct assays revealed that TR1 contains a positive regulator while the TR2 contains a negative regulator (**Figure 5.2**). Furthermore, the expression from the reporter gene constructs is directed by the CACNA1C minimal promoter and it is able to respond in a stimulus-inducible manner when subjected to either cocaine or lithium chloride (**Figure 5.3**).

The calcium current directed by the calcium channels is fundamental in both cardiac muscle tissue and in the brain (Christel and Lee, 2012, Striessnig *et al.*, 2014). These data on the *CACNA1C* promoter region and the GxE interaction displayed represent a step forward, although little, in the understanding of this gene

expression and its relation to external stimuli. In the future, the data presented in this thesis might be of great relevance for personalised medicine and shed light into the functioning and regulation of this elaborate gene.

7.5 The *REST* VNTR

The activity and levels of REST have been associated with several pathological conditions as previously highlighted in *section 1.10*. Recently, Lu *et al.* (2014) and Goetzl *et al.* (2015) reported a marked reduction of REST in AD patients.

In this investigation we reported a novel polymorphic microsatellite within the *REST* promoter region. Two variants of this microsatellite have been previously reported by the STR web browser and a third one has been experimentally found in our study. The reporter gene constructs, generated with the three variants, proved to be able to support expression. In addition, the expression can be differentially modulated depending on the copy variant (**Figure 6.5**).

Our genotype data suggest an association of this microsatellite with AD according to the Hardy–Weinberg (H-W) test. However, the relatively small sample size did not show enough statistical power to prove a significant association with AD through the clump test.

Nonetheless, these results, if further investigated, might be able to open a new path in the understanding of this pathology, better chances of a early diagnosis and better prevention strategies, if not a personalised genetic therapy.

7.6 Limitations of the work

In chapter 3 we have verified the expression of the *MAOA* mRNA utilising the uVNTR primer set in order to target the 5'UTR of this specific isoform. However, we have only been able to assess the presence of this *MAOA* mRNA in few human samples. As shown in **Figure 3.4,** clear differences are present in the expression of *MAOA* associated to different VNTRs (3R – 4R). An association of the cortisol levels for each sample to a possible differential expression would have strengthen the finding in the cell line model where the different *MAOA* alleles responded in a specific allelic manner (**Figure 3.5**)

In chapter 4, three distinct mRNA isoforms have been isolated utilising different primer sets that allowed us to distinguish each isoform. However, a deeper analysis at the protein level would be necessary in order to definitely prove the translation of these mRNA isoform into proteins. In addition, an activity study of the different MAOA isoforms would be necessary. If the FAD binding domain lacking in the short (**Figures 4.1 C – 4.3 – 4.7**) and medium (**Figures 4.1 B – 4.6**) MAOA is absolutely necessary for a correct activity of these isoforms the resulting proteins would be inactive. In addition a quantitative PCR of the different isoforms would strengthen the mRNA expression profile analysed in the chapter.

In chapter 6 the relative small cohort did not allow us enough statistical power for the REST VNTR analysis. The study should be repeated in a much larger cohort.

7.7 Future studies

For a complete analysis of *MAOA*, methylation and expression profile analysis in human samples associated to clinical data should be preferential. This will confirm and clarify a differential allelic expression due to the different uVNTR alleles.

After overexpression of the MAOA isoforms (short and medium) the activity should be assessed. In alternative to the overexpression, MAOA could be enriched through antybody precipitation. It would be important to confirm the transcription factor binding in the HAP1 KO cell lines through ChIP analysis and correlate these to the expression profile, preferentially with qPCR specifically targeting the different *MAOA* isoforms

In chapter 5 an activator and a repressor have been found in close proximity of the *CACNA1C* TSS in the P2 and P3 constructs respectively. The reporter gene expression assay with the different constructs should be repeated in association with EZH2 overexpression. Being EZH2 an important regulator of several developmental processes and its specific binding sequence present in several calcium channels, it might give insights for a better understanding of the CACNA1C and other calcium channels dysregulation found in schizophrenic patients.

Chapter 8

Reference List

ALI, F. R., VASILIOU, S. A., HADDLEY, K., PAREDES, U. M., ROBERTS, J. C., MIYAJIMA, F., KLENOVA, E., BUBB, V. J. & QUINN, J. P. 2010. Combinatorial interaction between two human serotonin transporter gene variable number tandem repeats and their regulation by CTCF. *J Neurochem,* 112**,** 296-306.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-402.

ALVAREZ, S., MAS, S., GASSO, P., BERNARDO, M., PARELLADA, E. & LAFUENTE, A. 2010. Lack of association between schizophrenia and polymorphisms in dopamine metabolism and transport genes. *Fundam Clin Pharmacol,* 24**,** 741-7.

ALZHEIMER'S ASSOCIATION 2015. 2015 Alzheimer's disease facts and figures. *Alzheimers Dement,* 11**,** 332-84.

ANDRES, A. M., SOLDEVILA, M., NAVARRO, A., KIDD, K. K., OLIVA, B. & BERTRANPETIT, J. 2004. Positive selection in MAOA gene is human exclusive: determination of the putative amino acid change selected in the human lineage. *Hum Genet,* 115**,** 377-86.

ANDRES, M. E., BURGER, C., PERAL-RUBIO, M. J., BATTAGLIOLI, E., ANDERSON, M. E., GRIMES, J., DALLMAN, J., BALLAS, N. & MANDEL, G. 1999. CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc Natl Acad Sci U S A,* 96**,** 9873-8.

ARMAS, P., MARGARIT, E., MOUGUELAR, V. S., ALLENDE, M. L. & CALCATERRA, N. B. 2013. Beyond the binding site: in vivo identification of tbx2, smarca5 and wnt5b as molecular targets of CNBP during embryonic development. *PLoS One,* 8**,** e63234.

ARNETT, C. D., FOWLER, J. S., MACGREGOR, R. R., SCHLYER, D. J., WOLF, A. P., LANGSTROM, B. & HALLDIN, C. 1987. Turnover of brain monoamine oxidase measured in vivo by positron emission tomography using L-[11C]deprenyl. *J Neurochem,* 49**,** 522-7.

ASLUND, C., NORDQUIST, N., COMASCO, E., LEPPERT, J., ORELAND, L. & NILSSON, K. W. 2011. Maltreatment, MAOA, and delinquency: sex differences in gene-environment interaction in a large population-based cohort of adolescents. *Behav Genet,* 41**,** 262-72.

BALASUBRAMANIAN, S., HURLEY, L. H. & NEIDLE, S. 2011. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov,* 10**,** 261-75.

BALLAS, N., BATTAGLIOLI, E., ATOUF, F., ANDRES, M. E., CHENOWETH, J., ANDERSON, M. E., BURGER, C., MONIWA, M., DAVIE, J. R., BOWERS, W. J., FEDEROFF, H. J., ROSE, D. W., ROSENFELD, M. G., BREHM, P. & MANDEL, G. 2001. Regulation of neuronal traits by a novel transcriptional complex. *Neuron,* 31**,** 353-65.

BALLAS, N., GRUNSEICH, C., LU, D. D., SPEH, J. C. & MANDEL, G. 2005. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell,* 121**,** 645-57.

BARDEN, N. 2004. Implication of the hypothalamic-pituitary-adrenal axis in the physiopathology of depression. *J Psychiatry Neurosci,* 29**,** 185-93.

BIGOS, K. L., MATTAY, V. S., CALLICOTT, J. H., STRAUB, R. E., VAKKALANKA, R., KOLACHANA, B., HYDE, T. M., LIPSKA, B. K., KLEINMAN, J. E. & WEINBERGER, D. R. 2010. Genetic variation in CACNA1C affects brain circuitries related to mental illness. *Arch Gen Psychiatry,* 67**,** 939-45.

BOCHMAN, M. L., PAESCHKE, K. & ZAKIAN, V. A. 2012. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet,* 13**,** 770-80.

BREEN, G., COLLIER, D., CRAIG, I. & QUINN, J. 2008. Variable number tandem repeats as agents of functional regulation in the genome. *IEEE Eng Med Biol Mag,* 27**,** 103-4, 108.

BROOKS, T. A. & HURLEY, L. H. 2009. The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat Rev Cancer,* 9**,** 849-61.

BRUCE, A. W., DONALDSON, I. J., WOOD, I. C., YERBURY, S. A., SADOWSKI, M. I., CHAPMAN, M., GOTTGENS, B. & BUCKLEY, N. J. 2004. Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc Natl Acad Sci U S A,* 101**,** 10458-63.

BRUCE, A. W., LOPEZ-CONTRERAS, A. J., FLICEK, P., DOWN, T. A., DHAMI, P., DILLON, S. C., KOCH, C. M., LANGFORD, C. F., DUNHAM, I., ANDREWS, R. M. & VETRIE, D. 2009. Functional diversity for REST (NRSF) is defined by in vivo binding affinity hierarchies at the DNA sequence level. *Genome Res,* 19**,** 994-1005.

CALDERONE, A., JOVER, T., NOH, K. M., TANAKA, H., YOKOTA, H., LIN, Y., GROOMS, S. Y., REGIS, R., BENNETT, M. V. & ZUKIN, R. S. 2003. Ischemic insults derepress the gene silencer REST in neurons destined to die. *J Neurosci,* 23**,** 2112-21.

CAPRA, J. A., PAESCHKE, K., SINGH, M. & ZAKIAN, V. A. 2010. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS Comput Biol,* 6**,** e1000861.

CARETTE, J. E., RAABEN, M., WONG, A. C., HERBERT, A. S., OBERNOSTERER, G., MULHERKAR, N., KUEHNE, A. I., KRANZUSCH, P. J., GRIFFIN, A. M., RUTHEL, G., DAL CIN, P., DYE, J. M., WHELAN, S. P., CHANDRAN, K. & BRUMMELKAMP, T. R. 2011. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature,* 477**,** 340-3.

CARREL, L. & WILLARD, H. F. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature,* 434**,** 400-4.

CASPI, A., MCCLAY, J., MOFFITT, T. E., MILL, J., MARTIN, J., CRAIG, I. W., TAYLOR, A. & POULTON, R. 2002. Role of genotype in the cycle of violence in maltreated children. *Science,* 297**,** 851-4.

CHEN, Z. Y., HOTAMISLIGIL, G. S., HUANG, J. K., WEN, L., EZZEDDINE, D., AYDIN-MUDERRISOGLU, N., POWELL, J. F., HUANG, R. H., BREAKEFIELD, X. O., CRAIG, I. & ET AL. 1991. Structure of the human gene for monoamine oxidase type A. *Nucleic Acids Res,* 19**,** 4537-41.

CHEN, Z. Y., POWELL, J. F., HSU, Y. P., BREAKEFIELD, X. O. & CRAIG, I. W. 1992. Organization of the human monoamine oxidase genes and long-range physical mapping around them. *Genomics,* 14**,** 75-82.

CHESTER, D. S., DEWALL, C. N., DEREFINKO, K. J., ESTUS, S., PETERS, J. R., LYNAM, D. R. & JIANG, Y. 2015. Monoamine oxidase A (MAOA) genotype predicts greater aggression through impulsive reactivity to negative affect. *Behav Brain Res,* 283**,** 97-101.

CHOI, J. K. & KIM, Y. J. 2008. Epigenetic regulation and the variability of gene expression. *Nat Genet,* 40**,** 141-7.

CHONG, J. A., TAPIA-RAMIREZ, J., KIM, S., TOLEDO-ARAL, J. J., ZHENG, Y., BOUTROS, M. C., ALTSHULLER, Y. M., FROHMAN, M. A., KRANER, S. D. & MANDEL, G. 1995. REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell,* 80**,** 949-57.

CHOUDHARY, C., KUMAR, C., GNAD, F., NIELSEN, M. L., REHMAN, M., WALTHER, T. C., OLSEN, J. V. & MANN, M. 2009. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science,* 325**,** 834-40.

CHRISTEL, C. & LEE, A. 2012. Ca2+-dependent modulation of voltage-gated Ca2+ channels. *Biochim Biophys Acta,* 1820**,** 1243-52.

COHEN, I. L., LIU, X., SCHUTZ, C., WHITE, B. N., JENKINS, E. C., BROWN, W. T. & HOLDEN, J. J. 2003. Association of autism severity with a monoamine oxidase A functional polymorphism. *Clin Genet,* 64**,** 190-7.

CORNISH, K. M., MANLY, T., SAVAGE, R., SWANSON, J., MORISANO, D., BUTLER, N., GRANT, C., CROSS, G., BENTLEY, L. & HOLLIS, C. P. 2005. Association of the dopamine transporter (DAT1) 10/10-repeat genotype with ADHD symptoms and response inhibition in a general population sample. *Mol Psychiatry,* 10**,** 686-98.

DAPIC, V., ABDOMEROVIC, V., MARRINGTON, R., PEBERDY, J., RODGER, A., TRENT, J. O. & BATES, P. J. 2003. Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Res,* 31**,** 2097-107.

DE COLIBUS, L., LI, M., BINDA, C., LUSTIG, A., EDMONDSON, D. E. & MATTEVI, A. 2005. 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,* 102**,** 12684-9.

DE LA SERNA, I. L. & IMBALZANO, A. N. 2002. Unfolding heterochromatin for replication. *Nat Genet,* 32**,** 560-2.

DEJESUS-HERNANDEZ, M., MACKENZIE, I. R., BOEVE, B. F., BOXER, A. L., BAKER, M., RUTHERFORD, N. J., NICHOLSON, A. M., FINCH, N. A., FLYNN, H., ADAMSON, J., KOURI, N., WOJTAS, A., SENGDY, P., HSIUNG, G. Y., KARYDAS, A., SEELEY, W. W., JOSEPHS, K. A., COPPOLA, G., GESCHWIND, D. H., WSZOLEK, Z. K., FELDMAN, H., KNOPMAN, D. S., PETERSEN, R. C., MILLER, B. L., DICKSON, D. W., BOYLAN, K. B., GRAFF-RADFORD, N. R. & RADEMAKERS, R. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron,* 72**,** 245-56.

DELISI, L. E., MAURIZIO, A. M., SVETINA, C., ARDEKANI, B., SZULC, K., NIERENBERG, J., LEONARD, J. & HARVEY, P. D. 2005. Klinefelter's syndrome (XXY) as a genetic model for psychotic disorders. *Am J Med Genet B Neuropsychiatr Genet,* 135B**,** 15-23.

DELLA CORTE, L. & TIPTON, K. F. 1980. The turnover of the A- and B-forms of monoamine oxidase in rat liver. *Biochem Pharmacol,* 29**,** 891-5.

DEVYS, D., LUTZ, Y., ROUYER, N., BELLOCQ, J. P. & MANDEL, J. L. 1993. The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat Genet,* 4**,** 335-40.

DIETRICH, N., LERDRUP, M., LANDT, E., AGRAWAL-SINGH, S., BAK, M., TOMMERUP, N., RAPPSILBER, J., SODERSTEN, E. & HANSEN, K. 2012. REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS Genet,* 8**,** e1002494.

DOLPHIN, A. C. 2012. Calcium channel auxiliary alpha2delta and beta subunits: trafficking and one step beyond. *Nat Rev Neurosci,* 13**,** 542-55.

DOMSCHKE, K., TIDOW, N., KUITHAN, H., SCHWARTE, K., KLAUKE, B., AMBREE, O., REIF, A., SCHMIDT, H., AROLT, V., KERSTING, A., ZWANZGER, P. & DECKERT, J. 2012. Monoamine oxidase A gene DNA hypomethylation - a risk factor for panic disorder? *Int J Neuropsychopharmacol,* 15**,** 1217-28.

DU, L., FALUDI, G., PALKOVITS, M., SOTONYI, P., BAKISH, D. & HRDINA, P. D. 2002. High activity-related allele of MAO-A gene associated with depressed suicide in males. *Neuroreport,* 13**,** 1195-8.

EBERSBERGER, I., METZLER, D., SCHWARZ, C. & PAABO, S. 2002. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am J Hum Genet,* 70**,** 1490-7.

ECKART, N., SONG, Q., YANG, R., WANG, R., ZHU, H., MCCALLION, A. S. & AVRAMOPOULOS, D. 2016. Functional Characterization of Schizophrenia-Associated Variation in CACNA1C. *PLoS One,* 11**,** e0157086.

EGGER, G., LIANG, G., APARICIO, A. & JONES, P. A. 2004. Epigenetics in human disease and prospects for epigenetic therapy. *Nature,* 429**,** 457-63.

ENCODE PROJECT CONSORTIUM 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

FAN, M., LIU, B., JIANG, T., JIANG, X., ZHAO, H. & ZHANG, J. 2010. Meta-analysis of the association between the monoamine oxidase-A gene and mood disorders. *Psychiatr Genet,* 20**,** 1-7.

FELSENFELD, G. & GROUDINE, M. 2003. Controlling the double helix. *Nature,* 421**,** 448-53.

FERGUSSON, D. M., BODEN, J. M., HORWOOD, L. J., MILLER, A. & KENNEDY, M. A. 2012. Moderating role of the MAOA genotype in antisocial behaviour. *Br J Psychiatry,* 200**,** 116-23.

FERNANDEZ-CASTILLO, N., CABANA-DOMINGUEZ, J., SORIANO, J., SANCHEZ-MORA, C., RONCERO, C., GRAU-LOPEZ, L., ROS-CUCURULL, E., DAIGRE, C., VAN DONKELAAR, M. M., FRANKE, B., CASAS, M., RIBASES, M. & CORMAND, B. 2015. Transcriptomic and genetic studies identify NFAT5 as a candidate gene for cocaine dependence. *Transl Psychiatry,* 5**,** e667.

FISCHLE, W., WANG, Y. & ALLIS, C. D. 2003. Histone and chromatin cross-talk. *Curr Opin Cell Biol,* 15**,** 172-83.

FONDON, J. W., 3RD & GARNER, H. R. 2004. Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci U S A,* 101**,** 18058-63.

FONDON, J. W., 3RD, HAMMOCK, E. A., HANNAN, A. J. & KING, D. G. 2008. Simple sequence repeats: genetic modulators of brain function and behavior. *Trends Neurosci,* 31**,** 328-34.

FRAZZETTO, G., DI LORENZO, G., CAROLA, V., PROIETTI, L., SOKOLOWSKA, E., SIRACUSANO, A., GROSS, C. & TROISI, A. 2007. Early trauma and increased risk for physical aggression during adulthood: the moderating role of MAOA genotype. *PLoS One,* 2**,** e486.

FROMER, M., POCKLINGTON, A. J., KAVANAGH, D. H., WILLIAMS, H. J., DWYER, S., GORMLEY, P., GEORGIEVA, L., REES, E., PALTA, P., RUDERFER, D. M., CARRERA, N., HUMPHREYS, I., JOHNSON, J. S., ROUSSOS, P., BARKER, D. D., BANKS, E., MILANOVA, V., GRANT, S. G., HANNON, E., ROSE, S. A., CHAMBERT, K., MAHAJAN, M., SCOLNICK, E. M., MORAN, J. L., KIROV, G., PALOTIE, A., MCCARROLL, S. A., HOLMANS, P., SKLAR, P., OWEN, M. J., PURCELL, S. M. & O'DONOVAN, M. C. 2014. De novo mutations in schizophrenia implicate synaptic networks. *Nature,* 506**,** 179-84.

FUDA, N. J., ARDEHALI, M. B. & LIS, J. T. 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature,* 461**,** 186-92.

FUKS, F., BURGERS, W. A., BREHM, A., HUGHES-DAVIES, L. & KOUZARIDES, T. 2000. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat Genet,* 24**,** 88-91.

GALINDO, C. L., MCCORMICK, J. F., BUBB, V. J., ABID ALKADEM, D. H., LI, L. S., MCIVER, L. J., GEORGE, A. C., BOOTHMAN, D. A., QUINN, J. P., SKINNER, M. A. & GARNER, H. R. 2011. A long AAAG repeat allele in the 5' UTR of the ERR-gamma gene is correlated with breast cancer predisposition and drives promoter activity in MCF-7 breast cancer cells. *Breast Cancer Res Treat,* 130**,** 41-8.

GAO, Z., URE, K., DING, P., NASHAAT, M., YUAN, L., MA, J., HAMMER, R. E. & HSIEH, J. 2011. The master negative regulator REST/NRSF controls adult neurogenesis by restraining the neurogenic program in quiescent stem cells. *J Neurosci,* 31**,** 9772-86.

GARBER, K., SMITH, K. T., REINES, D. & WARREN, S. T. 2006. Transcription, translation and fragile X syndrome. *Curr Opin Genet Dev,* 16**,** 270-5.

GASCON, E. & GAO, F. B. 2012. Cause or Effect: Misregulation of microRNA Pathways in Neurodegeneration. *Front Neurosci,* 6**,** 48.

GAWESKA, H. & FITZPATRICK, P. F. 2011. Structures and Mechanism of the Monoamine Oxidase Family. *Biomol Concepts,* 2**,** 365-377.

GOETZL, E. J., BOXER, A., SCHWARTZ, J. B., ABNER, E. L., PETERSEN, R. C., MILLER, B. L., CARLSON, O. D., MUSTAPIC, M. & KAPOGIANNIS, D. 2015. Low neural exosomal levels of cellular survival factors in Alzheimer's disease. *Ann Clin Transl Neurol,* 2**,** 769-73.

GOLDMAN, N., GLEI, D. A., LIN, Y. H. & WEINSTEIN, M. 2010. The serotonin transporter polymorphism (5-HTTLPR): allelic variation and links with depressive symptoms. *Depress Anxiety,* 27**,** 260-9.

GRIMES, J. A., NIELSEN, S. J., BATTAGLIOLI, E., MISKA, E. A., SPEH, J. C., BERRY, D. L., ATOUF, F., HOLDENER, B. C., MANDEL, G. & KOUZARIDES, T. 2000. The co-repressor mSin3A is a functional component of the REST-CoREST repressor complex. *J Biol Chem,* 275**,** 9461-7.

GRIMSBY, J., CHEN, K., WANG, L. J., LAN, N. C. & SHIH, J. C. 1991. Human monoamine oxidase A and B genes exhibit identical exon-intron organization. *Proc Natl Acad Sci U S A,* 88**,** 3637-41.

GROCHANS, E., GRZYWACZ, A., JURCZAK, A., SAMOCHOWIEC, A., KARAKIEWICZ, B., BRODOWSKA, A., STARCZEWSKI, A. & SAMOCHOWIEC, J. 2013. The 5HTT and MAO-A polymorphisms associate with depressive mood and climacteric symptoms in postmenopausal women. *Prog Neuropsychopharmacol Biol Psychiatry,* 45**,** 125-30.

GROCHANS, E., JURCZAK, A., SZKUP, M., SAMOCHOWIEC, A., WLOSZCZAK-SZUBZDA, A., KARAKIEWICZ, B., GRZYWACZ, A., BRODOWSKA, A. & SAMOCHOWIEC, J. 2015. Evaluation of the relationship between 5-HTT and MAO gene polymorphisms, mood and level of anxiety among postmenopausal women. *Int J Environ Res Public Health,* 12**,** 268-81.

GUAN, F., ZHANG, B., YAN, T., LI, L., LIU, F., LI, T., FENG, Z., ZHANG, B., LIU, X. & LI, S. 2014. MIR137 gene and target gene CACNA1C of miR-137 contribute to schizophrenia susceptibility in Han Chinese. *Schizophr Res,* 152**,** 97-104.

GUINDALINI, C., HOWARD, M., HADDLEY, K., LARANJEIRA, R., COLLIER, D., AMMAR, N., CRAIG, I., O'GARA, C., BUBB, V. J., GREENWOOD, T., KELSOE, J., ASHERSON, P., MURRAY, R. M., CASTELO, A., QUINN, J. P., VALLADA, H. & BREEN, G. 2006. A dopamine transporter gene functional variant associated with cocaine abuse in a Brazilian sample. *Proc Natl Acad Sci U S A,* 103**,** 4552-7.

GUSELLA, J. F. & MACDONALD, M. E. 2006. Huntington's disease: seeing the pathogenic process through a genetic lens. *Trends Biochem Sci,* 31**,** 533-40.

GWAS CONSORTIUM BIPOLAR DISORDER WORKING GROUP, P. 2011. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet,* 43**,** 977-83.

GYMREK, M., WILLEMS, T., GUILMATRE, A., ZENG, H., MARKUS, B., GEORGIEV, S., DALY, M. J., PRICE, A. L., PRITCHARD, J. K., SHARP, A. J. & ERLICH, Y. 2016. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat Genet,* 48**,** 22-9.

HADDLEY, K., BUBB, V. J., BREEN, G., PARADES-ESQUIVEL, U. M. & QUINN, J. P. 2012. Behavioural genetics of the serotonin transporter. *Curr Top Behav Neurosci,* 12**,** 503-35.

HADDLEY, K., VASILIOU, A. S., ALI, F. R., PAREDES, U. M., BUBB, V. J. & QUINN, J. P. 2008. Molecular genetics of monoamine transporters: relevance to brain disorders. *Neurochem Res,* 33**,** 652-67.

HANNAN, A. J. 2010. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. *Trends Genet,* 26**,** 59-65.

HEILS, A., TEUFEL, A., PETRI, S., STOBER, G., RIEDERER, P., BENGEL, D. & LESCH, K. P. 1996. Allelic variation of human serotonin transporter gene expression. *J Neurochem,* 66**,** 2621-4.

HERMAN, J. G. & BAYLIN, S. B. 2003. Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med,* 349**,** 2042-54.

HILL, J., BREEN, G., QUINN, J., TIBU, F., SHARP, H. & PICKLES, A. 2013. 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 Behav,* 12**,** 388-96.

HINDORFF, L. A., SETHUPATHY, P., JUNKINS, H. A., RAMOS, E. M., MEHTA, J. P., COLLINS, F. S. & MANOLIO, T. A. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A,* 106**,** 9362-7.

HING, B., DAVIDSON, S., LEAR, M., BREEN, G., QUINN, J., MCGUFFIN, P. & MACKENZIE, A. 2012. A polymorphism associated with depressive disorders differentially regulates brain derived neurotrophic factor promoter IV activity. *Biol Psychiatry,* 71**,** 618-26.

HO, L. W., FURLONG, R. A., RUBINSZTEIN, J. S., WALSH, C., PAYKEL, E. S. & RUBINSZTEIN, D. C. 2000. Genetic associations with clinical characteristics in bipolar affective disorder and recurrent unipolar depressive disorder. *Am J Med Genet,* 96**,** 36-42.

HUANG, Y., MYERS, S. J. & DINGLEDINE, R. 1999. Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. *Nat Neurosci,* 2**,** 867-72.

HUANG, Y. Y., CATE, S. P., BATTISTUZZI, C., OQUENDO, M. A., BRENT, D. & MANN, J. J. 2004. An association between a functional polymorphism in the monoamine oxidase a gene promoter, impulsive traits and early abuse experiences. *Neuropsychopharmacology,* 29**,** 1498-505.

HUBLITZ, P., ALBERT, M. & PETERS, A. H. 2009. Mechanisms of transcriptional repression by histone lysine methylation. *Int J Dev Biol,* 53**,** 335-54.

HUNG, C. F., LUNG, F. W., HUNG, T. H., CHONG, M. Y., WU, C. K., WEN, J. K. & LIN, P. Y. 2012. Monoamine oxidase A gene polymorphism and suicide: an association study and meta-analysis. *J Affect Disord,* 136**,** 643-9.

HUPPERT, J. L. 2008. Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem Soc Rev,* 37**,** 1375-84.

HUPPERT, J. L. 2010. Structure, location and interactions of G-quadruplexes. *FEBS J,* 277**,** 3452-8.

HUPPERT, J. L., BUGAUT, A., KUMARI, S. & BALASUBRAMANIAN, S. 2008. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res,* 36**,** 6260-8.

HURD, M. D., MARTORELL, P., DELAVANDE, A., MULLEN, K. J. & LANGA, K. M. 2013. Monetary costs of dementia in the United States. *N Engl J Med,* 368**,** 1326-34.

JI, B., HIGA, K. K., KELSOE, J. R. & ZHOU, X. 2015. Over-expression of XIST, the Master Gene for X Chromosome Inactivation, in Females With Major Affective Disorders. *EBioMedicine,* 2**,** 907-16.

JOHNSON, D. S., MORTAZAVI, A., MYERS, R. M. & WOLD, B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science,* 316**,** 1497-502.

JOO, J. E., NOVAKOVIC, B., CRUICKSHANK, M., DOYLE, L. W., CRAIG, J. M. & SAFFERY, R. 2014. Human active X-specific DNA methylation events showing stability across time and tissues. *Eur J Hum Genet,* 22**,** 1376-81.

KARLSGODT, K. H., SUN, D. & CANNON, T. D. 2010. Structural and Functional Brain Abnormalities in Schizophrenia. *Curr Dir Psychol Sci,* 19**,** 226-231.

KASHI, Y. & KING, D. G. 2006. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet,* 22**,** 253-9.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

KHOSRAVANI, H. & ZAMPONI, G. W. 2006. Voltage-gated calcium channels and idiopathic generalized epilepsies. *Physiol Rev,* 86**,** 941-66.

KIM-COHEN, J., CASPI, A., TAYLOR, A., WILLIAMS, B., NEWCOMBE, R., CRAIG, I. W. & MOFFITT, T. E. 2006. MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis. *Mol Psychiatry,* 11**,** 903-13.

KING, D. G., SOLLER, M. & KASHI, Y. 1997. Evolutionary tuning knobs. *Endeavour,* 21**,** 36-40.

KLEIN, C. J., BOTUYAN, M. V., WU, Y., WARD, C. J., NICHOLSON, G. A., HAMMANS, S., HOJO, K., YAMANISHI, H., KARPF, A. R., WALLACE, D. C., SIMON, M., LANDER, C., BOARDMAN, L. A., CUNNINGHAM, J. M., SMITH, G. E., LITCHY, W. J., BOES, B., ATKINSON, E. J., MIDDHA, S., PJ, B. D., PARISI, J. E., MER, G., SMITH, D. I. & DYCK, P. J. 2011. Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat Genet,* 43**,** 595-600.

KLENOVA, E., SCOTT, A. C., ROBERTS, J., SHAMSUDDIN, S., LOVEJOY, E. A., BERGMANN, S., BUBB, V. J., ROYER, H. D. & QUINN, J. P. 2004. YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders. *J Neurosci,* 24**,** 5966-73.

KORNBERG, R. D. 2005. Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci,* 30**,** 235-9.

KUNUGI, H., ISHIDA, S., KATO, T., TATSUMI, M., SAKAI, T., HATTORI, M., HIROSE, T. & NANKO, S. 1999. A functional polymorphism in the promoter region of monoamine oxidase-A gene and mood disorders. *Mol Psychiatry,* 4**,** 393-5.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, Y., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LASKY-SU, J. A., FARAONE, S. V., GLATT, S. J. & TSUANG, M. T. 2005. Meta-analysis of the association between two polymorphisms in the serotonin transporter gene and affective disorders. *Am J Med Genet B Neuropsychiatr Genet,* 133B**,** 110-5.

LEE, T. I. & YOUNG, R. A. 2000. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet,* 34**,** 77-137.

LI, B., CAREY, M. & WORKMAN, J. L. 2007. The role of chromatin during transcription. *Cell,* 128**,** 707-19.

LIN, Y. M., DAVAMANI, F., YANG, W. C., LAI, T. J. & SUN, H. S. 2008. Association analysis of monoamine oxidase A gene and bipolar affective disorder in Han Chinese. *Behav Brain Funct,* 4**,** 21.

LIU, Z., HUANG, L., LUO, X. J., WU, L. & LI, M. 2015. MAOA Variants and Genetic Susceptibility to Major Psychiatric Disorders. *Mol Neurobiol*.

LOBANENKOV, V. V., NICOLAS, R. H., ADLER, V. V., PATERSON, H., KLENOVA, E. M., POLOTSKAJA, A. V. & GOODWIN, G. H. 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene,* 5**,** 1743-53.

LOESCH, D. Z., BUI, Q. M., DISSANAYAKE, C., CLIFFORD, S., GOULD, E., BULHAK-PATERSON, D., TASSONE, F., TAYLOR, A. K., HESSL, D., HAGERMAN, R. & HUGGINS, R. M. 2007. Molecular and cognitive predictors of the continuum of autistic behaviours in fragile X. *Neurosci Biobehav Rev,* 31**,** 315-26.

LU, R. B., LEE, J. F., KO, H. C., LIN, W. W., CHEN, K. & SHIH, J. C. 2002. No association of the MAOA gene with alcoholism among Han Chinese males in Taiwan. *Prog Neuropsychopharmacol Biol Psychiatry,* 26**,** 457-61.

LU, T., ARON, L., ZULLO, J., PAN, Y., KIM, H., CHEN, Y., YANG, T. H., KIM, H. M., DRAKE, D., LIU, X. S., BENNETT, D. A., COLAIACOVO, M. P. & YANKNER, B. A. 2014. REST and stress resistance in ageing and Alzheimer's disease. *Nature,* 507**,** 448-54.

LUCZAK, M. W. & JAGODZINSKI, P. P. 2006. The role of DNA methylation in cancer development. *Folia Histochem Cytobiol,* 44**,** 143-54.

LUNYAK, V. V., BURGESS, R., PREFONTAINE, G. G., NELSON, C., SZE, S. H., CHENOWETH, J., SCHWARTZ, P., PEVZNER, P. A., GLASS, C., MANDEL, G. & ROSENFELD, M. G. 2002. Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science,* 298**,** 1747-52.

LYKKE-ANDERSEN, S. & JENSEN, T. H. 2007. Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie,* 89**,** 1177-82.

MACKENZIE, A., HING, B. & DAVIDSON, S. 2013. Exploring the effects of polymorphisms on cis-regulatory signal transduction response. *Trends Mol Med,* 19**,** 99-107.

MAJOUNIE, E., RENTON, A. E., MOK, K., DOPPER, E. G., WAITE, A., ROLLINSON, S., CHIO, A., RESTAGNO, G., NICOLAOU, N., SIMON-SANCHEZ, J., VAN SWIETEN, J. C., ABRAMZON, Y., JOHNSON, J. O., SENDTNER, M., PAMPHLETT, R., ORRELL, R. W., MEAD, S., SIDLE, K. C., HOULDEN, H., ROHRER, J. D., MORRISON, K. E., PALL, H., TALBOT, K., ANSORGE, O., CHROMOSOME, A. L. S. F. T. D. C., FRENCH RESEARCH NETWORK ON, F. F. A., CONSORTIUM, I., HERNANDEZ, D. G., AREPALLI, S., SABATELLI, M., MORA, G.,

CORBO, M., GIANNINI, F., CALVO, A., ENGLUND, E., BORGHERO, G., FLORIS, G. L., REMES, A. M., LAAKSOVIRTA, H., MCCLUSKEY, L., TROJANOWSKI, J. Q., VAN DEERLIN, V. M., SCHELLENBERG, G. D., NALLS, M. A., DRORY, V. E., LU, C. S., YEH, T. H., ISHIURA, H., TAKAHASHI, Y., TSUJI, S., LE BER, I., BRICE, A., DREPPER, C., WILLIAMS, N., KIRBY, J., SHAW, P., HARDY, J., TIENARI, P. J., HEUTINK, P., MORRIS, H. R., PICKERING-BROWN, S. & TRAYNOR, B. J. 2012. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol,* 11**,** 323-30.

MANOLI, I., LE, H., ALESCI, S., MCFANN, K. K., SU, Y. A., KINO, T., CHROUSOS, G. P. & BLACKMAN, M. R. 2005. Monoamine oxidase-A is a major target gene for glucocorticoids in human skeletal muscle cells. *FASEB J,* 19**,** 1359-61.

MATTICK, J. S. & MAKUNIN, I. V. 2005. Small regulatory RNAs in mammals. *Hum Mol Genet,* 14 Spec No 1**,** R121-32.

MATYS, V. 2003. TRANSFAC(R): transcriptional regulation, from patterns to profiles. *Nucleic Acids Res,* 31**,** 374-378.

MELAS, P. A., WEI, Y., WONG, C. C., SJOHOLM, L. K., ABERG, E., MILL, J., SCHALLING, M., FORSELL, Y. & LAVEBRATT, C. 2013. Genetic and epigenetic associations of MAOA and NR3C1 with depression and childhood adversities. *Int J Neuropsychopharmacol,* 16**,** 1513-28.

MEYER, J. H., GINOVART, N., BOOVARIWALA, A., SAGRATI, S., HUSSEY, D., GARCIA, A., YOUNG, T., PRASCHAK-RIEDER, N., WILSON, A. A. & HOULE, S. 2006. Elevated monoamine oxidase a levels in the brain: an explanation for the monoamine imbalance of major depression. *Arch Gen Psychiatry,* 63**,** 1209-16.

MICHELOTTI, E. F., MICHELOTTI, G. A., ARONSOHN, A. I. & LEVENS, D. 1996. Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol Cell Biol,* 16**,** 2350-60.

MICHELOTTI, E. F., TOMONAGA, T., KRUTZSCH, H. & LEVENS, D. 1995. Cellular nucleic acid binding protein regulates the CT element of the human c-myc protooncogene. *J Biol Chem,* 270**,** 9494-9.

MIGNONE, F., GISSI, C., LIUNI, S. & PESOLE, G. 2002. Untranslated regions of mRNAs. *Genome Biol,* 3**,** REVIEWS0004.

MIYAJIMA, F., QUINN, J. P., HORAN, M., PICKLES, A., OLLIER, W. E., PENDLETON, N. & PAYTON, A. 2008. Additive effect of BDNF and REST polymorphisms is associated with improved general cognitive ability. *Genes Brain Behav,* 7**,** 714-9.

MOFFITT, T. E., CASPI, A. & RUTTER, M. 2005. Strategy for investigating interactions between measured genes and measured environments. *Arch Gen Psychiatry,* 62**,** 473-81.

MOOSMANG, S., HAIDER, N., KLUGBAUER, N., ADELSBERGER, H., LANGWIESER, N., MULLER, J., STIESS, M., MARAIS, E., SCHULLA, V., LACINOVA, L., GOEBBELS, S., NAVE, K. A., STORM, D. R., HOFMANN, F. & KLEPPISCH, T. 2005. Role of hippocampal Cav1.2 Ca2+ channels in NMDA receptor-independent synaptic plasticity and spatial memory. *J Neurosci,* 25**,** 9883-92.

MORISHIMA, M., HARADA, N., HARA, S., SANO, A., SENO, H., TAKAHASHI, A., MORITA, Y. & NAKAYA, Y. 2006. Monoamine oxidase A activity and

norepinephrine level in hippocampus determine hyperwheel running in SPORTS rats. *Neuropsychopharmacology,* 31**,** 2627-38.

MUGFORD, J. W., STARMER, J., WILLIAMS, R. L., JR., CALABRESE, J. M., MIECZKOWSKI, P., YEE, D. & MAGNUSON, T. 2014. Evidence for local regulatory control of escape from imprinted X chromosome inactivation. *Genetics,* 197**,** 715-23.

MURGATROYD, C., HOFFMANN, A. & SPENGLER, D. 2012. In vivo ChIP for the analysis of microdissected tissue samples. *Methods Mol Biol,* 809**,** 135-48.

MURGATROYD, C. & SPENGLER, D. 2011. Epigenetics of early child development. *Front Psychiatry,* 2**,** 16.

MURGATROYD, C. A., PENA, C. J., PODDA, G., NESTLER, E. J. & NEPHEW, B. C. 2015. Early life social stress induced changes in depression and anxiety associated neural pathways which are correlated with impaired maternal care. *Neuropeptides,* 52**,** 103-11.

NAKAMURA, M., UENO, S., SANO, A. & TANABE, H. 2000. The human serotonin transporter gene linked polymorphism (5-HTTLPR) shows ten novel allelic variants. *Mol Psychiatry,* 5**,** 32-8.

NAM, K., MUNCH, K., HOBOLTH, A., DUTHEIL, J. Y., VEERAMAH, K. R., WOERNER, A. E., HAMMER, M. F., GREAT APE GENOME DIVERSITY, P., MAILUND, T. & SCHIERUP, M. H. 2015. Extreme selective sweeps independently targeted the X chromosomes of the great apes. *Proc Natl Acad Sci U S A,* 112**,** 6413-8.

NARUSE, Y., AOKI, T., KOJIMA, T. & MORI, N. 1999. Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. *Proc Natl Acad Sci U S A,* 96**,** 13691-6.

NCIRI, R., BOUJBIHA, M. A., JBAHI, S., ALLAGUI, M. S., ELFEKI, A., VINCENT, C. & CROUTE, F. 2015. Cytoskeleton involvement in lithium-induced SH-SY5Y neuritogenesis and the role of glycogen synthase kinase 3beta. *Aging Clin Exp Res,* 27**,** 255-63.

NIKULINA, V., WIDOM, C. S. & BRZUSTOWICZ, L. M. 2012. Child abuse and neglect, MAOA, and mental health outcomes: a prospective examination. *Biol Psychiatry,* 71**,** 350-7.

NILSSON, K. W., COMASCO, E., ASLUND, C., NORDQUIST, N., LEPPERT, J. & ORELAND, L. 2011. MAOA genotype, family relations and sexual abuse in relation to adolescent alcohol consumption. *Addict Biol,* 16**,** 347-55.

NYEGAARD, M., DEMONTIS, D., FOLDAGER, L., HEDEMAND, A., FLINT, T. J., SORENSEN, K. M., ANDERSEN, P. S., NORDENTOFT, M., WERGE, T., PEDERSEN, C. B., HOUGAARD, D. M., MORTENSEN, P. B., MORS, O. & BORGLUM, A. D. 2010. CACNA1C (rs1006737) is associated with schizophrenia. *Mol Psychiatry,* 15**,** 119-21.

OBERMAIR, G. J., SZABO, Z., BOURINET, E. & FLUCHER, B. E. 2004. Differential targeting of the L-type Ca2+ channel alpha 1C (CaV1.2) to synaptic and extrasynaptic compartments in hippocampal neurons. *Eur J Neurosci,* 19**,** 2109-22.

OHNO, S. 1972. So much "junk" DNA in our genome. *Brookhaven Symp Biol,* 23**,** 366-70.

OHTSUKA, N., BADUREK, S., BUSSLINGER, M., BENES, F. M., MINICHIELLO, L. & RUDOLPH, U. 2013. GABAergic neurons regulate

lateral ventricular development via transcription factor Pax5. *Genesis,* 51**,** 234-45.
ORPHANIDES, G., LAGRANGE, T. & REINBERG, D. 1996. The general transcription factors of RNA polymerase II. *Genes Dev,* 10**,** 2657-83.
ORR, H. T. & ZOGHBI, H. Y. 2007. Trinucleotide repeat disorders. *Annu Rev Neurosci,* 30**,** 575-621.
OTTER, M., SCHRANDER-STUMPEL, C. T. & CURFS, L. M. 2010. Triple X syndrome: a review of the literature. *Eur J Hum Genet,* 18**,** 265-71.
OU, X. M., CHEN, K. & SHIH, J. C. 2006. Glucocorticoid and androgen activation of monoamine oxidase A is regulated differently by R1 and Sp1. *J Biol Chem,* 281**,** 21512-25.
PAESCHKE, K., SIMONSSON, T., POSTBERG, J., RHODES, D. & LIPPS, H. J. 2005. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol,* 12**,** 847-54.
PAI, C.-Y., CHOU, S.-L. & HUANG, F. F.-Y. 2007. Assessment of the role of a functional VNTR polymorphism in MAOA gene promoter: a preliminary Study. *FORENSIC SCIENCE JOURNAL,* 6**,** 37-43.
PALM, K., BELLUARDO, N., METSIS, M. & TIMMUSK, T. 1998. Neuronal expression of zinc finger transcription factor REST/NRSF/XBR gene. *J Neurosci,* 18**,** 1280-96.
PAN, T., LI, X., XIE, W., JANKOVIC, J. & LE, W. 2005. Valproic acid-mediated Hsp70 induction and anti-apoptotic neuroprotection in SH-SY5Y cells. *FEBS Lett,* 579**,** 6716-20.
PAULUS, F. M., BEDENBENDER, J., KRACH, S., PYKA, M., KRUG, A., SOMMER, J., METTE, M., NOTHEN, M. M., WITT, S. H., RIETSCHEL, M., KIRCHER, T. & JANSEN, A. 2014. Association of rs1006737 in CACNA1C with alterations in prefrontal activation and fronto-hippocampal connectivity. *Hum Brain Mapp,* 35**,** 1190-200.
PAYSEUR, B. A., JING, P. & HAASL, R. J. 2011. A genomic portrait of human microsatellite variation. *Mol Biol Evol,* 28**,** 303-12.
PEDROSA, E., YE, K., NOLAN, K. A., MORRELL, L., OKUN, J. M., PERSKY, A. D., SAITO, T. & LACHMAN, H. M. 2007. Positive association of schizophrenia to JARID2 gene. *Am J Med Genet B Neuropsychiatr Genet,* 144B**,** 45-51.
PHIEL, C. J., ZHANG, F., HUANG, E. Y., GUENTHER, M. G., LAZAR, M. A. & KLEIN, P. S. 2001. Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen. *J Biol Chem,* 276**,** 36734-41.
PHILIBERT, R. A., WERNETT, P., PLUME, J., PACKER, H., BRODY, G. H. & BEACH, S. R. 2011. Gene environment interactions with a novel variable Monoamine Oxidase A transcriptional enhancer are associated with antisocial personality disorder. *Biol Psychol,* 87**,** 366-71.
PHILLIPS, T. 2008. The role of methylation in gene expression. 1**,** 116.
PICKLES, A., HILL, J., BREEN, G., QUINN, J., ABBOTT, K., JONES, H. & SHARP, H. 2013. 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. *J Child Psychol Psychiatry,* 54**,** 1308-17.

PLATH, K., MLYNARCZYK-EVANS, S., NUSINOW, D. A. & PANNING, B. 2002. Xist RNA and the mechanism of X chromosome inactivation. *Annu Rev Genet,* 36**,** 233-78.

PURCELL, S. M., MORAN, J. L., FROMER, M., RUDERFER, D., SOLOVIEFF, N., ROUSSOS, P., O'DUSHLAINE, C., CHAMBERT, K., BERGEN, S. E., KAHLER, A., DUNCAN, L., STAHL, E., GENOVESE, G., FERNANDEZ, E., COLLINS, M. O., KOMIYAMA, N. H., CHOUDHARY, J. S., MAGNUSSON, P. K., BANKS, E., SHAKIR, K., GARIMELLA, K., FENNELL, T., DEPRISTO, M., GRANT, S. G., HAGGARTY, S. J., GABRIEL, S., SCOLNICK, E. M., LANDER, E. S., HULTMAN, C. M., SULLIVAN, P. F., MCCARROLL, S. A. & SKLAR, P. 2014. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature,* 506**,** 185-90.

QIU, H. T., MENG, H. Q., SONG, C., XIU, M. H., CHEN DA, C., ZHU, F. Y., WU, G. Y., KOSTEN, T. A., KOSTEN, T. R. & ZHANG, X. Y. 2009. Association between monoamine oxidase (MAO)-A gene variants and schizophrenia in a Chinese population. *Brain Res,* 1287**,** 67-73.

QUINN, J. P., WARBURTON, A., MYERS, P., SAVAGE, A. L. & BUBB, V. J. 2013. Polymorphic variation as a driver of differential neuropeptide gene expression. *Neuropeptides,* 47**,** 395-400.

RAIBER, E. A., KRANASTER, R., LAM, E., NIKAN, M. & BALASUBRAMANIAN, S. 2012. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res,* 40**,** 1499-508.

RAJ, A. & VAN OUDENAARDEN, A. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell,* 135**,** 216-26.

REIF, A., RICHTER, J., STRAUBE, B., HOFLER, M., LUEKEN, U., GLOSTER, A. T., WEBER, H., DOMSCHKE, K., FEHM, L., STROHLE, A., JANSEN, A., GERLACH, A., PYKA, M., REINHARDT, I., KONRAD, C., WITTMANN, A., PFLEIDERER, B., ALPERS, G. W., PAULI, P., LANG, T., AROLT, V., WITTCHEN, H. U., HAMM, A., KIRCHER, T. & DECKERT, J. 2014. MAOA and mechanisms of panic disorder revisited: from bench to molecular psychotherapy. *Mol Psychiatry,* 19**,** 122-8.

REIF, A., WEBER, H., DOMSCHKE, K., KLAUKE, B., BAUMANN, C., JACOB, C. P., STROHLE, A., GERLACH, A. L., ALPERS, G. W., PAULI, P., HAMM, A., KIRCHER, T., AROLT, V., WITTCHEN, H. U., BINDER, E. B., ERHARDT, A. & DECKERT, J. 2012. Meta-analysis argues for a female-specific role of MAOA-uVNTR in panic disorder in four European populations. *Am J Med Genet B Neuropsychiatr Genet,* 159B**,** 786-93.

REITZ, C. 2015a. Genetic diagnosis and prognosis of Alzheimer's disease: challenges and opportunities. *Expert Rev Mol Diagn,* 15**,** 339-48.

REITZ, C. 2015b. Novel susceptibility loci for Alzheimer's disease. *Future Neurol,* 10**,** 547-558.

RENTON, A. E., MAJOUNIE, E., WAITE, A., SIMON-SANCHEZ, J., ROLLINSON, S., GIBBS, J. R., SCHYMICK, J. C., LAAKSOVIRTA, H., VAN SWIETEN, J. C., MYLLYKANGAS, L., KALIMO, H., PAETAU, A., ABRAMZON, Y., REMES, A. M., KAGANOVICH, A., SCHOLZ, S. W., DUCKWORTH, J., DING, J., HARMER, D. W., HERNANDEZ, D. G., JOHNSON, J. O., MOK, K., RYTEN, M., TRABZUNI, D., GUERREIRO, R. J., ORRELL, R. W., NEAL, J., MURRAY, A., PEARSON, J., JANSEN, I. E., SONDERVAN, D., SEELAAR, H., BLAKE, D., YOUNG, K.,

HALLIWELL, N., CALLISTER, J. B., TOULSON, G., RICHARDSON, A., GERHARD, A., SNOWDEN, J., MANN, D., NEARY, D., NALLS, M. A., PEURALINNA, T., JANSSON, L., ISOVIITA, V. M., KAIVORINNE, A. L., HOLTTA-VUORI, M., IKONEN, E., SULKAVA, R., BENATAR, M., WUU, J., CHIO, A., RESTAGNO, G., BORGHERO, G., SABATELLI, M., CONSORTIUM, I., HECKERMAN, D., ROGAEVA, E., ZINMAN, L., ROTHSTEIN, J. D., SENDTNER, M., DREPPER, C., EICHLER, E. E., ALKAN, C., ABDULLAEV, Z., PACK, S. D., DUTRA, A., PAK, E., HARDY, J., SINGLETON, A., WILLIAMS, N. M., HEUTINK, P., PICKERING-BROWN, S., MORRIS, H. R., TIENARI, P. J. & TRAYNOR, B. J. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron,* 72**,** 257-68.

RIFFO-CAMPOS, A. L., CASTILLO, J., TUR, G., GONZALEZ-FIGUEROA, P., GEORGIEVA, E. I., RODRIGUEZ, J. L., LOPEZ-RODAS, G., RODRIGO, M. I. & FRANCO, L. 2015. Nucleosome-specific, time-dependent changes in histone modifications during activation of the early growth response 1 (Egr1) gene. *J Biol Chem,* 290**,** 197-208.

RIGGS, A. D. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet,* 14**,** 9-25.

RIGGS, A. D. 2002. X chromosome inactivation, differentiation, and DNA methylation revisited, with a tribute to Susumu Ohno. *Cytogenet Genome Res,* 99**,** 17-24.

ROBERTS, J., SCOTT, A. C., HOWARD, M. R., BREEN, G., BUBB, V. J., KLENOVA, E. & QUINN, J. P. 2007. 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. *J Neurosci,* 27**,** 2793-801.

RODRIGUEZ, S., GAUNT, T. R. & DAY, I. N. 2009. Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. *Am J Epidemiol,* 169**,** 505-14.

ROGOZIN, I. B., KOCHETOV, A. V., KONDRASHOV, F. A., KOONIN, E. V. & MILANESI, L. 2001. Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. *Bioinformatics,* 17**,** 890-900.

RONAN, J. L., WU, W. & CRABTREE, G. R. 2013. From neural development to cognition: unexpected roles for chromatin. *Nat Rev Genet,* 14**,** 347-59.

ROOPRA, A., SHARLING, L., WOOD, I. C., BRIGGS, T., BACHFISCHER, U., PAQUETTE, A. J. & BUCKLEY, N. J. 2000. Transcriptional repression by neuron-restrictive silencer factor is mediated via the Sin3-histone deacetylase complex. *Mol Cell Biol,* 20**,** 2147-57.

ROSENBERG, S. S. & SPITZER, N. C. 2011. Calcium signaling in neuronal development. *Cold Spring Harb Perspect Biol,* 3**,** a004259.

SABOL, S. Z., HU, S. & HAMER, D. 1998. A functional polymorphism in the monoamine oxidase A gene promoter. *Hum Genet,* 103**,** 273-9.

SAINSBURY, S., BERNECKY, C. & CRAMER, P. 2015. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol,* 16**,** 129-43.

SAMOCHOWIEC, J., HAJDUK, A., SAMOCHOWIEC, A., HORODNICKI, J., STEPIEN, G., GRZYWACZ, A. & KUCHARSKA-MAZUR, J. 2004. Association studies of MAO-A, COMT, and 5-HTT genes polymorphisms in

patients with anxiety disorders of the phobic spectrum. *Psychiatry Res,* 128**,** 21-6.

SANDOVAL, P. C., SLENTZ, D. H., PISITKUN, T., SAEED, F., HOFFERT, J. D. & KNEPPER, M. A. 2013. Proteome-wide measurement of protein half-lives and translation rates in vasopressin-sensitive collecting duct cells. *J Am Soc Nephrol,* 24**,** 1793-805.

SAWAN, C., VAISSIERE, T., MURR, R. & HERCEG, Z. 2008. Epigenetic drivers and genetic passengers on the road to cancer. *Mutat Res,* 642**,** 1-13.

SCHIZOPHRENIA PSYCHIATRIC GENOME-WIDE ASSOCIATION STUDY, C. 2011. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet,* 43**,** 969-76.

SCHOENHERR, C. J. & ANDERSON, D. J. 1995. Silencing is golden: negative regulation in the control of neuronal gene transcription. *Curr Opin Neurobiol,* 5**,** 566-71.

SCHULZE, T. G., MULLER, D. J., KRAUSS, H., SCHERK, H., OHLRAUN, S., SYAGAILO, Y. V., WINDEMUTH, C., NEIDT, H., GRASSLE, M., PAPASSOTIROPOULOS, A., HEUN, R., NOTHEN, M. M., MAIER, W., LESCH, K. P. & RIETSCHEL, M. 2000. Association between a functional polymorphism in the monoamine oxidase A gene promoter and major depressive disorder. *Am J Med Genet,* 96**,** 801-3.

SEN, S., BURMEISTER, M. & GHOSH, D. 2004. Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *Am J Med Genet B Neuropsychiatr Genet,* 127B**,** 85-9.

SHAM, P. C. & CURTIS, D. 1995. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Hum Genet,* 59**,** 97-105.

SKLAR, P., SMOLLER, J. W., FAN, J., FERREIRA, M. A., PERLIS, R. H., CHAMBERT, K., NIMGAONKAR, V. L., MCQUEEN, M. B., FARAONE, S. V., KIRBY, A., DE BAKKER, P. I., OGDIE, M. N., THASE, M. E., SACHS, G. S., TODD-BROWN, K., GABRIEL, S. B., SOUGNEZ, C., GATES, C., BLUMENSTIEL, B., DEFELICE, M., ARDLIE, K. G., FRANKLIN, J., MUIR, W. J., MCGHEE, K. A., MACINTYRE, D. J., MCLEAN, A., VANBECK, M., MCQUILLIN, A., BASS, N. J., ROBINSON, M., LAWRENCE, J., ANJORIN, A., CURTIS, D., SCOLNICK, E. M., DALY, M. J., BLACKWOOD, D. H., GURLING, H. M. & PURCELL, S. M. 2008. Whole-genome association study of bipolar disorder. *Mol Psychiatry,* 13**,** 558-69.

SMITH, C. L. & PETERSON, C. L. 2005. ATP-dependent chromatin remodeling. *Curr Top Dev Biol,* 65**,** 115-48.

SPERLING, R. A., AISEN, P. S., BECKETT, L. A., BENNETT, D. A., CRAFT, S., FAGAN, A. M., IWATSUBO, T., JACK, C. R., JR., KAYE, J., MONTINE, T. J., PARK, D. C., REIMAN, E. M., ROWE, C. C., SIEMERS, E., STERN, Y., YAFFE, K., CARRILLO, M. C., THIES, B., MORRISON-BOGORAD, M., WAGSTER, M. V. & PHELPS, C. H. 2011. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement,* 7**,** 280-92.

STEVENS, H. C., CHAM, K. S., HUGHES, D. J., SUN, R., SAMPLE, J. T., BUBB, V. J., STEWART, J. P. & QUINN, J. P. 2012. CTCF and Sp1 interact with

the Murine gammaherpesvirus 68 internal repeat elements. *Virus Genes,* 45**,** 265-73.

STRIESSNIG, J., PINGGERA, A., KAUR, G., BOCK, G. & TULUC, P. 2014. L-type Ca2+ channels in heart and brain. *Wiley Interdiscip Rev Membr Transp Signal,* 3**,** 15-38.

SUN, D. & HURLEY, L. H. 2009. The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J Med Chem,* 52**,** 2863-74.

SZUTORISZ, H., DILLON, N. & TORA, L. 2005. The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem Sci,* 30**,** 593-9.

SZYF, M., PAKNESHAN, P. & RABBANI, S. A. 2004. DNA methylation and breast cancer. *Biochem Pharmacol,* 68**,** 1187-97.

TAKAI, D. & JONES, P. A. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A,* 99**,** 3740-5.

TAMMINGA, C. A. & ZUKIN, R. S. 2015. Schizophrenia: Evidence implicating hippocampal GluN2B protein and REST epigenetics in psychosis pathophysiology. *Neuroscience,* 309**,** 233-42.

TAPIA-RAMIREZ, J., EGGEN, B. J., PERAL-RUBIO, M. J., TOLEDO-ARAL, J. J. & MANDEL, G. 1997. A single zinc finger motif in the silencing factor REST represses the neural-specific type II sodium channel promoter. *Proc Natl Acad Sci U S A,* 94**,** 1177-82.

TIKKANEN, R., DUCCI, F., GOLDMAN, D., HOLI, M., LINDBERG, N., TIIHONEN, J. & VIRKKUNEN, M. 2010. MAOA alters the effects of heavy drinking and childhood physical abuse on risk for severe impulsive acts of violence among alcoholic violent offenders. *Alcohol Clin Exp Res,* 34**,** 853-60.

TOTHOVA, Z., KOLLIPARA, R., HUNTLY, B. J., LEE, B. H., CASTRILLON, D. H., CULLEN, D. E., MCDOWELL, E. P., LAZO-KALLANIAN, S., WILLIAMS, I. R., SEARS, C., ARMSTRONG, S. A., PASSEGUE, E., DEPINHO, R. A. & GILLILAND, D. G. 2007. FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell,* 128**,** 325-39.

TREANGEN, T. J. & SALZBERG, S. L. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet,* 13**,** 36-46.

VAN DER VELDEN, A. W. & THOMAS, A. A. 1999. The role of the 5' untranslated region of an mRNA in translation regulation during development. *Int J Biochem Cell Biol,* 31**,** 87-106.

VASILIOU, S. A., ALI, F. R., HADDLEY, K., CARDOSO, M. C., BUBB, V. J. & QUINN, J. P. 2012. The SLC6A4 VNTR genotype determines transcription factor binding and epigenetic variation of this gene in response to cocaine in vitro. *Addict Biol,* 17**,** 156-70.

VERMA, D., CHAKRABORTI, B., KARMAKAR, A., BANDYOPADHYAY, T., SINGH, A. S., SINHA, S., CHATTERJEE, A., GHOSH, S., MOHANAKUMAR, K. P., MUKHOPADHYAY, K. & RAJAMMA, U. 2014. Sexual dimorphic effect in the genetic association of monoamine oxidase A (MAOA) markers with autism spectrum disorder. *Prog Neuropsychopharmacol Biol Psychiatry,* 50**,** 11-20.

VIRE, E., BRENNER, C., DEPLUS, R., BLANCHON, L., FRAGA, M., DIDELOT, C., MOREY, L., VAN EYNDE, A., BERNARD, D., VANDERWINDEN, J. M., BOLLEN, M., ESTELLER, M., DI CROCE, L., DE LAUNOIT, Y. & FUKS, F. 2006. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature,* 439**,** 871-4.

WARBURTON, A., BREEN, G., RUJESCU, D., BUBB, V. J. & QUINN, J. P. 2015. Characterization of a REST-Regulated Internal Promoter in the Schizophrenia Genome-Wide Associated Gene MIR137. *Schizophr Bull,* 41**,** 698-707.

WARBY, S. C., GRAHAM, R. K. & HAYDEN, M. R. 1993. Huntington Disease. *In:* PAGON, R. A., ADAM, M. P., ARDINGER, H. H., WALLACE, S. E., AMEMIYA, A., BEAN, L. J. H., BIRD, T. D., FONG, C. T., MEFFORD, H. C., SMITH, R. J. H. & STEPHENS, K. (eds.) *GeneReviews(R).* Seattle (WA).

WARGELIUS, H. L., MALMBERG, K., LARSSON, J. O. & ORELAND, L. 2012. Associations of MAOA-VNTR or 5HTT-LPR alleles with attention-deficit hyperactivity disorder symptoms are moderated by platelet monoamine oxidase B activity. *Psychiatr Genet,* 22**,** 42-5.

WEBB, B. T., SULLIVAN, P. F., SKELLY, T. & VAN DEN OORD, E. J. 2008. Model-based gene selection shows engrailed 1 is associated with antipsychotic response. *Pharmacogenet Genomics,* 18**,** 751-9.

WEICK, J. P., GROTH, R. D., ISAKSEN, A. L. & MERMELSTEIN, P. G. 2003. Interactions with PDZ proteins are required for L-type calcium channels to activate cAMP response element-binding protein-dependent gene expression. *J Neurosci,* 23**,** 3446-56.

WHEELER, D. G., BARRETT, C. F., GROTH, R. D., SAFA, P. & TSIEN, R. W. 2008. CaMKII locally encodes L-type channel activity to signal to nuclear CREB in excitation-transcription coupling. *J Cell Biol,* 183**,** 849-63.

WHITE, J. A., MCKINNEY, B. C., JOHN, M. C., POWERS, P. A., KAMP, T. J. & MURPHY, G. G. 2008. Conditional forebrain deletion of the L-type calcium channel Ca V 1.2 disrupts remote spatial memories in mice. *Learn Mem,* 15**,** 1-5.

WITCHER, M. & EMERSON, B. M. 2009. Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. *Mol Cell,* 34**,** 271-84.

WONG, C. C., CASPI, A., WILLIAMS, B., CRAIG, I. W., HOUTS, R., AMBLER, A., MOFFITT, T. E. & MILL, J. 2010. A longitudinal study of epigenetic variation in twins. *Epigenetics,* 5**,** 516-26.

WURST, W., AUERBACH, A. B. & JOYNER, A. L. 1994. Multiple developmental defects in Engrailed-1 mutant mice: an early mid-hindbrain deletion and patterning defects in forelimbs and sternum. *Development,* 120**,** 2065-75.

YU, Y. W., TSAI, S. J., HONG, C. J., CHEN, T. J., CHEN, M. C. & YANG, C. W. 2005. Association study of a monoamine oxidase a gene promoter polymorphism with major depressive disorder and antidepressant response. *Neuropsychopharmacology,* 30**,** 1719-23.

ZHANG, J., CHEN, Y., ZHANG, K., YANG, H., SUN, Y., FANG, Y., SHEN, Y. & XU, Q. 2010. A cis-Phase Interaction Study of Genetic Variants Within the MAOA Gene in Major Depressive Disorder. *Biological Psychiatry,* 68**,** 795-800.

ZHANG, M. M., XIAO, C., YU, K. & RUAN, D. Y. 2003. Effects of sodium valproate on synaptic plasticity in the CA1 region of rat hippocampus. *Food Chem Toxicol,* 41**,** 1617-23.

ZHU, Q. S., GRIMSBY, J., CHEN, K. & SHIH, J. C. 1992. Promoter organization and activity of human monoamine oxidase (MAO) A and B genes. *J Neurosci,* 12**,** 4437-46.