Does G-quadruplex DNA have a functional role in regulation of the transcriptome?

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

September 2014

Acknowledgements

I would like to thank:

My UK supervisor Professor John Quinn for his initial encouragement and continued support, tolerance, laughter and understanding in matters of science and otherwise, without this I would not have a thesis to submit.

Quinn Lab members Dr. Kate Haddley, Dr. Abigail Savage, Miss Alix Warburton, Mrs Veridiana Pessoa, Dr Jill Bubb, Mr Maurizio Manca and Miss Olympia Gianfrancessco for creating a lively, fun, friendly and stimulating environment in which to work.

My NIH supervisor Dr David Levens for sharing insight and knowledge, challenging my abilities and developing me as a scientist.

Levens Lab members Fedour Kouzine, Suzanne Sanford, Eric Batchelor, Brian Fisher, Louisa Ho and Nie Z, for welcoming me into their lab. A special thanks to my friends; Weixin Zhou for sharing ice cream and bubble tea, Ashutosh Gupta for sharing fantastic food and conversation and Laura Baranello for sharing coffee, fun drinks and generally looking after me.

Nicola Sater and Otto for friendship advice and support during our morning walk in dogwood park.

My friend Kevin Garnet for coffee and conversation to keep my mind sharp.

An extra special thank you to my wife Elizabeth Myers for her belief, love, support and occasional harsh words of encouragement.

I would like to thank the Wellcome Trust & National Institutes of Health NCI for funding my Ph.D.

I would like to dedicate my thesis to my beautiful daughter Bella Marjory Myers.

Abstract

All DNA transactions involving overtwisting or undertwisting of the double helix alter the topological state of DNA, yielding distortions known as DNA supercoiling or DNA torsional stress. Under conditions of negative supercoiling susceptible sequences within B-form DNA may be transformed to stable, non-B-DNA structures such as G-quadruplex. G-quadruplex DNA has the potential to create or destroy transcription factor binding sites or influence nucleosome positioning and so may contribute to gene regulation, making it an attractive therapeutic target. Although *in vitro* and in silico studies support the notion that G-quadruplex is regulatory, a compelling *in vivo* demonstration of G-quadruplex function is lacking.

Using computational and experimental approaches previously developed to identify DNA segments that actually adopt non-B DNA conformations *in vivo*, I have identified sequences with a high probability of secondary structure formation. Identified sequences are located in the promoters of three genes central to cancer biology where much of the work in G-Quadruplex has been focused. I have expanded this focus into the field of neuroscience by identifying a further three sequences in the promoters of genes implicated in a range of neurological disorders. Preliminary biochemical analysis of these regions assessed methylation status and transcription factor binding profiles corresponding to a dynamic model of G-quadruplex formation at the well characterised CT element in the c-MYC promoter.

In order to provide definitive structure function analysis of G-quadruplex I designed a gene editing strategy to reengineer the well characterized CT element of c-MYC. The initial strategy used AAV- targeting vectors to introduce strategic mutations that destabilize G-quadruplex DNA conformations in the CT element of c-MYC in the HCT-116 cell line via homologous recombination. In parallel to these structure-destabilizing mutations, neutral base changes were targeted to the 5´-UTRs of the targeted alleles, distinguishing them from their respective unmodified homologous partners. qPCR and ChIP analysis of the relative expression and factor binding profiles of targeted and wild-type alleles will enable a definitive structure function analysis of basal or induced expression; determining whether these secondary structures have a functional role in regulating gene expression and the factors involved.

Low targeting efficiencies of the AAV gene editing strategy prompted me combine the recent technology of the CRISPR-Cas system with traditional AAV targeting to produce a highly efficient, high fidelity and streamlined approach to gene editing.

Contents

Chapter 3: Repeat sequences in the promoter of selected genes have the potential to form the secondary DNA Structure G-quadruplex…….102-142

Chapter 5: CRISPR & AAV, an efficient approach to gene editing.178-205

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

List of Tables

List of Abbreviations

A	Adenine
AAV	Adeno-associated virus
ATP	Adenosine-5'-triphosphate
BLAT	BLAST-like alignment tool
bp	Base pairs
C	Cytosine
cDNA	complementary deoxyribonucleic acid
ChIP	Chromatin immunoprecipitation
CNS	Central nervous system
CpG	CG dinucleotides
CTCF	CCCTC binding-protein
DAT1	Dopamine transporter (SLC6A3)
dATP	deoxyadenosine trisphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dI/dC	deoxyinosinic-deoxycytidylic acid
dTTP	Deoxythymidine triphosphate
DHSs	DNase-hypersensitive sites
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediamine tetraacetic acid
EGTA	Ethylene glycol tetraacetic acid,
FBS	Fetal bovine serum
G	Guanine
×g	times gravity
GTFs	General transcription factors
H3K4m2	di-methylation of lysine 4 in histone H3
H3K9m2	di-methylation of lysine 9 in histone H3
H3K36m3	tri-methylation of lysine 36 in histone H3
HDAC	Histone deacetylase complex
HeLa	Henrietta Lacks clonal cell line
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hnRNA	Heterogeneous nuclear ribonucleic acid
JAr	human placental choriocarcinoma
kb	Kilobases
KCl	Potassium chloride
LiCl	Lithium chloride
l-allele	Long variants of the 5’ promoter VNTR of the 5-HTT gene
LB	Luria-bertani broth
LPR	Linked polymorphic region
MeCPs	Methyl-CpG-binding proteins
µg	Micro-grams
µM	Micro-molar
µl	Micro-litre
ml	Milli-litres

mM	mill-molar
mRNA	Messenger ribonucleic acid
NCBI	National center for biotechnology information
ng	nano-grams
nm	nanometer
ORF	Open reading frame
OCD	Obsessive compulsive disorder
p300	E1A binding protein p300
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
Pmol	Pico-molar
Pol II	polymerase II
PFC	Prefrontal cortex
PIC	Preinitiation complex
qPCR	Quantitative PCR
RNA	Ribonucleic acid
RPM	Revolutions per minute
RPMI	Roswell Park Memorial Institute
RT-PCR	Reverse transcriptase PCR
s-allele	short variants of the 5’ promoter VNTR of the 5-HTT gene
SDS	Sodium dodecyl sulphate
SeqChIP	Sequential chromatin immunoprecipitation
SERT	Serotonin transporter
SLC6A4	Serotonin transporter
SLC	Solute carrier gene family
SNP	Single nucleotide polymorphisms
Stin2	VNTR in Intron 2 of 5-HTT
SPECT	Sngle photon emission computed tomography
SSRI	selective serotonin reuptake inhibitors
SV40	Simian virus 40
TBE	Tris/Borate/EDTA
TBP	TATA binding protein
TCA	Tricyclic antidepressants
TF	Transcription factor
TFBS	Transcription factor binding sites
TRD	Transcription repression domain
UCSC	University of California Santa Cruz
UTRs	Untranslated regions
UV	Ultraviolet irradiation
V	voltages
VNTR	Variable number of tandem repeat
5HT	Serotonin
5HTT	Serotonin transporter
ZFs	Zinc fingers

Chapter 1

General Introduction

1.1 Gene regulation

1.1.1 Transcription

Transcription is the key step in regulation of gene expression; a fundamentally stochastic process with randomness in transcription and translation leading to cell-cell variations in mRNA and protein levels (Raj and van Oudenaarden, 2008).

Eukaryotic gene expression is driven by a series of complex and dynamic events. The process is initiated by the assembly of a set of protein complexes TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH generally referred to as the general transcription factors (GTFs) (Orphanides et al., 1996). These complexes assemble into a ~2 MDa complex which is then anchored to the core promoter sequence to form the preinitiation complex (PIC) (Smale and Kadonaga, 2003, Kornberg, 2005). The PIC is then recognised by RNA polymerase II (Pol II). Three distinct classes of DNA sequence elements direct RNA Pol II. The first includes DNA sequences within the core promoter generally a TATA-box (consensus TATAAA); the second is a proximal-promoter element *cis*-regulatory element and the third and distal-promoter element termed enhancer. Enhancers are positioned at distant sites in chromatin and modify the rate of initiation complex formation by mechanisms that are still poorly understood (Roeder, 2005). PIC formation is followed by promoter melting and formation of an "open" initiation complex (Figure 1.1). Once RNA Pol II has recognised the PIC the synthesis of RNA is initiated although recent studies estimate only 1 in 90 polymerases proceed to elongation (Darzacq et al., 2007) suggesting polymerase and transcription complex assembly is an inefficient process (Hager et al., 2009). Promoter escape, pausing, elongation, and transcriptional

termination are all additional steps in eukaryotic gene expression reviewed in (Fuda et al., 2009, Lee and Young, 2000, Orphanides and Reinberg, 2000). It should be noted that although the PIC been implicated in selecting the accurate transcription start site (TSS), promoter melting, and Pol II promoter escape (Matsui et al., 1980, Roeder, 1996, Goodrich et al., 1996) the details as to how this essential complex assembles remain elusive (He et al., 2013).

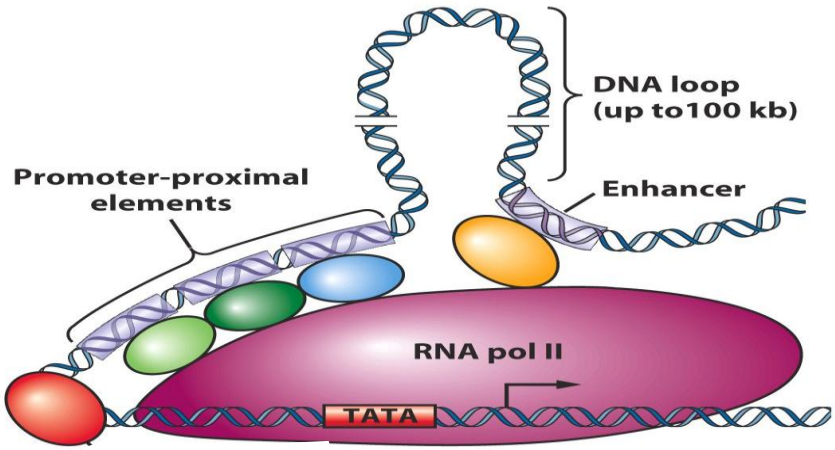


Figure 1.1: Transcription initiation; association of RNA polymerase II and GTFs to core promoter.
A schematic showing the core promoter elements: TATA-box, promoter-proximal elements and promoter-distal elements interacting with RNA polymeraseII and the general transcription factors during transcription (adapted from B. Lewin, Genes IX).

Holoenzyme recruitment to promoter regions and polymerase pausing were believed to be the key regulatory steps in the expression of most protein-coding genes (Margaritis and Holstege, 2008, Ptashne and Gann, 1997, Roeder, 2005, Core and Lis, 2008, Muse et al., 2007, Seila et al., 2009, Zeitlinger et al., 2007). This view was recently extended to include promoter melting as an additional step to globally regulate transcription in eukaryotes based on studies mapping single-stranded DNA (ssDNA) across the genome (Kouzine et al., 2013b).

1.1.2 Transcription Dynamics

Transcription factors can move rapidly through the nucleus by diffusion, in no specific direction and requiring no cellular energy (Misteli, 2001). Diffusion coefficients range from ~0.5 to 5 $\mu m^2 s^{-1}$ depending on the transcription factors shape, size, and the nature of its chromatin interactions (Misteli, 2001, Gorski et al., 2006). This rapid rate of diffusion means a molecule can cover the length of a typical mammalian nucleus in a few seconds and enables a single transcription factor to visit the volume of the nucleus in minutes (Hager et al., 2009). Transcription factor – chromatin binding is highly transient and dynamic; with interactions occurring in the order of seconds. Recent quantitative fluorescence recovery after photobleaching (FRAP) analysis suggests a range of milliseconds to ~ 100 seconds (Sprague et al., 2004, Farla et al., 2004, Gorski et al., 2006, Phair et al., 2004). Data has emerged to suggest that transcription factors may spend most of their time bound to chromatin (Elf et al., 2007, Phair et al., 2004). The estimated time between interactions is on the order of 50 – 250ms (Hager et al., 2009) although the majority of these interactions are non-specific, highly transient and as such not functional. This 3D hopping continues until the molecule finds a specific site, binding for an extended period in order to elicit a functional response. This has been described as the random 3D scanning model (Misteli, 2001), (Figure 1.2). It has been proposed that once bound the transcription factor is then able to perform a local search of the DNA through scanning, hopping or directed motor driven motion (Gorman and Greene, 2008, Gowers et al., 2005, Halford and Marko, 2004).

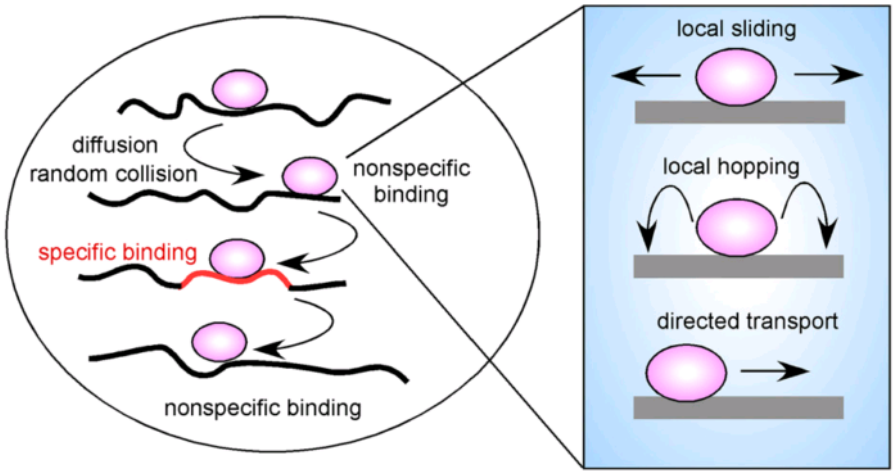


Figure 1.2 Random 3D scanning model of transcription factor binding

Transcription factors diffuse rapidly through the nuclear space and by random collision associate with chromatin. The majority of encounters are at non-target sites and as such are highly transient. When a specific binding site (red) is found a prolonged interaction occurs to elicit a functional response. When the transcription factor binds it might embark upon local motion along the chromatin fibre by sliding along the DNA, hopping or by directed, motor driven motion.(Adapted from Hager et al., 2009)

1.1.3 Chromatin organisation

In order to fit the 2m length of DNA into the 6μm nucleus of eukaryotic cells, DNA must be densely packaged as chromatin. The fundamental unit of chromatin is the nucleosome, in which 147 DNA base pairs are wrapped 1.65 turns around the core histone octamer. Histones are small basic proteins consisting of globular domains and a charged amino-terminus (N-terminus) tail. The core histone octamer consists of two copies of each of the histones H2A, H2B, H3 and H4 (Luger et al., 1997, Kornberg and Thomas, 1974). These strong and repetitive interactions between DNA and histones occur as a result of the negatively charged phosphate backbone of the DNA molecule, and the charged amino-terminus (N-terminus) tail of the histone; one fifth of these amino acids are positively charged (lysine and arginine). More than 14 contact points exist between histones and DNA which make the nucleosome a stable protein-DNA interaction under physiological conditions. Short stretches of naked DNA termed linker regions connect nucleosomes to one another. Linker DNA ranges in length between 20 and 90 bp and varies between different species and even within a single cellular genome (Woodcock et al., 2006, Godde and Widom, 1992). Decondensed chromatin; 11-nm fibre is often referred to as ‘beads on a string’, with each ‘bead’ a nucleosome and the ‘string’ the linker DNA. The binding of a fifth histone creates a more condensed higher-ordered structure. The next level of compaction is termed ’30-nm fibre’ which is organised by the linker histone H1 binding to the linker DNA and a nucleosome on either side. The 30-nm fibre is further condensed forming looped domains termed ‘300-nm fiber’. The final level of compaction is the looped ‘300-nm fibre’ being further folded, giving rise to the characteristic metaphase chromosome (Felsenfeld and Groudine, 2003).

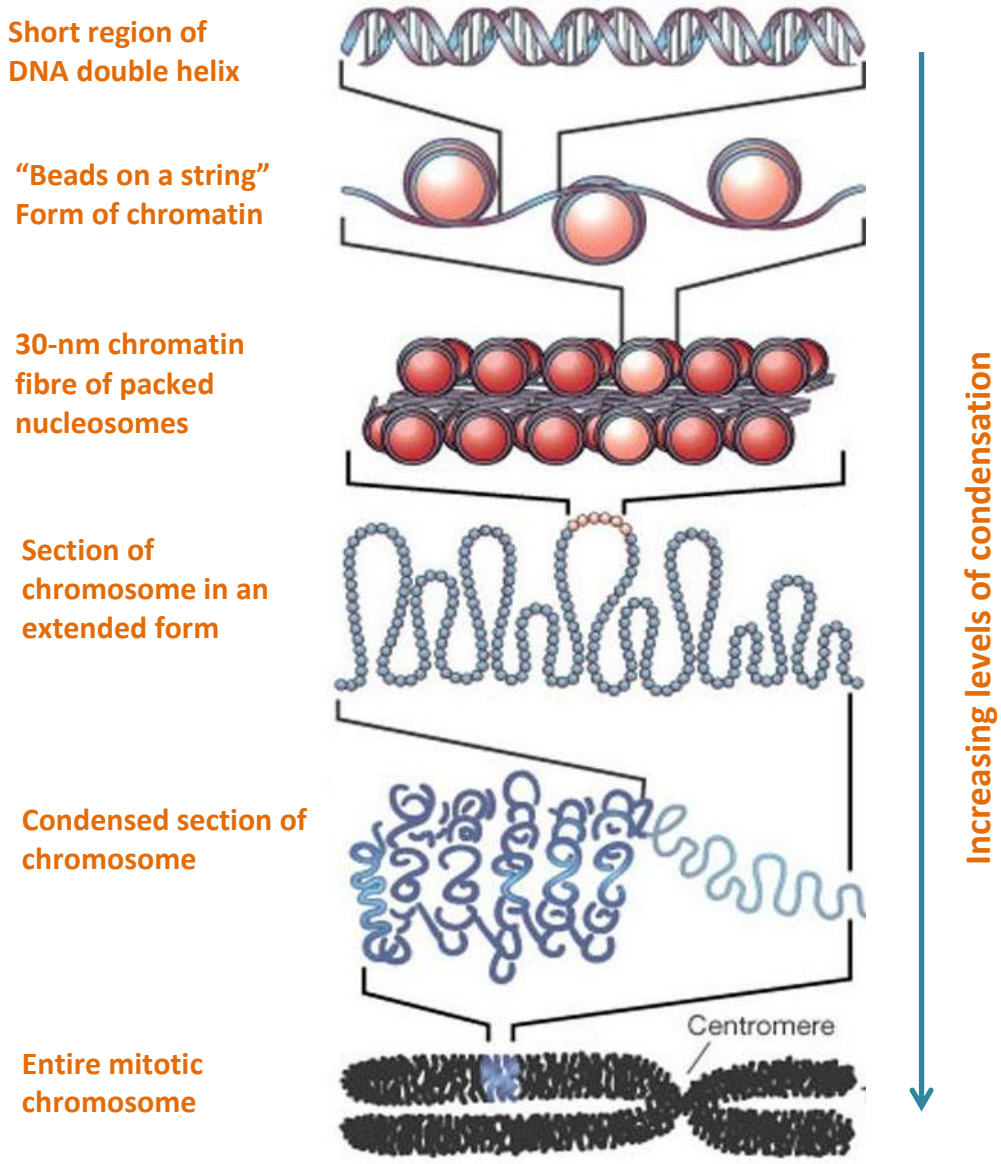


Figure 1.3 Levels of chromatin organisation.
Organization of DNA into the various levels of chromatin structure, a. Duplex DNA, b. 11nm nucleosome, c. 30nm nucleosome, d. 300nm nucleosome e. Section of condensed chromosome, f. mitotic chromosome. (Adapted from Felsenfeld and Groudine, 2003).

1.1.4 Chromatin Remodelling

As transcription factors are traversing the nuclear landscape it is essential they can access the DNA in order to interact with it and orchestrate their regulatory function. Due to the strong binding energy between nucleosomes and DNA, reorganisation of the nucleosome structure must accompany any transcription factor-DNA interaction (Hager et al., 2009). Further, *in-vivo* experiments have provided significant evidence indicating that disruption of nucleosome structure is required for proper elongation; identifying the fact that nucleosome disruption travels along the gene faster than the rate of RNA polymerase II translocation (25 nucleotides per second) (Petesch and Lis, 2008, Izban and Luse, 1992).

The original studies (Wu et al., 1979) describe DNaseI hypersensitive sites (DHS), which represent chromatin regions sensitive to nucleolytic agents specifically DNaseI and micrococcal nuclease (MN), indicating that these regions have disrupted nucleosome structures. Global studies now estimate 2% of the mammalian genome has disrupted nucleosome structure (Hesselberth et al., 2009, Boyle et al., 2008). Transcriptional regulatory elements are strongly associated with these DHS sites (John et al., 2008) therefore chromatin remodelling is implicitly related with binding of transcription factors.

The traditional way to view these nuclear disruptions was in terms of switching between static positions (Becker, 2002, Mellor, 2005, Aoyagi et al., 2002), however studies have since indicated that nucleosome structural remodelling at DHS sites is a continuous and dynamic process. Therefore data generated using static techniques that utilize fixed or broken cell preparations in fact represents alterations in the equilibrium distribution of dynamic states (Fletcher et al., 2002, Nagaich and Hager, 2004, Boeger et al., 2008).

1.1.5 Genetic variability

1.1.5.1 Allelic variation: micro-, mini-satellites, repeats and Variable Number Tandem Repeats (VNTR)

Over 500,000 micro- or mini-satellites have been identified in the human genome. 10–100 nucleotides repeated several times in tandem and bordered by unique DNA sequences are defined as mini-satellites. Micro-satellites also called Short Tandem Repeats (STRs) have smaller sequence repeats e.g., tetra and di-nucleotide sequence repeats (Figure 1.4).

Some degree of degeneration is usually observed in the repeating units; where one repeat may be slightly different to the next but the overall core consensus sequence is maintained. In order to determine whether a repeat unit is variable, genotyping is used; this enables identification of different allelic copy numbers existing between individuals in a population, for example at the same allelic loci, one individual may have 8 copies of a repeat, whilst another may have 10 copies (Warburton et al., 2014).

Variable Number Tandem Repeats (VNTR) are defined as having sufficient DNA sequence in the repeat to act as a sequence specific DNA binding site for proteins. Proteins able to bind these sites include transcription factors suggesting VNTRs can act as transcriptional regulatory domains (Haddley et al., 2008) (Figure 1.5). These domains are similar to the regulatory repeat domains seen in retroviruses or herpesviruses (Quinn et al., 1987, Quinn et al., 1998, Quinn et al., 1989). Most VNTRs with the potential to act as transcriptional regulatory domains are located in non-coding regions of the genome. These repeat regions contain potential cis-regulatory elements or transcription factor binding sites which produce localised hubs. These multiple binding sites could have implications for gene expression

(Soeby et al., 2005, Lovejoy et al., 2003, Klenova et al., 2004). In support of the hypothesis that VNTRs are able to act as modulators of gene expression many are located at higher density in gene enriched areas compared to non-genic regions (Breen et al., 2008). Potential regulatory VNTRs have been identified in the serotonin transporter (5-HTT), dopamine transporter (DAT1), N-methyl-Daspartame receptor 1 (GRIN1/NMDAR), dopamine D4 receptor (DRD4), and monoamine oxidase A (MAOA). VNTRs are predicted to act as endogenous modulators of transcription or to alter post transcriptional properties of the gene such as mRNA stability. Based on published data by the Quinn group, VNTRs can function in both a tissue-specific and stimulus-inducible manner to fine-tune gene expression (Haddley et al., 2012, Vasiliou et al., 2012). This control over gene expression could be mechanistically correlated with normal physiological function, variation between individual, and a number of diseases and disorders (Ali et al., 2010). A further implication is that if stimulus inducible expression is variable based on a specific polymorphism associated with a disorder then that may in turn have implications on the response to specific pharmacological treatment (Fiskerstrand et al., 1999, MacKenzie et al., 2000, MacKenzie and Quinn, 1999, Lovejoy et al., 2003, Klenova et al., 2004, Roberts et al., 2007).

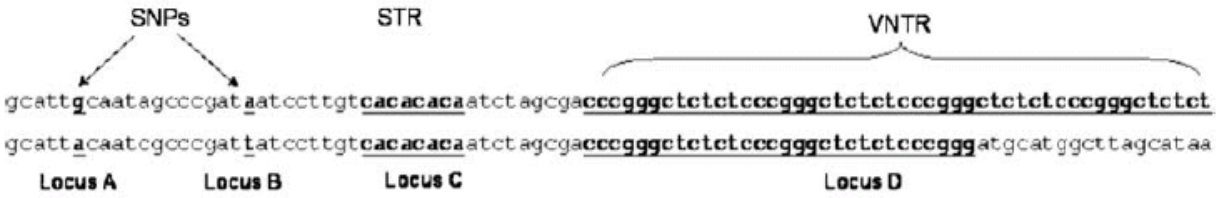


Figure 1.4 Examples of the various polymorphisms

Polymorphisms at locus A and locus B are Single Nucleotide Polymorphisms (SNPs). Locus C is a micro-satellite polymorphism, or Short Tandem Repeat (STR). Locus D is a Variable Number Tandem Repeat (VNTR).

1.2 DNA structure and topology

1.2.1 DNA structure and topology

The dynamics and function of proteins interacting with DNA are indeed central to transcriptional regulation, however evidence is emerging to suggest that DNA itself in terms of its structure and topology may play a role in gene regulation and as such should no longer be thought of solely as a passive information repository (Kouzine et al., 2013a, Lavelle, 2009).

Genetic processes such as transcription and replication apply physical forces to DNA (Figure 1.4), and in turn alter the molecules structure (Baranello et al., 2012). The structural changes that distort the double helix are known as supercoiling or DNA torsional stress and occur as over twisting, under twisting and associated loop-like plectoneme structures (Vologodskii et al., 1992). According to the "twin domain" theory DNA is thread through the transcription machinery, forcing it to revolve around its axis (Osborne et al., 2004, Iborra et al., 2001, Boeger et al., 2005, Cook, 1995); leaving negative supercoils (untwisted) in its wake and driving positive supercoils (over twisted) ahead of the translocating RNA polymerase (Liu and Wang, 1987, Vologodskii et al., 1992). Supercoiling is mediated by topoisomerases which transiently break and re-join the DNA backbone in turn removing positive and negative supercoils (Roca, 2011a). Topoisomerases play an essential role in maintenance of the double helix topology as without them torsional stress would rise to such an extent that all genomic processes would be disrupted (Roca, 2011b, Lavelle, 2009).

1.2.2 The role of DNA supercoiling

DNA supercoiling has a direct influence on nucleosome structure, with positive DNA supercoiling facilitating unwrapping of DNA from the histones, and negative supercoiling promoting nucleosome formation (Freeman and Garrard, 1992). It has therefore been suggested that a burst of transcription which drives positive supercoiling ahead of the translocating RNA polymerase (Liu and Wang, 1987, Vologodskii et al., 1992) could in fact be "clearing the way" for polymerase to elongate through a nucleosome dense landscape (Recouvreux et al., 2011, Zlatanova and Victor, 2009). Negatively supercoiled DNA left behind the transcription machinery would be conducive of rapid nucleosome re-formation in the wake of the elongating complex (Baranello et al., 2012). Taking these data into account, it is likely that by varying the intensity and polarity of supercoiling, chromatin formation could be directly regulated to meet transcriptional demands in real-time. This assertion fits with the modern view of chromatin remodelling being a continuous and dynamic process (Hager et al., 2009).

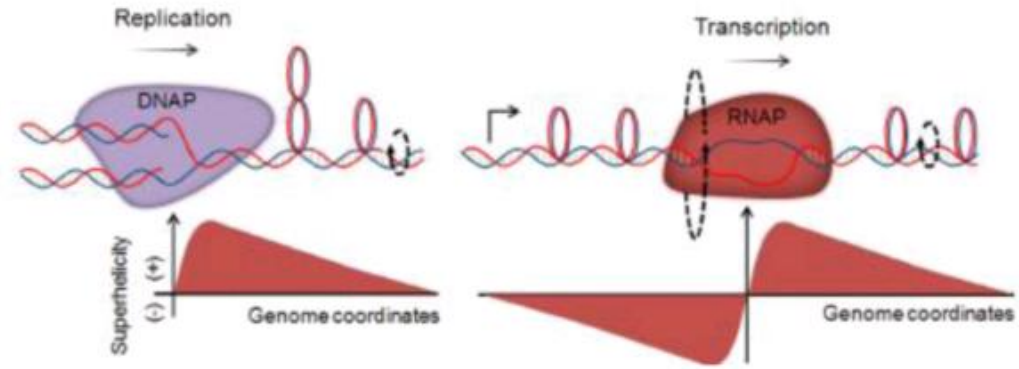


Figure 1.5 Supercoiling generated during transcription and replication.

Polymerases moving without rotation screw the DNA through the protein complexes, causing the template to rotate around its axis (adapted from Baranello et al., 2012).

1.2.3 Alternative DNA structure (Non-B DNA)

In addition to nucleosome arrangement and the covalent modification of histones, alternative DNA conformation provides another source of structural diversity in determining gene activity. Further these alternative DNA structures may contribute to the transmission of epigenetic states (Kouzine and Levens, 2007), through their ability to modify transcription factor binding sites and influence nucleosome structure. Non-B DNA structures include single-stranded DNA (Michelotti et al., 1996b), H-DNA (Htun and Dahlberg, 1989), Z-DNA (left handed double helix) (Rich and Zhang, 2003), G-quadruplex DNA (Simonsson, 2001), triplex DNA (Gilbert and Feigon, 1999), slipped-strand DNA (Pearson and Sinden, 1996), sticky DNA (Son et al., 2006), S(stretched)-DNA (Cluzel et al., 1996) and cruciforms (Shlyakhtenko et al., 1998).

Although the double helix is dynamic and breathes (opening of DNA bubbles) at physiological temperatures, the opening of bubbles is localized and short lived. Hydrophobic and electrostatic base-stacking and base-pairing create thermodynamic and kinetic barriers which prevent large segments of DNA from adopting alternative DNA conformations (Yakovchuk et al., 2006). However these barriers can be overcome and duplex destabilization can occur through high levels of transcription induced negative supercoiling (Kouzine et al., 2013a). This makes it possible for the double helix to adopt a variety of alternative structures, depending on the sequence composition (Kouzine and Levens, 2007). The promoter region of c-MYC is the best characterised region in terms of its ability to form either a single-stranded or a non-B DNA conformation. (Chung and Levens, 2005, Levens et al., 1997, Marcu, 1987, Michelotti et al., 1996b, Kouzine et al., 2008)

1.2.3.1 G-Quadruplex

G-quadruplexes are four-stranded DNA structures, the sequence of DNA that enables these structures to form is (G3+ N1-7 G3+ N1-7 G3+ N1-7 G3). This sequence arrangement creates two or more G-tetrads; four hydrogen-bonded guanines in a planar arrangement (Figure 1.7). Transcription induced negative supercoiling makes it energetically possible for duplex DNA to adopt the G-quadruplex formation, stabilized by monovalent cations, such as K+ and Na+, that intercalate between the G-tetrads and coordinate bonds with the guanine carbonyl groups (Dapic et al., 2003, Sun and Hurley, 2009).

Mounting evidence suggests that G-quadruplex may play an important physiological role. Putative G-quadruplex motifs are highly prevalent in human promoter regions with up to 40% of human gene promoters containing at least one G-quadruplex (Huppert and Balasubramanian, 2005, Huppert and Balasubramanian, 2007, Verma et al., 2008) and analyses of human cells revealed a correlation between the extent of gene expression and predicted G quadruplexes at promoters (Du et al., 2008). Furthermore, at skeletal muscle promoters, formation of G-quadruplexes impacts the recruitment of MyoD during differentiation (Shklover et al., 2010, Yafe et al., 2008). A disproportionatly high incidence of G-quadruplex motifs have been observed in the promoters of oncogenes, in contrast to the promoters of tumour suppressors which exhibit an extremely low G-quadruplex formation potential (Eddy and Maizels, 2006). Of particular interest is the structure-specific antibody engineered by the Balasubramanian group which has allowed them to visualize quantitatively DNA G-quadruplex structures in human cells. They also show that G-quadruplex formation in DNA is modulated during cell cycle progression (Biffi et al., 2013).

1.2.3.2 G-quadruplex DNA as therapeutic targets

A wide range of G-Quadruplex structures are possible (Figure 1.6) this topological diversity is a result of variations in sequence, strand directionality, loop length, and the number of tetrad stacks (Gonzalez and Hurley, 2010a). This diverse spectrum provides an opportunity to development specific molecules that can modulate the formation or stabilize these structures and in turn regulate gene expression.

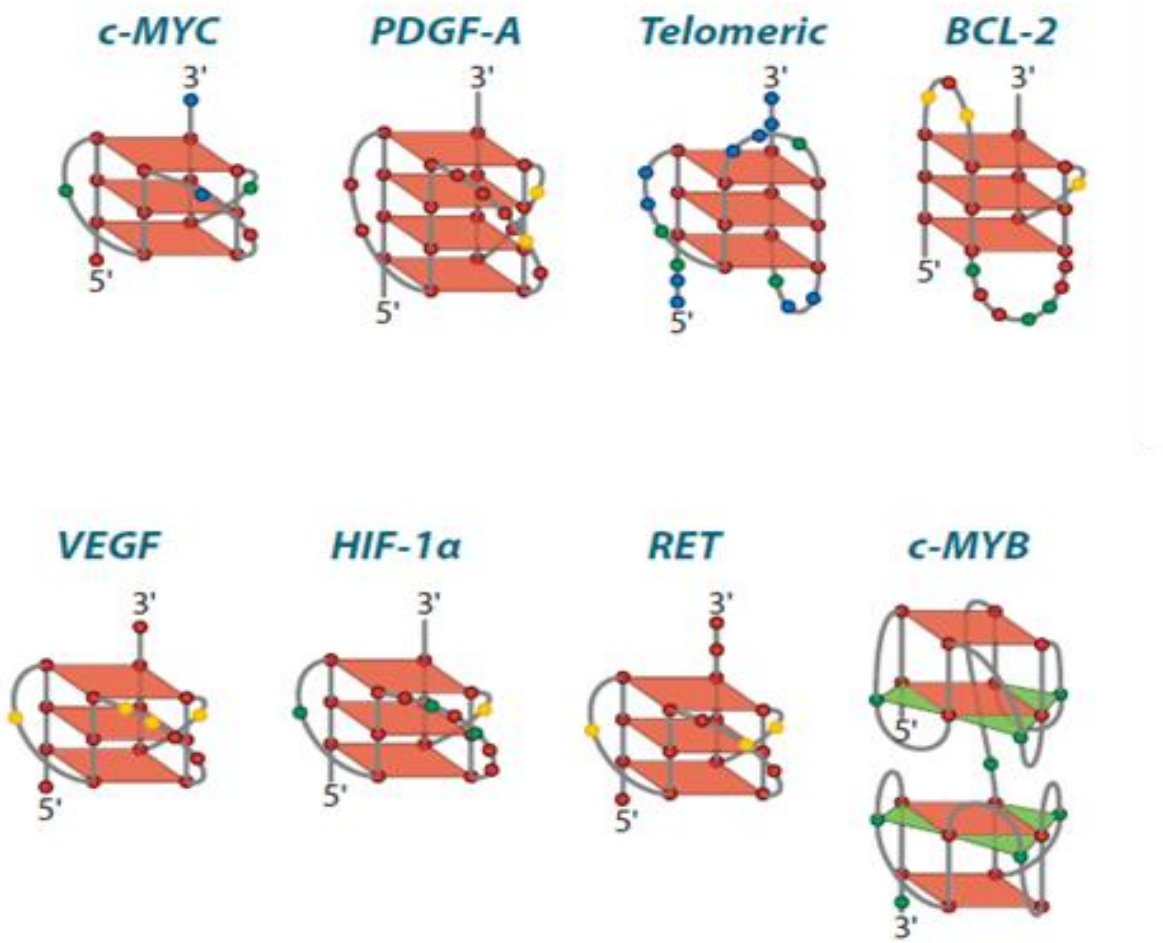


Figure 1.6 Possible of G-quadruplex formations

Models of the conformations of potential G-quadruplexes, determined by varying sequence arrangement, coloured dots refer to specific base pairs: red – G, blue-C yellow – A, green -T the gene in which they are located is detailed above (adapted from (Gonzalez and Hurley, 2010a))

For example the G-quadruplex that forms within the CT element of c-MYC has been demonstrated to function as a repressor element (Siddiqui-Jain et al., 2002). Therefore developing compounds that can stabilize this structure could be used to specifically repress c-MYC expression. c-MYC has been implicated in most aspects of tumour biology with deregulation generally due to gene amplification, translocation, altered ploidy and increased transcription (Marcu et al., 1992, Spencer and Groudine, 1991). Therefore selectively targeting its aberrant expression would be an attractive therapeutic intervention. In fact a number of molecules have been explored as potential G-quadruplex stabilising agents, including TMPyP4, which was shown to decrease c-MYC expression at both the mRNA and protein levels, as well as lower the level of several c-MYC–regulated genes (Grand et al., 2002). More recently the Balasubramanian group have identified pyridine-2,6-bis-quinilodicarboxamide derivative, pyridostatin5 (PDS) as an exciting G-quadruplex stabilising ligand.(Murat et al., 2013, McLuckie et al., 2013, Muller et al., 2012)

Cancer is not the not the only disease where molecules targeting G-quadruplex are applicable; a recent study has identified the G-quadruplex present in the promoter of HIV-1and states that the structure may be critical for HIV-1 gene expression and is potentially a novel target for anti-HIV-1 drug development (Amrane et al., 2014).

1.3 Genes analysed for G-quadruplex DNA

1.3.1 c-MYC.

c-MYC is a transcription factor and 10–15% of genes are regulated by this protein (Fernandez et al., 2003, Li et al., 2003, Zeller et al., 2006). c-MYC has been shown to interact with numerous transcription and chromatin regulators (Cheng et al., 1999, Cowling and Cole, 2006, Eilers and Eisenman, 2008, Rahl et al., 2010, Wasylishen and Penn, 2010) and is implicated in most aspects of tumour biology, including apoptosis, proliferation metastasis, angiogenesis and changes in the tumour microenvironment (Oster et al., 2002, Eisenman, 2001). c-MYC is the key example used by Weinstein to describe the concept of "oncogenic addiction" whereby some cancers appear to rely on one or a few genes to maintain the malignant phenotype, therefore genes such as c-MYC may be the "Achilles heel" in specific cancers (Weinstein et al., 1997, Weinstein, 2002). A seminal paper by the Levens Lab addresses the "addiction" to c-MYC, through a simple rule which explains the majority of c-MYC biology: c-MYC is not an on-off specifier of gene activity, but a nonlinear amplifier of expression. This means it acts universally at active genes except for a class of genes termed 'immediate early' genes that are induced before c-MYC (Nie et al., 2012). c-MYC transcriptional regulation is a complex process involving numerous promoters (P0, P1, P2, P3) and start transcriptional start sites (Gonzalez and Hurley, 2010a), (Figure 1.7).

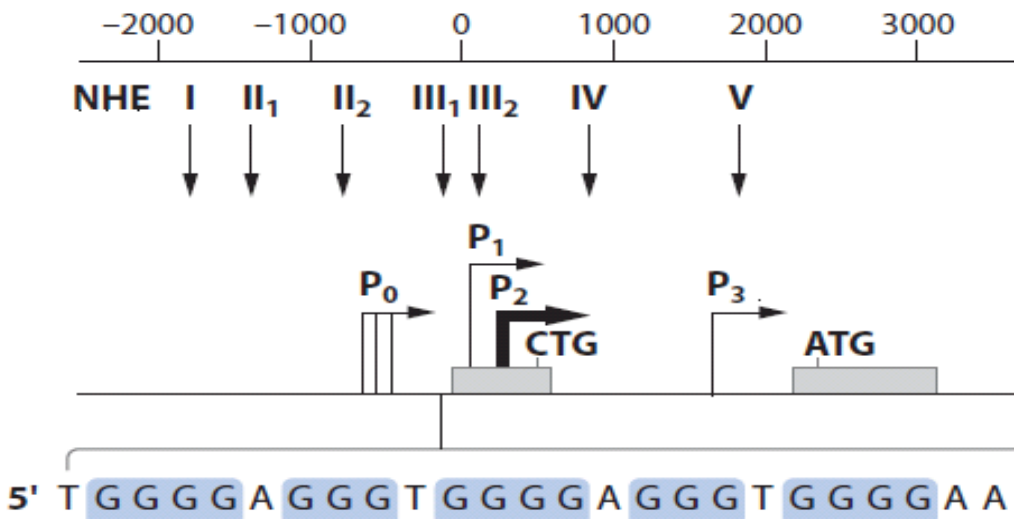


Figure 1.7 c-MYC promoter structure

c-MYC promoter contains 5 NHE domains NHE I – V, with arrows to identify location within the promoter. NHE III_1 is the CT element, expanded in the insert below to show G-rich and C-rich regions of the sequence, (adapted from (Gonzalez and Hurley, 2010a)

The c-MYC promoter contains seven nuclease hypersensitive elements (NHEs) (Figure 1.8), the NHE III_1 region (CT-element), is able to form non–B-DNA structures in supercoiled DNA as well as in its endogenous location (Travers and Muskhelishvili, 2007, Liu and Wang, 1987, Michelotti et al., 1996b, Kohwi and Kohwi-Shigematsu, 1991). The CT element is located –142 to –115 base pairs upstream of the P1 promoter and studies have attributed up to 90% control of the total c-MYC transcription to it (Davis et al., 1989, Berberich and Postel, 1995). The CT element is made up of a cytosine-rich (C-rich) coding strand and a guanine-rich (G-rich) noncoding strand that are able to engage in a slow equilibrium between B-form duplex DNA, single-stranded DNA, and tetra-stranded DNA (Gonzalez and Hurley, 2010a). The guanine rich strand is able to adopt G-quadruplex structure (Figure 1.8) and the complementary cytosine strand is able to adopt an i-motif or i tetraplex.

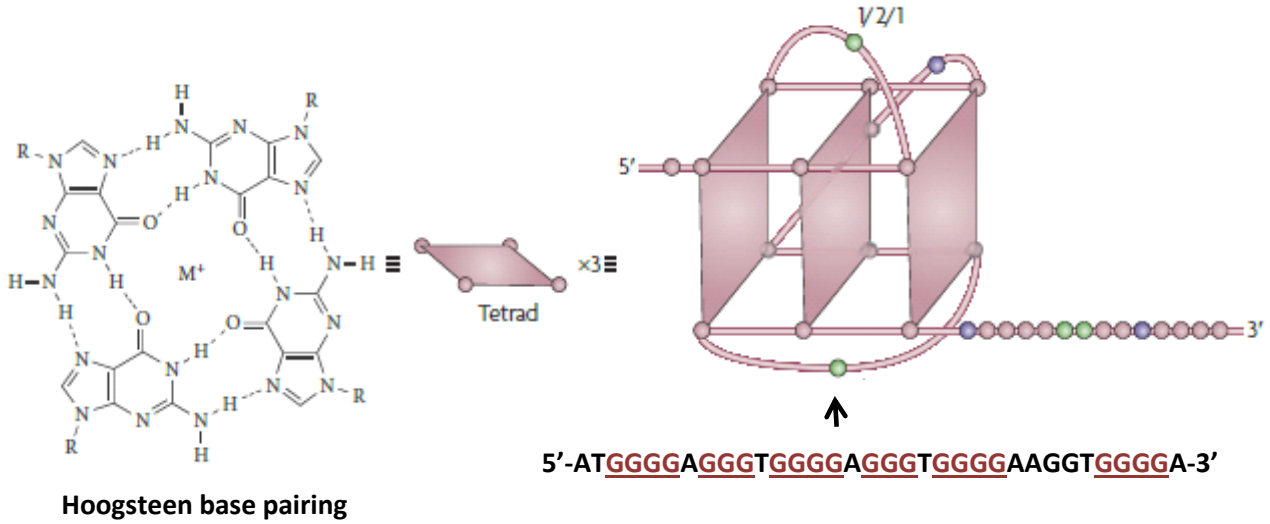


Figure 1.8 Sequence and structures of the G‑quadruplex the CT element

Four guanine residues form a tetrad, and three tetrads can stack to form an intramolecularly linked G-quadruplex with a 1/2/1 looping pattern. (Figure adapted from Brooks and Hurley, 2009).

1.3.1.1 G-quadruplex formation within the CT element of c-MYC

The discovery of the Non-B DNA structures; G-quadruplex and i-motif within the CT element of the c-MYC promoter (Sun and Hurley, 2009) led to the suggestion that these structures may play a role in the regulation of c-MYC. One model that has been proposed describes the transition of the CT element through three DNA structural motifs, two of which lead to activation and one that results in repression of c-MYC (Figure 1.19).

When in duplex form, the transcriptional factor Sp1 binds to the CT element and activates transcription (DesJardins and Hay, 1993). This increased level of transcription could result in the accumulation of supercoiling and flip the CT element into the single-stranded conformation. This would allow the single strand DNA binding proteins hnRNP K and CNBP to bind the purine-rich and pyrimidine-rich strands, respectively in turn maintaining the active state (Brooks and Hurley, 2009, Tomonaga and Levens, 1996, Michelotti et al., 1995). In addition to single-stranded conformation, the CT element can adopt stable non-B DNA structures, a G-quadruplex on the purine-rich strand and an i-motif on the pyrimidine-rich strand (Sun and Hurley, 2009). These globular structures sequester the standard B-DNA transcription factor binding sites and potentially silence transcription.

This model describes a situation in which the same sequence of DNA is able to bind different sets of transcriptional regulators, based not on the sequence specificity but on structural specificity. The result is that gene specific responses could be achieved using ubiquitous transcriptional factors. The DNA itself in terms of its structure and topology is playing an active role in gene regulation, through local flipping between different DNA conformations induced by torsional stress. The structural elements forming within the CT element act like molecular switches

selecting which transcriptional factors to employ based on the physical signals from the cell.

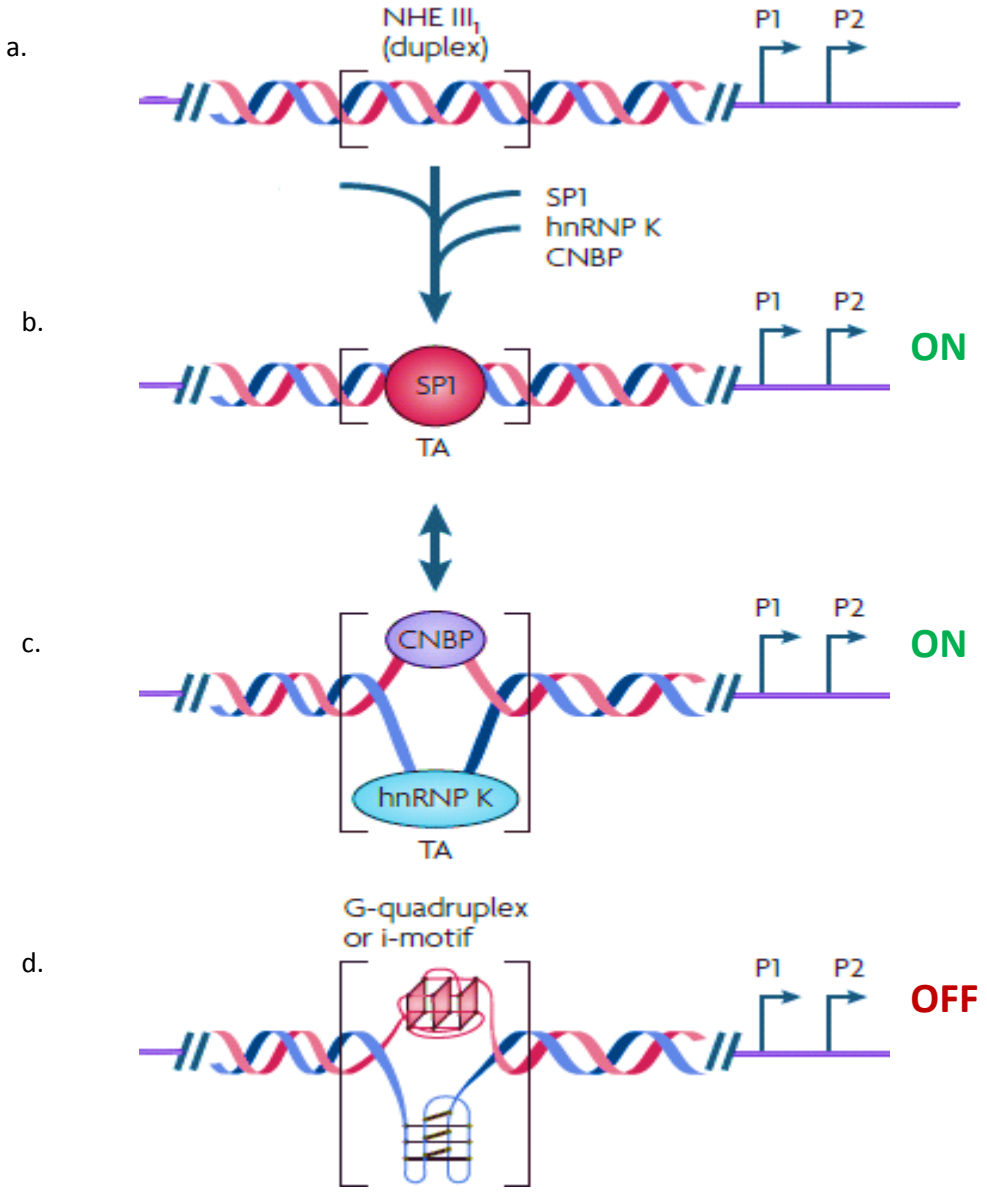


Figure 1.9 A model describing the dynamic transition trough alternative DNA conformations at the CT element.

a.CT element in standard duplex DNA without proteins bound. b. Binding of Sp1 to the duplex structure, resulting in activation of c-MYC expression. c. Binding of single stranded binding proteins hnRNP K and CNBP to the single stranded C and G rich regions, respectively, results in activation of c-MYC transcription. d. The formation of the G-quadruplex and i-motif sequestering duplex DNA binding sites resulting in repression of c-MYC transcription.

1.3.2 Monoamine oxidase A (MAOA)

MAOA is a key brain enzyme which metabolises monoamines, preferentially serotonin (5-HT), noradrenaline (NA) and dopamine (DA). Modulation of the expression of MAOA, resulting in differential levels of active protein, is implicated in the pathophysiology of numerous neuropsychiatric disorders (Reif et al., 2013, Fan et al., 2010). The gene is considered an excellent model for gene x environment (GxE) interactions that underlie mental health (Caspi et al., 2002).

1.3.2.1 MAOA polymorphic regions (VNTRs)

The polymorphic repetitive domain termed the µVNTR (variable number tandem repeat) is considered a major mediator of MAOA gene expression (Sabol et al., 1998). Approximately 20% of MAOA manuscripts address the action of the µVNTR . The µVNTR is a 30 base pair repeat element that is found as 2, 3, 3.5, 4 and 5 copy number variants in the general population. Alleles containing 2, 3 and 5 copy VNTRs are defined as low expressers of MAOA whilst the 3.5 and 4 repetition VNTRs are high expression variants, demonstrating a 2 to 10 fold increased expression. The µVNTR s accepted mode of action is as a transcriptional regulator. These low and high expression alleles can represent either a risk factor or a protective factor depending on the disorder studied (Melas et al., 2013). However such associations are further complicated as the MAOA gene is located on the X chromosome, which makes it potentially more susceptible to sex-linked differences. This is reflected in the fact that genetic associations for MAOA in many disorders have a tendency to demonstrate 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). Consequently many studies of MAOA GxE interactions study only males due to the

compounding effects of gene dosage especially in females heterozygous for the μVNTR (Caspi et al., 2002, Kim-Cohen et al., 2006). This would be consistent with the studies that have found sex differences utilising the μVNTR association. The original hypothesis of X chromosome inactivation of one MAOA allele in the female has more recently been questioned with potentially 15% of genes on the X chromosome escaping inactivation (Mugford et al., 2014, Joo et al., 2014). Whether MAOA escapes X chromosome inactivation is not fully resolved in the literature however there is evidence to support a model in which it escapes inactivation (Carrel and Willard, 2005).

1.3.3 The serotonin transporter (SLC6A4 / 5-HTT)

Serotonin (5-hydroxytryptamine / 5-HT) is a monoamine neurotransmitter and a known neuromodulator of emotional behaviour and psychological states. Instabilities in the serotonergic system are associated with the pathophysiology of many CNS-related disorders including depression (Ressler and Nemeroff, 2000, Lesch et al., 2003), aggression and antisocial behaviour (Lesch and Merschdorf, 2000), anxiety (Ressler and Nemeroff, 2000, Lesch et al., 2003), suicide (Arango et al., 2002) and addiction (Kreek et al., 2005).

The serotonin transporter also termed (5-HTT, SLC6A4, *SERT*) is a Na^+/Cl^- dependent transporter member of the solute carrier gene family (SLC). In the CNS, 5-HT and 5-HTT localise predominantly in the perisynaptic membrane of neurons within the raphe nuclei. Raphe nuclei processes innervate a number of brain areas thought to be involved in behaviour regulation and cognition (McLaughlin et al., 1996, Zhou et al., 2000). Once 5-HT has been released into the synaptic cleft, the 5-HTT mediates its presynaptic reuptake to be recycled by the cell or degraded by monoamine oxidases resulting in serotonergic neurotransmission termination (Rudnick and Wall, 1993). The 5-HTT plays an essential role in regulating the spatiotemporal fine-tuning of the 5-HT neurotransmission by determining the duration and amplitude of 5-HT synaptic signal (Blakely et al., 1991).

Modulation in the expression or action of 5-HTT is therefore associated with behaviour and mood having been strongly linked to depression (Owens and Nemeroff, 1994, Owens and Nemeroff, 1998).

1.3.3.1 Serotonergic signalling pathway a drug target

Antidepressants known as selective 5-HT reuptake inhibitors or SSRIs target the 5-HTT to increase the bioavailability of 5-HT in the synaptic cleft (Roman et al., 2003). Tricyclic antidepressants (TCA) such as amtriptyline also target the 5-HTT and despite the undesirable side effects are still in widespread (Roman et al., 2003).

1.3.3.2. 5-HTT 5' linked polymorphic region (LPR) VNTR

The 5-HTT gene encompasses 14 exons spanning ~ 39 Kb on chromosome 17q11.2, encoding a 630 amino acids protein that has twelve hydrophobic transmembrane domains (Lesch et al., 1994). Two non-coding polymorphic regions are present within the 5-HTT gene. The linked polymorphic region within the 5'-promoter region is present as 14 (short/deletion) or 16 (long/insertion) copies of 22 or 23 bp elements. Stin2 within intron 2 is present as three variants containing; 9, 10 or 12 copies of a 16 or 17 bp element (Figure 1.11). The LPR has been studied extensively for GxE associations (Caspi et al., 2003, Uher and McGuffin, 2008), both have been suggested as predisposing genetic factors for a number of neurological disorders, reviewed in (Haddley et al., 2008). The 5'LPR and Stin2 VNTRs have been identified as transcriptional regulatory domains, despite conflicting evidence in the literature (Lotrich and Pollock, 2004). Modulation of their function could therefore affect levels of 5-HTT expression in a tissue-specific and stimulus inducible manner as reviewed in (Haddley et al., 2008). The transcription factor CTCF has been characterized as a regulator of both VNTRs (Klenova et al., 2004, Roberts et al., 2007, Ali et al., 2010).

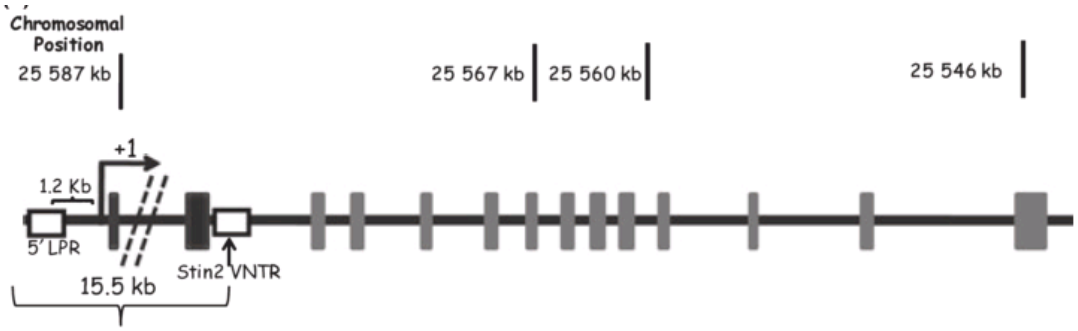


Figure 1.10 Schematic representation of the 5-HTT gene.

Diagram highlights chromosome position, the 5'-LPR and Stin2 VNTRs (white boxes), transcription start site, black arrow with +1 and coding exons and (grey boxes) and non-coding exons (black boxes), (adapted from (Ali et al., 2010)).

1.4 Transcription factors involved in gene regulation via G-quadruplex

1.4.1 Sp1

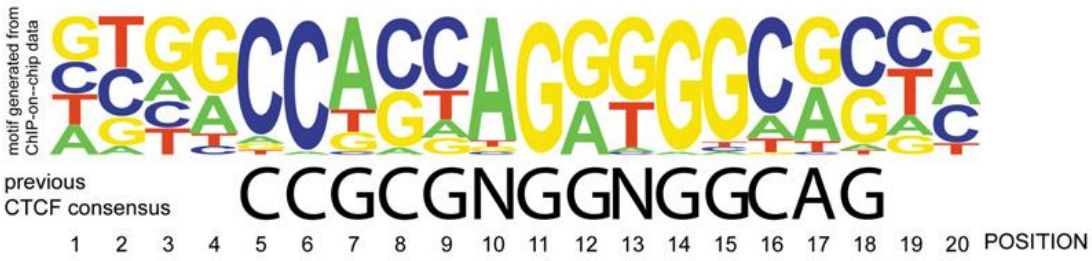
The transcription factor Sp1 is ubiquitously expressed in mammalian cells and plays an important role in the regulation of a number of genes central to many cellular activities including metabolism, cell growth, differentiation, angiogenesis and apoptosis (Deniaud et al., 2009). Sp1 is a 785-amino-acid, 100- to 110-kDa DNA binding protein and is characterized by three Cys2His2-type zinc finger motifs. These zinc fingers interact with the decanucleotide consensus sequence 5'(G/T)GGGCGG(G/A)(G/A)(C/T)-3' in double-stranded DNA (Song et al., 2001). However its binding sequence is known to be degenerate; binding a number of GC-rich motifs (such as 5' -G/T-GGGCGG-G/A-G/A-C/T-3' or 5'-G/T-G/A-GGCG-G/T-G/A-G/A-C/T-3') with high affinity (Briggs et al., 1986, Kadonaga and Tjian, 1986, Kadonaga et al., 1987). Sp1 is able to mediate the expression of TATA-containing and TATA-less genes through protein-protein interactions or via dynamic interplay with additional transcription factors (Olofsson et al., 2007). Such genes include c-MYC (Parisi et al., 2007), c-Jun (McDonough et al., 1997), and Stat1 (Canaff et al., 2008). Interactions with chromatin-modifying factors such as p300 (Suzuki et al., 2000) and histone deacetylases (HDACs) (Zhao et al., 2003), suggest a role for Sp1 in chromatin remodelling.

There are five Sp1 binding sites within the c-MYC promoter, three upstream of P1 and two upstream of P2. The binding sites at the CT element and the one located at –44 base-pairs from the P1 promoter are particularly high-affinity Sp1 binding sites (Geltinger et al., 1996, Wierstra and Alves, 2006). Sp1 binds to the CT-element after serum stimulation of c-MYC, suggesting that it is involved in the serum-induced activation of c-MYC transcription (Liu and Levens, 2006).

Interestingly, it has been reported that for promoters containing multiple Sp1 binding sites, such as the c-MYC promoter, Sp1 exerts its transcriptional synergism through direct protein-protein interaction, in which Sp1 forms higher-order complexes that are able to bind to multiple sites (Mastrangelo et al., 1991, Su et al., 1991). A recent study of the Sp1 binding site within the promoter region of the c-KIT oncogene has suggested that it is able to bind both, canonical Sp1 duplex DNA as well as G-quadruplex structures *in vitro* (Raiber et al., 2012)

1.4.2 CTCF

CTCF is a ubiquitously expressed zinc finger DNA-binding protein which is highly conserved in higher eukaryotes. The full length protein contains an eleven zinc finger central DNA-binding domain that displays close to 100% homology between mouse, chicken, and human (Ohlsson et al., 2001). Combinatorial use of different zinc fingers give CTCF its ability to bind to a wide range of variant sequences in addition to specific co-regulatory proteins leading to its description as a "multivalent factor" (Filippova et al., 1996). This unique structural feature confers a versatile role in genome regulation (Phillips and Corces, 2009), including promoter repression, enhancer insulation and more recently, long-range interactions (Dixon et al., 2012). CTCF functions also extend to roles in differentiation and development (Torrano et al., 2005), regulation of cell proliferation (Rasko et al., 2001), tumour suppression (Filippova et al., 1998, Filippova et al., 2002) and apoptosis (Docquier et al., 2005).Thousands of CTCF DNA-binding sequences are present in the genome, including the CT element within the promoter of c-MYC, (Lobanenkov et al., 1990, Klenova et al., 1993) and the 5' LPR VNTR within the promoter of 5-HTT (Ali et al., 2010, Vasiliou et al., 2012).



1.14 DNA logo representing CTCF-binding motif.

Traditional CTCF consensus is shown in black, colored letters of various height represent the relative frequency at which the nucleotide occurs at each position, (adapted from (Kim et al., 2007)).

1.4.2 Nucleolin

Nucleolin is a 110-kDa multifunctional nucleolar phosphoprotein with an active role in numerous cellular processes; including ribosome biogenesis, cell proliferation, chromatin decondensation, differentiation and maintenance of neural tissue, transcriptional regulation, and apoptosis (Storck et al., 2007, Mongelard and Bouvet, 2007, Grinstein et al., 2007, Grinstein et al., 2002, Ginisty et al., 1999). Nucleolin is described as a modular protein as it can be structurally divided into three discrete domains: 1) The N-terminal, consisting of highly acidic regions punctuated by basic sequences housing multiple phosphorylation sites. 2) The central domain, which contains four RNA-binding motifs. 3) The C-terminal, defined by spaced RGG repeats punctuated by amino acids. It is the association of these structural domains that allow nucleolin to interact with different proteins and RNA sequences (Ginisty et al., 1999).

Of particular note are the circular dichroism (CD) spectroscopy studies which have shown that nucleolin will bind to G-quadruplex DNA. It has been shown that the interaction of nucleolin with the CT element within the c-MYC promoter induces the formation of a G-quadruplex from single-stranded DNA (Gonzalez and Hurley, 2010b). Gonzalez *et al.*,2002 also demonstrated G-quadruplex selectivity as nucleolin appears to preferentially bind to parallel G-quadruplexes with short loops as in the c-MYC G-quadruplex over other G-quadruplex formations. The group also suggest that nucleolin may regulate c-MYC transcription by modulating the structure of the CT element based on the high affinity and selectivity of nucleolin for the c-MYC G-quadruplex structure.

1.4.4 hnRNP K

hnRNP K is a 463-residue modular nucleic acid binding protein characterized by the presence of KH domains. Unlike most transcription factors this class of binding protein does not target double stranded DNA. *In vivo* and *in vitro* studies have shown these proteins to specifically bind and transactivate single-stranded cis elements within the c-MYC promoter. (Matunis et al., 1992, Bomsztyk et al., 1997, Ostareck-Lederer et al., 1998).

Interactions with nucleic acids (ssDNA and RNA) are mediated by three KH domains (Tomonaga and Levens, 1995). hnRNP K interacts with single-stranded nucleic acids via folding of the KH domains to form an elongated groove on the proteins surface. This groove is the site at which the protein – DNA interaction occurs via hydrogen bonds and van der Waals contacts (Tomonaga and Levens, 1995). hnRNP K KH1 (residues 32–112) and KH2 (residues 142–217) are located at the N-terminal end of the protein, separated by a 30-residue linker (Braddock et al., 2002). hnRNP K KH3 is isolated from the other two KH domains and located at the C-terminal end of the protein (residues 389–459). The intervening 172-residue stretch between KH2 and KH3 is involved in protein–protein interactions with multiple partners that include other transcription factors (such as TATA binding protein and various zinc finger-containing transcriptional repressors). This is in addition to proteins involved in diverse signal transduction pathways (such as a number of tyrosine and serine/threonine kinases, and the Vav proto-oncoprotein) (Bomsztyk et al., 1997). hnRNP K therefore has regulatory roles in transcription and translation and participates in a number of signalling systems (Bomsztyk et al., 1997, Ostareck-Lederer et al., 1998).

Given the fact that hnRNP K binds preferentially to single-stranded nucleic acids any interactions with DNA are associated with duplex melting due to increased levels of negative supercoiling (Duncan et al., 1994). In support of this, hnRNP K has been identified as binding to the pyrimidine-rich single strand of the CT element and activating the promoter of c-MYC *in vivo* and *in vitro* (Duncan et al., 1994, Michelotti et al., 1996a, Tomonaga and Levens, 1996).

1.4.5 CCHC-type zinc finger nucleic acid binding protein (CNBP)

CCHC-type zinc finger nucleic acid binding protein (CNBP) is a multifunctional protein comprised of seven cysteine-cysteine histidine-cysteine zinc knuckles and an arginine-glycine-glycine (RGG) domain. The protein has been shown to control cell proliferation and survival, implicating it in embryonic development, especially in forebrain and craniofacial development (Chen et al., 2003).

CNBP has been shown to bind to the purine-rich single strand of the c-MYC CT element followed by induction of c-MYC expression, it has therefore been suggested that CNBP may function to mediate c-MYC expression (Michelotti et al., 1995, Armas et al., 2008b). A function for CNBP in the regulation of c-MYC is supported by studies into CNBP−/− mouse embryos, which have been shown to have a substantial reduction in cell proliferation, correlating with the absence of c-MYC expression (Chen et al., 2003). The RGG domain of CNBP has been identified as essential for its nucleic acid binding activity and induction of c-MYC expression. The zinc knuckles of the protein only partially contribute towards its function in this capacity, as identified by functional mutation analysis (Armas et al., 2008a). Interestingly proteolytic mechanisms have been proposed as the way in which CNBP's biochemical activity is regulated owing to the identification of truncated forms of the protein (lacking the RGG domain) naturally occurring (Flink and Morkin, 1995).

Aims and objectives:

1. Address the potential for repeat elements in the promoter of selected genes to form G-quadruplex using both in-silico and preliminary biochemical analysis; transcription factor binding and methylation status at the promoters of selected genes.

2. Use AAV targeting vectors to destabilise G-quadruplex structures to assess directly the capacity of and mechanism through which G-quadruplex influences gene expression, initially validating the approach in the c-MYC promoter with the well characterized CT element.

3. Combine emerging technology CRISPR-Cas system with the traditional AAV targeting vectors to produce a highly efficient, high fidelity and streamlined approach to gene editing.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Commonly used solutions and reagents

2.1.1.1 General Molecular Biology

5x TBE buffer

54 g Tris(hydroxymethyl)methylamine (Tris) (BDH), 27.5 g boric acid (Sigma), 20 ml 0.5 M Ethylenediamine Tetraacetic Acid (EDTA) (Sigma) and dH_2O up to 1 litre.

6x agarose gel loading buffer

0.25% (w/v) Bromophenol blue, 0.25% (w/v) xylene cyanol, , 1 mM EDTA, 30% glycerol and 70% water.

2.1.1.2 Chromatin Immunoprecipitation (ChIP)

Cell lysis buffer

50mM Hepes-KOH pH7. 5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, protease inhibitor cocktail (Sigma Cat No. P8340)

Nuclear lysis buffer

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

ChIP dilution buffer

16.7 mM Tris–HCl, pH 8.1, 167 mM NaCl, 1.1%Triton X-100, 0.01% SDS, 1.2 mM, EDTA, protease inhibitor cocktail (Sigma)

Magnetic beads

Dynabeads Protein G (Invitrogen™ 10003D)

Low-salt wash buffer

20 mM Tris–HCl, pH 8.1, 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA,

High-salt wash buffer

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

Lithium chloride (Li Cl) wash buffer

10 mM Tris–HCl, pH 8.1, 250 mM LiCl, 1% Igepal, 1% sodium deoxycholate, 1 mM EDTA.

TE buffer

10 mM Tris–HCl, pH 8.0, 1 mM EDTA.

Elution buffer

50 mM Tris–HCl, pH 8 1 mM EDTA, 1%SDS, 50 mM NaHCO 3, 2. 5 M NaCl.

RNAse

RNAse A: 20 mg/ml. (Sigma)

SDS 10%.

Sodium dodecyl sulfate (Sigma) 10% w/v in dH_2O

Proteinase K.

Proteinase K solution (Qiagen)

2.1.1.3 Microbiology

Super optimal growth media with catabolite repression (SOC)

20 g/L Tryptone, 5 g/L Yeast Extract, 4.8 g/L MgSO4, 3.603 g/L dextrose, 0.5g/L NaCl, 0.186 g/L KCl.

Luria broth (LB) media

10.0 g/L trypton, 5.0 g/L yeast extract and 10.0 g/L NaCl.

LB agar

10.0g/L trypton, 5.0 g/L yeast extract and 12 g/L agar.

2.1.1.4 Chemical stimuli given to cells

1 M Lithium chloride (Sigma Cat. No. L9650)

0.42 g in 10 ml of dH_2O, used at 1 mM (1 µl in 1 ml of appropriate tissue culture media).

1 mM Cocaine hydrochloride (Sigma Cat. No. C5776)

0.0034 g in 10 ml of dH_2O, use at 1 µM (1 µl in 1 ml of appropriate tissue culture media) or at 10 µM (10 µl in 1 ml of appropriate tissue culture media).

10mM Valproic acid sodium salt (Sigma Cat. No. P4543)

0.16g in 10ml of dH_2O sterile filtered and used at 2µM (8µl in 40ml appropriate tissue culture media).

2.1.1.5 Cell and tissue culture media

JAr cell line media

RPMI-1640 medium (Sigma) supplemented with 10% heat-inactivated foetal calf serum (Sigma), 2 mg/ml glucose, 1 mM sodium pyruvate, 2 mM L-glutamine, 10 mM HEPES and 1% (v/v) 100x penicillin/streptomycin (equates to a final concentration of 100 units penicillin/100 µg streptomycin (Sigma-Aldrich; Cat. No. P0781).

SH-SY5Y cell line media

1:1 solution of Eagle's minimal essential medium (EMEM) media and nutrient mixture F12 HAM media supplemented with 10% (v/v) heat-inactivated foetal bovine serum (Sigma), 1% of L-Glutamine 2mM, 1% of Sodium Pyruvate 1mM and 1% penicillin 100 units/ml / streptomycin 0.1 mg/ml (Sigma).

HCT-116 cell line media

Dulbecco's Modified Eagle's medium (DMEM), (Sigma Cat No. 5546) supplemented with 2mM Glutamine and 10% Foetal Bovine Serum (Sigma Cat. No. F3018) 1% penicillin 100 units/ml / streptomycin 0.1 mg/ml (Sigma).

HCT-116 selection media

DMEM (Sigma Cat No. 5546) supplemented with 2mM Glutamine and 10% Foetal Bovine Serum (Sigma Cat. No. F3018), 0.05uM G418

HEK-293 cell line media

Dulbecco's minimum essential media (DMEM) F12, supplemented with 10% foetal bovine serum, 1% l-glutamine [200 mM], 1% sodium pyruvate [100 mM], and 1% penicillin [10,000 u/mL]/streptomycin [10 µg/mL].

2.1.1.6 Plasmids

Reporter gene plasmids cloned and used during this study are listed in Table 2 of the appendix.

2.2 Methods

2.2.1 General cloning methods

2.2.1.1 PCR primer design

Primers were designed with the aid of primer design programme; primer3 http://biotools.umassmed.edu/bioapps/primer3_www.cgi, secondary structure and primer dimer formation were prevented with the aid of integrated DNA technologies (IDT), oligonucleotide analysis software; oligoanalizer http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/. Primers were generally designed between 20-25 bp in length, with a melting temperature of 50-65°C and GC content between 40-60%.

2.2.1.2 Standard PCR

PCR was used as a method for analysing mRNA expression and to amplify specific DNA fragments for use in molecular cloning. PCR amplifications were carried out on a peqSTAR 2X Gradient Thermocycler (PeqLab). 25µl PCR reactions contained 5-100 ng DNA template, 0.5mM dNTP (Sigma), 0.4 µM primer (Eurofins or IDT), 0.625u GoTaq® DNA Polymerase / HotStart GoTaq® DNA Polymerase (Promega), 5× Green GoTaq® Reaction Buffer (Promega) specific primer and PCR conditions are detailed in Table 1 of the appendix.

2.2.1.3 PCR purification

To purify double-stranded DNA fragments post PCR (Section 2.2.1.2), restriction enzyme digestion (Section 2.2.6) and ChIP (Section 2.6) for accurate quantification or downstream reactions such as Gibson Isothermal Assembly, Wizard® SV Gel and PCR Clean-Up System was used following manufacturer's instructions. Briefly; equal volume of membrane binding solution was added to the PCR amplification, an SV minicolumn was inserted into the collection tube, to which the prepared PCR product was added, incubated at room temperature for 1 minute, and centrifuged at 16,000 × g for 1 minute. The flow through was discarded and the minicolumn reinserted into collection tube, which was washed by adding 700µl of membrane wash solution and centrifuged at 16,000 × g for 1 minute. The flow through was discarded and minicolumn reinserted into the Collection Tube, this wash step was then repeated with 500µl membrane wash solution and centrifuged at 16,000 × g for 5 minutes. The collection tube was emptied and the column assembly re-centrifuged for 1 minute with the microcentrifuge lid open to allow evaporation of any residual ethanol. To elute; the minicolumn was transferred to a clean 1.5ml microcentrifuge tube and 50µl of nuclease-free water was added to the minicolumn. This was then incubated at room temperature for 1 minute and centrifuged at 16,000 × g for 1 minute, the minicolumn was discarded and DNA stored at 4°C or –20°C.

2.2.1.4 Analysis of DNA using agarose gel-electrophoresis

Agarose gel-electrophoresis was used to analyse PCR products or fragments generated by restriction digestion, 1-2% agarose (multi-purpose agarose, Bioline, Cat. No. BIO-41025) was melted in 120ml 0.5x TBE buffer and supplemented with 6µl ethidium bromide (10 mg/ml aqueous solution, Sigma E-5134). Gels of 120 ml were cast in 12x14 cm or 20.5x10 cm trays and the appropriate combs inserted. Gels were left at room temperature for 30 minutes to set, and were then submerged in horizontal gel electrophoresis tanks (Hybaid turn and cast submarine gel system, Hybaid, or Savant HG 350 tank) containing 0.5x TBE buffer. Samples were mixed with 6x loading buffer (1x final concentration) and loaded into the wells. The size of a PCR product or restriction digest fragments was determined by loading a DNA ladder (100bp ladder; Promega Cat. No. G2101 or 1Kb ladder; Promega, Cat. No. G5711). In general, gels were run for between 45-180 minutes at 120V (Hybaid) depending on the fragment or combination of fragments (as in the case of VNTR's) Electrophoretically separated DNA was visualised using a UV transilluminator (BioDoc-it Imaging System).(BioDoc-it Imaging System) at a wavelength of 302 nm.

2.2.1.5 Recovery of DNA from agarose-gels

PCR products and restriction digestion fragments were isolated by running the PCR reaction or the restriction digest on agarose gels. The bands which corresponded to products of the predicted size were excised from the agrose gel under long wave UV translumination using a clean blade. The DNA was recovered from the gel slice using the Wizard SV Gel and PCR Clean-Up System, following manufacturer's instructions. Briefly; the gel slice was placed in a 1.5ml microcentrifuge tube, 10µl Membrane Binding Solution was added per 10mg of gel slice and vortexed and incubated at 50–65°C until the gel slice completely dissolved. Once dissolved an SV Minicolumn was inserted into the Collection Tube, to which the gel mixture was added. This was incubated at room temperature for 1 minute, and centrifuged at 16,000 × g for 1 minute, the flow through was discarded and the Minicolumn reinserted into Collection Tube. The Collection Tube was washed by adding 700µl of Membrane Wash Solution and centrifuged at 16,000 × g for 1 minute. The flow through was discarded and Minicolumn reinserted into the Collection Tube, this wash step was then repeated with 500µl Membrane Wash Solution and centrifuged at 16,000 × g for 5 minutes. The Collection Tube was emptied and the column assembly re-centrifuged for 1 minute with the microcentrifuge lid open to allow evaporation of any residual ethanol. To elute; the Minicolumn was transferred to a clean 1.5ml microcentrifuge tube and 50µl of nuclease-free water was added to the Minicolumn. This was incubated at room temperature for 1 minute and centrifuged at 16,000 × g for 1 minute, the Minicolumn was discarded and DNA stored at 4°C or –20°C.

2.2.2 Gibson Isothermal Assembly

2.2.2.1. Gibson Isothermal Assembly design

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 1hour reaction. The technique was used to construct recombinant rAAV vectors, a detailed description of the biochemistry and approach is given in chapter 4.3.2.2.

2.2.2.2 Removal of 3' overhangs – T4 DNA polymerase treatment

When assembling DNA fragments linearized by or isolated from vectors using restriction enzymes which leave a 3' overhang, the ends were blunted using T4 DNA polymerase. DNA was dissolved in 1x NEBuffer supplemented with 100 µM of each dNTP, 1 unit of T4 DNA Polymerase per microgram of DNA was added and then Incubated for 15 minutes at 12°C, The reaction was stopped by adding EDTA to a final concentration of 10 mM and heating for 20 minutes at 75°C.

2.2.2.3. Gibson Isothermal Assembly 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.

2.2.3 Clustered Regularly Interspaced Short Palindromic Repeats - CRISPR associated system (CRISPR-Cas) specific cloning methods.

2.2.3.1 Annealing oligonucleotides

In order to anneal the sense and anti-sense single stranded oligonucleotides which become the guide RNA (gRNA) 5 µl of each oligonucleotides (1 µg/µl) were added to 85 µl of nuclease free H_2O and 5 µl of ligase buffer NEB, heated at 95 °C for 5 minutes and allowed to cool to room temperature (~60 min).

2.2.3.2 Golden Gate cloning

Golden Gate cloning was used as strategy to insert gRNA (annealed oligos) into the pGUIDE vector (a kind gift from Dr Patrick Harrison; University College Cork) rather than gel purifying the backbone. 150ng of pGUIDE plasmid was mixed with 150ng CRISPR gRNA oligos (annealed in Section 2.2.3.1), 1 µl *Bse*RI (NEB), 1µl T4 ligase (Promega), 2 µl ligase buffer (Promega) and dH_2O to a total volume of 20µl. The reaction was incubated at 37 °C for 5 min, then 16 °C for 10 min for 10 cycles, followed by 30 mins at 37°C and finally 20 mins at 80°C to heat inactivate the enzymes.

2.2.3.3 Transformation of chemically competent *E. Coli* cells DH5α™

Once the generation of recombinant plasmid DNA was confirmed by enzymatic digest (Section 2.2.5) and sequencing (Section 2.2.6), plasmids were transformed into strains of competent *E. coli* cells (DH5α™, Invitrogen™ Cat No.18265-017). Briefly, 50µl aliquot of competent cells was defrosted on ice, the ligation reaction (10µl) or 10ng of plasmid DNA were added to the defrosted cells and subsequently incubated on ice for 30 minutes. The cells were subjected to heatshock in a water bath for 45 seconds at 42°C and then incubated on ice for 2 minutes. 950 µl of pre-warmed LB broth was added to the cells, and the culture incubated at 37 °C for 1 hour on a shaker at 225 rpm. 50-200 µl of this culture was spread onto LB agar plates supplemented with 100 µg/ml ampicillin, and grown at 37 ˚C overnight.

2.2.4 Isolation of DNA constructs from bacteria

2.2.4.1 Mini-preparation of plasmid DNA

A small scale preparation of plasmid DNA, for up to 20 μg, was used for screening plasmids after manipulation for molecular cloning. The QIAprep Spin Miniprep Kit (Qiagen, Cat No. 27106) was used for this purpose. This kit uses a modified alkaline lysis method, and the lysate is neutralised and adjusted to high salt binding conditions. The neutralised lysate is cleared by centrifugation, before being applied to a silica-gel membrane which selectively absorbs DNA in high-salt conditions. Endonucleases are removed by a wash with buffer PB, and salts are removed by a wash with buffer PE (both part of the kit). The plasmid DNA was eluted in nuclease-free water.

2.2.4.2 Maxi-preparation of plasmid DNA

For the isolation of up to 500 μg of plasmid DNA the QIAGEN Plasmid Maxi Kit (Qiagen, Cat No. 12165) was used. The kit utilizes a modified alkaline lysis procedure which results in a cell lysate containing plasmid DNA among protein, chromosomal DNA and other cell debris. Debris is cleared from the lysate in a neutralising potassium acetate buffer, and the plasmid DNA contained in the supernatant is bound to an anion-exchange column under high salt and low pH conditions. Medium-salt washes remove RNA, proteins etc. and the plasmid DNA is eluted with a high-salt wash. The eluted DNA is then precipitated with isopropanol, washed with 70% ethanol, and resuspended in nuclease-free water.

2.2.5 Analytical restriction enzyme digests

Restriction enzymes were used for molecular cloning and to verify the insertion and position of DNA fragments into plasmid vectors. Restriction enzyme (~5 unit/1 µg DNA) digests were carried out in 1x restriction enzyme buffer. The digests were carried out at the appropriate temperature for the respective enzyme for a minimum time of 1 hour. DNA double digestion sequentially using two restriction endonuclease enzymes was performed when the two enzymes buffer salt concentration were not compatible, in such cases the initial reaction was performed using the enzyme that functioned in a low salt buffer, followed by digestion with the second enzyme that function in the high salt buffer. The second digest was set up adjusted to the volume of the first reaction. Enzymes were obtained from Promega and NEB. The fragments generated by restriction enzyme reaction were visualized after gel electrophoresis in a UV light transllluminator.

2.2.6 Sequencing

DNA sequencing was performed by Source Bioscience Life Sciences Sanger sequencing service (Source BioScience LifeSciences, 1 Orchard Place, Nottingham, Business Park, Nottingham, NG8 6PX).

2.2.7 Measurement of nucleic acid concentration by spectrophotometry

Nucleic acid concentrations were quantified using the NanoDrop™ 8000 Spectrophotometer (Thermo Scientific) with samples showing 260/280 ratios of ~1.8. for DNA and ~2 for RNA. A NanoDrop™ 8000 Spectrophotometer was calibrated using 1 µl dH_2O or TE buffer (TE buffer 10 mM Tris-Cl, pH 7.5 1 mM EDTA) as blank, 1 µl of sample was then loaded onto the pedestal, the arm was closed to form a sample column, the measurement was performed.

2.2.8 Cell and tissue culture

2.2.8.1 Culture of JAr (human placental chorocarcinoma) cells

JAr cells (ATCC HTB-144; American Type Culture Collection, Manassas,VA) were maintained as monolayers in supplemented RPMI-1640 medium (Sigma) (Section 2.1.1.5). Sub-confluent cultures were split at (70-80%) 1:3 to 1:10 i.e. seeding at 1-3x10,000 cells/cm^2 using 0.25% trypsin or trypsin/EDTA; 5% CO_2; 37°C.

2.2.8.2 Culture of SH-SY5Y (human derived neuroblastoma) cells

SH-SY5Y cells were maintained as monolayers in 1:1 solution of Dulbecco's EMEM media and nutrient mixture F12 HAM (Section 2.1.1.5) Sub-confluent cultures were split at (70-80%) 1:3 to 1:10 i.e. seeding at 1-3x10,000 cells/cm^2 using 0.25% trypsin or trypsin/EDTA; 5% CO_2; 37°C.

2.2.8.3 Culture of HEK-293 (Human embryonic kidney) cells

HEK-293 cells, were maintained in monolayers in DMEM, (Section 2.1.1.5) Sub-confluent cultures were split at (70-80%) 1:3 to 1:10 i.e. seeding at 1-3x10,000 cells/cm^2 using 0.25% trypsin or trypsin/EDTA; 5% CO_2; 37°C.

2.2.8.4 Culture of HCT-116 (human colorectal carcinoma) cells

HCT-116 cells (Sigma Cat. No. 91091005) were maintained as monolayers in McCoy′s 5a (Described in Section 2.1.1.5) sub-confluent cultures were split at (70-80%) 1:3 to 1:10 i.e. seeding at 1-3x10,000 cells/cm^2 using 0.25% trypsin or trypsin/EDTA; 5% CO_2; 37°C.

2.2.8.5 Lithium, cocaine and sodium valproate cell treatments

Cells were incubated in appropriate medium (see Section 2.1.1.5) for 24 hrs and supplemented with either 1 mM lithium chloride, 2µM sodium valproate or 10 µM cocaine (described in Section 2.1.1.4) for 1hr. Following treatment the cells were harvested. Lithium, cocaine and sodium valproate treated cells were subsequently used in, standard PCR, chromatin immunoprecipitation (ChIP) and methylation (MethylQuest) assays.

2.2.9 Co-transfection of rAAV and CRISPR-Cas plasmids.

2.2.9.1 Lipofectamine LTX transfection (rAAV)

Lipofectamine LTX transfection reagent (Invitrogen) was used to co-transfect recombinant AAV vectors rAAV-G4-KO-CT or rAAV-LoxP-CT with the helper plasmids, pDG or -pDP2rs (Plasmid Factory), into HEK-293 cells, enabling assembly of functional AAV. For transfection, x10 T75 flasks containing $4x10^6$ cells in 10 ml DMEM were incubated for 72 hours. Following incubation, the media was replaced with 9 ml of new media and incubated for a further 30 minutes. For each transfection, 1 ml serum free DMEM was supplemented with plasmid DNA, 18.75 µg pAAV-G4-KO-CT or pAAV-LoxP-CT and 18.75 µg pDG or -pDP2rs in addition to 37.5 µl Plus reagent (Invitrogen). This was followed by a 5 minute incubation at room temperature. 56 µl Lipofectamine LTX was then added to the serum-free DMEM and incubated for a further 30 minutes at room temperature. Following incubation 1 ml of the Lipofectamine LTX/media mix was aliquot across each flask, to ensure even distribution flasks were briefly rocked back and forth; cells were incubated for a further 72 hours.

2.2.9.2 TurboFect transfection (CRISPR-Cas)

The TurboFect transfection reagent (Thermo scientific) was used for transfection of CRISPR plasmids pGUIDE and Cas9-WT DNA into SH-SY5Y cells. The TurboFect transfection reagent is comprised of a proprietary cationic polymer in water, the polymer forms positively-charged, complexes with DNA, these complexes are compact and stable thus protecting the DNA from degradation and facilitate efficient plasmid delivery into eukaryotic cells.

24 hrs prior to transfection, $3x10^5$ cells were seeded into 12-well plates in 400 µl of growth medium to give a confluency of 70-90% at the time of transfection. Equal molar concentrations of pGUIDE and Cas9-WT (3µg) plasmid DNA was diluted in 20 µl of serum-free DMEM. TurboFect reagent was briefly vortexed and 12µl was added to the diluted DNA, this was then mixed immediately by vortexing, followed by a 20 minute incubation at room temperature. 20 µl of the transfection reagent/DNA mixture was added in a drop-wise manner to each well. The plate was gently rocked to achieve even distribution of the complexes; plates were then incubated at 37 °C in a CO_2 incubator.

2.2.10 mRNA analysis

2.2.10.1 Extraction of total RNA from JAr, SH-SY5Y and HCT-116 cells.

Total RNA was extracted from cells grown in 6-well culture plates to 70-80% confluency. Cell media was removed, followed by two washes with 1x PBS, the cells were lysed directly by adding 1 ml of TRIzol reagent (Invitrogen), transferred to a 1.5 ml microcentrifuge tube and incubated for 5 minutes at 15-30 °C to permit the complete dissociation of nucleoprotein complexes. DNA and protein were removed through phase separation by adding 200 µl of chloroform solution; samples were mixed by inversion for 15 seconds and incubated for 2-3 minutes at 21 °C. Samples were then centrifuged at 12,000x *g* for 15 minutes at 4 °C. Following centrifugation the mixture separated into lower red, phenol-chloroform phase, an interphase, and a colourless upper aqueous phase. The RNA containing aqueous phase was transferred to a sterile-RNase free microcentrifuge tube containing 500 µl 100% isopropanol and incubated for 10 minutes at 21°C. The total RNA was precipitated by centrifugation at 12,000 x *g* for 10 minutes at 4°C. Total RNA pellet was washed with 75% ethanol, and centrifuged again at 7,500x *g* for 5 minutes at 4°C. The RNA pellet was then left to dry for 15 minutes at room temperature to allow evaporation of residual ethanol. To dissolve the RNA, the pellet was resuspended in 20 µl of RNase free water and incubated for 10 minutes at 55°C; RNA samples were stored at -80˚C.

2.2.10.2 DNase digestion of total RNA extraction

Harvested total RNA was digested with RQ1 RNase-free DNase (Promega), in order to ensure that all DNA contaminants were removed. 1 µg of total RNA was digested with 1 µl (1unit) RQ1 RNase-free DNase in a 10 µl volume reaction containing 1 µl of RQ1 10x buffer (Promega). The reaction was performed at 37 ˚C for 30 minutes. 1 µl of Stop-solution (Promega) was added, and the enzyme was inactivated at 65 ˚C for 10 minutes. Successful removal of DNA was determined by visualisation of RNA and DNase treated RNA on a 1% agarose gel in electrophoresis. The RNA concentration was determined with NanoDrop™ 8000 Spectrophotometer (Section 2.2.7)

2.2.10.3 cDNA synthesis

Promega's Reverse Transcription System (Cat. No. A3500) was used for the reverse transcription step following manufacturer's instructions. Briefly 1µg of DNase treated RNA was used as template in each RT- PCR reaction.

Standard reverse transcription (RT) reaction:

$MgCl_2$, 25mM	4 µl
Reverse Transcription 10x Buffer	2 µl
dNTP Mixture, 10 mM	2 µl
Recombinant RNasin Ribonuclease Inhibitor	0.5 µl
AMV Reverse Transcriptase	15u
Random Primers	0.5 µg
Total RNA	1 µg
Nuclease-Free water to a final volume of	20µl

In order to achieve first strand cDNA synthesis the reverse transcription reaction mix was incubated at room temperature for 10 minutes, followed by 15 minutes at 42 ˚C, then 95˚C for 5 minutes and rapidly cooled to 4 ˚C for 5 minutes. As a negative control the reaction was carried out with dH_2O alone in the absence of RNA. The cDNA concentration was determined using the NanoDrop™ 8000 Spectrophotometer (Section 2.2.7). For quantitative PCR the resulting cDNA was diluted to10 ng/µl in nuclease-free water and subsequently 0.5-10 µl (5-100 ng) of this dilution were used as template in PCR's.

2.2.10.4 PCR to analyse mRNA expression

PCR was used as a method for analysing mRNA expression. PCR was performed in a peqSTAR 2X Gradient Thermocycler. Promega's GoTaq DNA polymerase and dNTPs were generally used unless otherwise stated. Primers (IDT or Eurofins genomics) were prepared as a 100 μM stock and stored at -20 ˚C (see Table 1 in Appendix for primer sequences). 5-100 ng of first-strand cDNA was used in the PCR reaction as below:

30ng first-strand cDNA	3 μl
dNTPs Mixture, 10mM	0.5 μl
$MgCl_2$, 25mM	1.5 μl
5x Buffer	5 μl
Forward primer	50 pmol
Reverse primer	50 pmol
Taq DNA polymerase	2.5 units
Nuclease-Free water to a final volume of	25 μl

PCR programmes were adapted from a standard protocol. The annealing temperature varied according to the melting temperature of the primer pairs used and the extension time was adapted to the size of the expected product, with 1 minute extension time for each 1000bp as a guideline. The most commonly used reaction conditions were; one cycle of 94 ˚C for 10 min, followed by 35 cycles of: denaturation at 94 ˚C for 1 min, annealing: 60 ˚C for 1 min, extension: 72 ˚C for 1 min and final extension to complete strands: 72 ˚C for 10 min.

2.2.10.5 Quantitative real time PCR

Quantitative real-time PCR was use to quantify viral DNA to obtain a viral titre. The reaction was performed in a qPCR machine (Bio-rad CFX Connect™ Real-Time PCR Detection System) using the GoTaq® qPCR Master Mix (Promega). This formulation includes a proprietary dsDNA-binding dye, which excites in the presence of dsDNA and the level of fluorescence is directly proportional to the copy number of amplicons produced. Standards were prepared by diluting rAAV plasmid in nuclease free H_2O in a series of 5-fold serial dilutions, ranging from 1 µg/ ml to $2x10^{-8}$ µg/ ml. 12.5 µl GoTaq® qPCR Master Mix master mix (containing GoTaq® Hot Start DNA polymerase, optimized PCR buffer, 5 mM MgCl, dNTP mix including dUTP) and 0.18 µM of each primer (see Table 1 of the appendix for primer sequences) were added to each of the diluted standards or AAV DNA sample, performed in triplicate. The GoTaq® Hot Start DNA polymerase incorporates a non-specific DNA-binding domain that lends physical stability to the polymerase-DNA complex. qPCR cycling steps were carried out as outlined in the BioRad qPCR machine (block pre heating 95 °C for 10 min, denaturation 94°C for 30 seconds, annealing 60 °C 40 seconds for 45 cycles.) Data was analysed using BioRad CFX manager software.

2.2.10.6 PCR for amplification of DNA fragments enriched by ChIP and methylation assays

A range of PCR reaction conditions were used to amplify the various VNTR's and fragments of interest, specific cycling and reaction conditions can be found in Table 2 of the appendix. In general PCR reactions contained 5ng DNA template, 0.5mM dNTP, 0.4 μM primer, 0.625u GoTaq® DNA Polymerase, 5× Green GoTaq® Reaction Buffer and betaine (Sigma) at the final concentration of 0.5M. The PCR amplifications were carried out on a peqSTAR 2X Gradient Thermocycler (PeqLab) under the following touch-down protocol: initial denaturation step (95 °C for 10 min); first cycles of denaturation (95 °C for 0.5 min); annealing (65 °C for 0.5 min); elongation (72 °C for 1.5 min), followed by repeated cycles with a 1°C decrease in annealing until 56 °C is reached followed by 30 cycles of denaturation (95 °C for 0.5 min); annealing (55 °C for 0.5 min); elongation (72 °C for 1.5 min), final elongation step (72 °C for 30 min). PCR products were analysed using a 1.5% agarose gels stained with ethidium bromide solution (Sigma) and visualised using a UV transilluminator (BioDoc-it Imaging System).

2.2.11 Chromatin Immunoprecipitation (ChIP)

2.2.11.1 Cell fixation and chromatin isolation

Human SH-SY5Y cells were grown to 80% confluency (~$3x10^7$ cells) in a T175 culture flask, when ready to harvest, cells were fixed by adding 1% (v/v) of 37% formaldehyde (Sigma) directly to culture media and incubated on a shaking platform for 8 minutes, cross linking was quenched by adding 4ml of 1.25 M glycine to a final concentration of 125mM. Cells were washed three times with 10ml ice-cold 1x PBS (Sigma) supplemented with 10µl protease inhibitor cocktail (PIC) (Sigma) and 100µl 0.1M PMSF (Sigma). Cells were resuspended in 5 ml cellular lysis buffer (Section 2.1.1.2) including PIC and incubated end over end for 10 minutes at 4°C followed by centrifugation at 3,500 rpm for 5 minutes at 4°C, supernatent was aspirated off and this step was repeated using nuclear lysis buffer (Section 2.1.1.2). Following centrifugation the supernatant was once again aspirated off and the lysate pellet resuspended in 1.5 ml sonication buffer (Section 2.1.1.2)

2.2.11.2 Shearing of chromatin using sonication.

The 1.5ml chromatin aliquots were sheared by sonication, using a BioRuptor (Diagenode) set at 50 cycles of 30 second pulses followed by 30 second pauses on high power. The sonicated lysate was transferred to a 1.5ml microcentrifuge tube on ice and centrifuged at 10,000 × *g* for 10 min at 4°C to pellet the cell debris. The supernatant (containing the chromatin) was transferred to a new 1.5ml microcentrifuge tube. 50ul was taken from the supernatant, for fragment size analysis and the remaining sonicated sample stored at −80°C.

To estimate the sonicated DNA concentration and fragments size, 50ul nuclease free water, 6 ul 5M NaCl and 2ul RNase1 (promega) (10u/ul) was added to

the 50ul sample of sonicated chromatin. This was then vortexed and incubated at 37°C for 30mins. 2ul Proteinase K was added to the RNase digested sample which was then vortexd and incubated at 65 °C for 2hrs. DNA was then purified using the *Wizard* SV *Gel* and *PCR* Clean-Up System (Promega). 1µl of the purified DNA was then quantified using the NanoDrop™ 8000 Spectrophotometer (Thermo Scientific). Fragment size was determined by gel electrophoresis; running 5µl of the purified DNA on a 1% agarose gel with 1Kb DNA ladder.

2.2.11.3 Chromatin capture using magnetic beads.

In order to capture chromatin on magnetic beads 20 µg sheared chromatin was incubated overnight with the appropriate antibody (see Table 3 of the appendix) Chromatin Antibody mix was then incubated with Dynabeads Protein G (invitrogen) for 2hrs.

2.2.11.4 Chromatin elution, cross-link reversal and protein degradation.

Immunoprecipitated samples were briefly spun to collect any liquid in the microcentrifuge tube lid. Dynabead-protein-DNA complexes were captured using a magnetic rack and washed with 1ml low salt, high salt, lithium Chloride and TE wash buffers. Immune complexes were then eluted by adding 100 µl of elution buffer containing 2.5µl (50 µg/ml) proteinase K to the magnetic Dynabead-protein–DNA complexes. These were mixed at 62°C for 2 h to release the protein bound DNA and reverse the cross-linking. The supernatant was then removed and transferred to a new tube using the magnetic racks. The protein–DNA complex mix was then incubated at 95°C for 10 min to denture the proteins and inactivate the proteinase K.

DNA was recovered using The *Wizard*® SV *Gel* and *PCR* Clean-Up System (Promega) and quantified using a NanoDrop™ 8000 Spectrophotometer (Thermo Scientific) followed by normalization with nuclease free H_20 to 1ng/µl.

2.2.12 Methylation analysis

2.2.12.1 Isolation of genomic DNA

Genomic DNA was isolated from cells using the QIAmp® DNA Mini Kit (Qiagen) following manufacturers guidelines. Briefly, cell pellet was resuspended in 200 µl PBS, to which 20 µl proteinase K was added and mixed by pipetting. 200 µl Buffer AL was then added to the sample and mixed by pulse-vortexing for 15 seconds. Followed by incubation at 56°C for 10 min, the 1.5 ml microcentrifuge tube containing the sample was then centrifuged briefly to remove drops from the inside of the lid. 200 µl ethanol (96–100%) was added to the sample and mixed again by pulse-vortexing for 15 s. After mixing, the 1.5 ml microcentrifuge tube containing the sample was once again centrifuged briefly to remove drops from the inside of the lid and the contents applied to the QIAamp Mini spin column (in a 2 ml collection tube) and centrifuged at 6000 x *g* for 1 min. The QIAamp Mini spin column was then placed in a clean 2 ml collection tube and the tube containing the filtrate discarded. 500 µl Buffer AW1 was then added to the QIAamp Mini spin column and centrifuged at 6000 x *g* for 1 min. The QIAamp Mini spin column was placed in a clean 2 ml collection tube and the collection tube containing the filtrate was discarded. 500 µl Buffer AW2 was added to the QIAamp Mini spin column and centrifuged at 20,000 x *g* for 3 min. The QIAamp Mini spin column was then placed in a clean 1.5 ml microcentrifuge tube, and the collection tube containing the filtrate was discarded. 200 µl Buffer AE was then added to the QIAamp Mini spin column and incubated at room temperature (15–25°C) for 1 min, and then centrifuged at 6000 x *g* (8000 rpm) for 1 min.

2.2.12.2 Isolation of methylated DNA

Genomic DNA isolated using QIAmp® DNA Mini Kit (Qiagen) was then fragmented by sonication using the Bioruptor Plus (Diagenode) set at 10 cycles of 30s 30 second pulses followed by 30 second pauses on high power. 300ng of fragmented gDNA was then incubated with CpG MethylQuest beads (GST-MBD fusion protein) and binding buffer at room temperature for 1hour with gentle mixing. Beads were then washed to remove DNA fragments not bound by the methyl domain binding protein (supernatant stored for subsequent analysis) methylated DNA was then eluted, both methylated DNA (eluted) and unmethylated DNA (supernatant) were analysed by PCR of the specific region, with primer sets identified in Table 1 of the appendix.

2.2.13 Adeno-associated virus (AAV) gene editing

2.2.13.1 Antibiotic death curve

The antibiotic resistance Neo^R gene driven by a thymidine kinase (TK) promoter was integrated into the viral genome which would be incorporated into the genome of SH-SY5Y cells via homologous recombination to isolate resistant cell colonies. It was necessary to determine the concentration of the antibiotic which killed wild-type SH-SY5Y cells; the death curve. The death curve determined the minimum concentration of the antibiotic used to culture potentially targeted cells and kill any cells which did not express the resistance gene and left targeted cells unaffected. G418 (Sigma) was dissolved to a working concentration of 100 mg/ml, which was subsequently diluted in DEMEM to give 11 concentrations ranging from 0.1 mg/ml to 1.5 mg/ml of G418. SH-SY5Y cells were diluted to 15,000 cells/ ml in DEMEM media, and 100 µl was aliquoted across a 96 well plate. Each row of the 96 well plate was supplemented with 100 µl of one of the 11 neomycin concentrations with the final row containing 100 µl antibiotic free media, this was repeated in triplicate and plates were incubated at 37^oC in 5% CO_2 for 10 days.

2.2.13.2 Generation of infectious AAV

2.2.13.2.1 Harvesting AAV from HEK-293 cells

72 hours post co-transfection both cells and media should contain live recombinant virus, the cell suspension media (containing virus) was harvested from the 10 T75 flask's, pooled and transferred to 50 ml conical tubes. The 10 T75 flasks were washed with 3 ml PBS to remove any residual media. The PBS wash (containing virus) was transferred into a separate 50 ml conical tube. Cells (containing virus) were then harvested by adding 2 ml trypsin to each T75 flask and incubated for 5 minutes; during the incubation stage, the three tubes containing the cell suspensions were centrifuged at 1000 rpm for 5 minutes. The trypsinized cells were then re-suspended in 2 ml DMEM, pooled and transferred to a 50 ml conical tube which was centrifuged for 5 minutes at 1000 rpm. The PBS wash solution was removed and discarded, leaving the cell pellet intact. The clarified supernatant from the initial centrifugation step was decanted into clean 50 ml conical tubes. The cell pellets and PBS wash pellet were re-suspended in 10 ml DMEM centrifuged for 5 minutes at 1000 rpm and decanted into a clean 50 ml conical tube, clarified supernatants and cell pellets were stored at -80^{o}C prior to purification.

2.2.13.3 AAV purification

To purify the virus from the lysate, the virions had to be released. This is achieved by repetitive freeze thaw cycles of the lysate. Viral supernatant and cell pellet were transferred to a 37^{o}C water bath. After 10 minutes, the cell pellet was thawed and re-suspended in 3 ml lysis buffer (Section 2.1.1.6). The tube containing the cell pellet was snap frozen in a dry ice ethanol bath for 10 minutes. Once frozen the tube was returned to the water bath for an additional 10 minutes, followed by vortexing to disrupt the cells; the freeze-thaw cycles were repeated twice. The cell suspensions and thawed supernatant were combined and aliquoted between 50 ml conical tubes, before being centrifuged at 2900 rpm for 30 minutes. After centrifugation, the supernatants were transferred to a sterile T75 and 1 µl Benzonase nuclease (Sigma) was added to each 10 ml viral supernatant (to remove any contaminating DNA in the supernatant). The mixture was then incubated for 30 minutes. The virus was purified using the AAV purification kit (VIRAPUR, Cat. No 003063). The kit was set up according to manufacturer's instructions; briefly the fibre-glass disc was washed with 4 ml PBS prior to use, then a vacuum pump was applied to the filter bottle. The flask with the Benzonase-treated virus was gently poured into the top of the filter and was allowed to be collected in the reservoir bottle below the filter. After filtration, 1 part dilution buffer 1 was added to 9 parts filtered supernatant. The purification filter assembly was attached to a 5 ml syringe filled with sterile PBS, and PBS was passed through the syringe. After the PBS was discarded, the 5 ml syringe was removed and the purification filter assembly was rearranged and attached to a 20 ml syringe. The supernatant was pulled through the filter assembly at rate of 20 ml/minute, discarding the supernatant when the syringe was filled. After the remaining supernatant had passed through, a new 20 ml syringe

was used to pass 30 ml of wash buffer 2 over the filter. Finally, 1.5 ml of elution buffer 3 was passed between two 5 ml syringes to elute it from the filter.

2.2.13.4 Virus quantification

The purified virus was quantified by quantitative real-time PCR (Section 2.2.10.5). To obtain linear rAAV DNA from the viral capsid a number of enzymatic steps were employed. To digest residual and contaminating DNA, 5 µl of virus was added to 5 µl DNaseI buffer (10x) and 10 U DNaseI (Promega). The solution was made up to 50 µl with nuclease free water and was incubated at 37^{o}C for 30 minutes, then at 95^{o}c for a further 2 minutes. To digest the viral capsid 100 µl of proteinase K solution (Qiagen), was added to each sample and incubated at 56^{o}C for 1 hour, then at 95^{o}c for a further 5 minutes to inactivate the proteinase K. This mix was then diluted 1:10 in nuclease free H_2O (an important step as undiluted samples failed to be amplified in qRT-PCR possibly due to the presence of residual inhibitory substances).

To ensure accurate quantification through RT-QPCR a standard curve was created using standards prepared by diluting rAAV-G4-CT (containing the Neo^{R} gene) in nuclease free H_2O in a series of 5-fold serial dilutions, ranging from 1 µg/ ml to 2x10^{-8} µg/ ml as shown in the table below.

Standard	µg/ml	copies neomycin^{R}/ml
1	1.000 e+0	2.821 e+11
2	2.000 e-1	5.642 e+10
3	4.000 e-2	1.128 e+10
4	8.000 e-3	2.257 e+09
5	1.600 e-3	4.514 e+08
6	3.200 e-4	9.027 e+07
7	6.400 e-5	1.805 e+07
8	1.280 e-5	3.611 e+06
9	2.560 e-6	7.222 e+05
10	5.120 e-7	1.444 e+05
11	1.024 e-7	2.889 e+04
12	2.048 e-8	5.777 e+03

Table 2.1 standard curve using dilutions of rAAV-G4-CT.

Obtaining a viral titre using RT-QPCR does not quantify infectious viral particles but quantifies copies of single stranded copies of neomycin^{R}. The calculation of single stranded copies neomycinR/ml makes the following assumption:

Number of single stranded copies of neomycinR in 1 μg double stranded rAAV = (2 x 6.02 x 1023) / (610x 7000 x 106) = 2.821 x 1011 where:

Number of copies of single stranded neomycinR gene per double stranded plasmid

6.02 x 10^{23} = number of molecules (copies) per mole of a substance

610 = average MW (g/mole) of one nucleotide bp

7000 = average size of a rAAV plasmid in bp

10^{6} = number of μg per g

For each reaction a 20 x PCR master mix was assembled:

0.225 μl forward primer (100 μM)
0.225 μl reverse primer (100 μM)
0.05l probe (100 μM)
0.75 μl H_20

For each reaction a final qRT-PCR mix was assembled:

1.25μl 20 x PCR master mix (above)
12.5μl GoTaq® qPCR Master Mix
10.25μl H20

Reactions were performed under the following cycling conditions:

Temperature	Time	Cycles
95°C	3 min	1
95°C	15 s	40
60°C	60 s	40

The data was analysed using BioRad CFX manager software, following the basic guidelines: Baseline and threshold values were set in the region of the exponential phase, with the replicates within 0.5 (threshold cycle) Cts. To obtain the standard curve (a plot of initial copies of single stranded neomycin^{R} gene/ml vs Ct), the following specifications were: i.) correlation coefficient $\geq$ 0.99; ii.) efficiency: 95-105%. To calculate the virus titre in genome copies (GC)/ml, the value of initial copies of single stranded neomycin^{R} gene/ml were multiplied by 300 (the dilution factor used in the initial preparation of the virus for qRT-PCR; DNaseI and proteinase K treatment).

2.2.13.5 Infection of HCT-116 cells with rAAV-G4KO-CT and rAAV-LoxP-CT

HCT-116 cells were seeded at a density of $5x10^5$ in 6x T75 flasks 24 hours prior to infection, to ensure cells were below 40% confluence at time of infection. A range of multiplicities of infection (MOI's) were used to identify the optimum ratio of viral particle to cell; 25, 50 and 100 µl of purified rAAV-G4KO-CT viral particles were added to 5 ml DMEM media in separate conical tubes, alongside this 25, 50 and 100 µl of purified rAAV-LoxP-CT viral particles were added to 5 ml media in separate conical tubes. Existing media was aspirated from the cells and was replaced with 5ml of virus-treated media; the cells were then incubated for 72 hours.

Infected cells were trypsinised, counted and plated at a range of cell densities (1000, 500, 250, 100, 50, 10 cells/ well) in 200µl selection media (0.5uM G418) across x6 96 well plates (1 plate per cell density) this was repeated for each of the three MOI's and for both targets (rAAV-G4KO-CT and rAAV-LoxP-CT) to give a total of x18, 96 well Cells were then returned to the incubator at 37 ^oC 5% CO_2 for 14 days.

2.2.13.6 Screen of infected cells to identify recombination events

2.2.13.6.1 Visual screen of 96 well plates to identify colony formation

After 14 days, the plates were visually screened to identify wells with successful colony formation, wells with 1 or more colonies were marked as potential positives and returned to the incubator for an addition 2-4 weeks to reach 50-70% confluence.

2.2.13.6.2 PCR screen of positive wells identified through visual screen.

Once infected cells reached 50-70% confluency they were lysed to release gDNA which was then PCR screened; to identify successful homologous recombination events. A positive and negative primer set was used to screen the gDNA. The positive set was specific for the targeted allele with the forward primer in the neomycin resistance cassette while the reverse primer was in located at the genomic region outside of the homology arms. Using this primer set only wells containing cells which had undergone homologous recombination of the neomycin selection cassette and therefore targeted sequence would give a PCR product. The negative primer set was wild-type specific. (Specific primers can be found in Table 1 of the appendix).

2.2.13.6.2.1 Release of genomic DNA (gDNA) from cells

To release the gDNA from the cells Direct PCR Lysis reagent (Viagen biotech Cat. No. 302-c) was used. The following reaction buffer was prepared per 96 well plate:

Reaction size	20 µl
Volume of cells	2 µl
Number of reactions	110
Direct PCR Buffer	2100 µl
Proteinase K	21 µl

20µl of Lysis buffer was aliquoted in to each well of a 96-well PCR plate. HCT-116 media (sectionwas removed, cells were then washed with PBS. Following aspiration of PBS, 25µl of Trypsin was added to each well, plates were then incubated at 37^{o}C for 10 minutes. Following incubation cell suspensions were mixed by gentle pipetting, 5µl of cell suspension was then transferred to the PCR plate containing 20 µl lysis buffer. To the remaining cells 200µl of media was added and 96 well plates were returned to the incubator.

The 96 well plates containing cell lysis suspension were sealed using a heat seal and incubated in the peqSTAR 2X Gradient Thermocycler (PeqLab) under the following conditions

Step	Cycles	Temp (^{o}C)	Incubation (min)
1	1	55	15
2	1	85	45

Lysed cells were transferred to 4^{o}C for short term storage until subsequent PCR Screen.

2.2.13.6.3 PCR screen setup

The 96-well plate containing direct PCR and control DNA was thawed on ice, the below master mix was then prepared on ice reaction size (µl) 20:

Volume of template 2 µl
number of reactions 100
buffer (5x) 400 µl
primer 1 (100µM) 15 µl
primer 2 (100µM) 15 µl
dNTP's (25mM each) 20 µl
MgCl2 (25mM) 120 µl
DMSO 120 µl
H2O, PCR grade 1090 µl
enzyme (5µ/µl) 20 µl

18µl of the PCR master mix was aliquoted across each well of the 96-well PCR plate, followed by 2µl of template DNA, the PCR plate was then sealed and with a heat sealer, briefly centrifuged to collect the content in the bottom of the well and transferred to the peqSTAR 2X Gradient Thermocycler (PeqLab) PCR machine, touchdown PCR cycling conditions are detailed bellow:

Step	Cycles	Temp (^{o}C)	Incubation (secs)
1	1	94	180
2	3	94	15
		64	30
		70	90
3	3	94	15
		61	30
		70	90
4	3	94	15
		58	30
		70	90
5	35	94	15
		57	30
		70	90
6	1	70	300
7		4	Hold

The plate was then removed from the PCR machine and stored at 4^{o}C. PCR analysis was performed using Gel electrophoresis, loading 10 µl of PCR product and 1.8 µl 6X loading dye into a 1% agarose gel.

2.2.13.6.4 Expansion of positive targeted pools

The initial PCR screen only identified pools of positive clones therefore it was necessary to isolate single cells from this pool to ensure a clonal population giving rise to a truly isogenic cell line. Once pools were identified as positive for targeting, the cells within that well were expanded to a 48-well plate. Expansion typically occurred 1-4 days post-harvest for the PCR screen, once the 24 well plate reached 70% confluence each well was expanded across 24 well plates ensuring the cells were robust enough to undergo single cell dilution (SCD).

2.2.13.6.5 Single cell dilution of targeted pools

Each of the 9x positive pools identified in the PCR screen were diluted across 3x 96 well plates 1x plate at a concentration of 1 cell/well x1 plate at 0.5 cells/well, and x1 plate at 3 cells/well, plates were transferred to an incubator at 37 ^{o}C 5% CO_2 for 14 days.

2.2.13.6.6 Visual screen of SCD cells to identify clonal cell colony formation

After 14 days incubation at 37 ^{o}C, 5% CO_2 SCD plates were visually screened for clear single colonies, these were marked by circling the well on the lid and record in a spreadsheet. When colonies reached 70% confluence, if there were greater than 20 colonies per plate, they were harvested for a PCR screen. Plates with less than 20 colonies per plate were consolidated onto a new 96-well plate and allowed to grow to 50-70% confluency before repeating the PCR screen.

2.2.14 CRISPR-Cas targeting efficiency – T7 endonuclease 1 (T7E1) assay

PCR and T7 endonuclease assay was used to determine efficiency of CRISPR-Cas targeting. CRISPR-Cas targeting introduces double strand breaks at or around the target region (Mir-137), the natural cellular process of NHEJ will repair these double strand breaks leaving bp errors, these can be detected by the mismatch sensitive enzyme T7E1. PCR amplicons of the target region Mir-137 (primer set can be found in Table 1 of the appendix) were isolated from the agarose gel using Wizard® SV Gel and PCR Clean-Up System (Section 2.2.1.5) and resuspended in 100 µl of nuclease free H_2O. the samples were heated at 95 °C for 5 minutes and allowed to cool to room temperature (~60 min). DNA annealed to form heteroduplex DNA, which was treated with 5 units of mismatch-sensitive T7 endonuclease 1 (NEB) for 20 mins at 37°C and then analyzed by 2% agarose gel electrophoresis.

Chapter 3

Repeat sequences in the promoter of selected genes have the potential to form the secondary DNA Structure G-quadruplex

3.1 Introduction

To date the vast majority of work surrounding G-quadruplex and its potential functionality in gene regulation has been in relation to cancer (Figure 3.15)

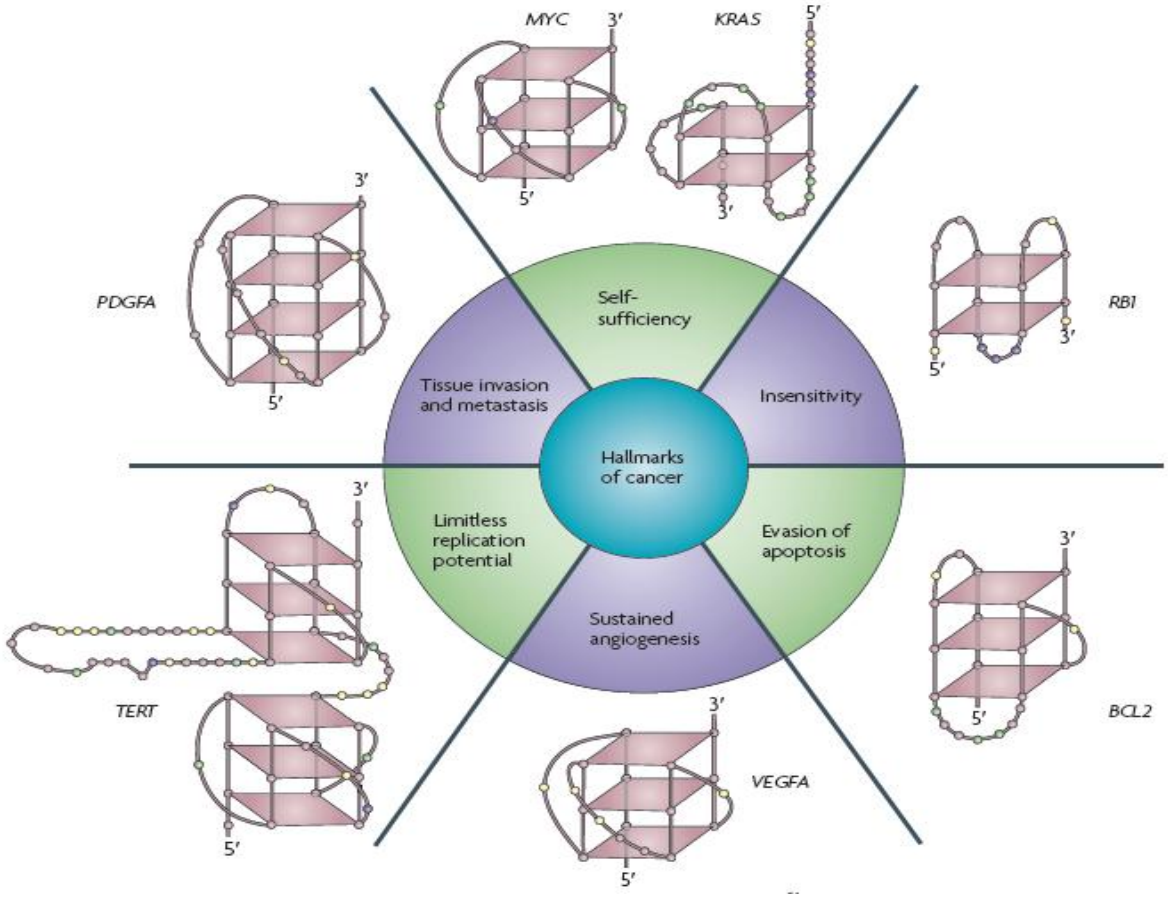


Figure 3.1 G-Quadruplex present in genes implicated in cancer

A schematic diagram showing the six hallmarks of cancer as defined in (Hanahan and Weinberg, 2000) correlated to a specific G-Quadruplex element in the promoter of a Cancer gene (Figure adapted from (Brooks and Hurley, 2009).

The G-quadruplex located within the CT element of c-MYC is the most well studied G-quadruplex as described in Section 1.3.1. Given the prevalence of c-MYC in so many cancers as outlined in Section 1.3.1.1 the ability to selectively target its aberrant expression would be an incredibly attractive intervention. If the G-quadruplex elements do indeed function as on off switches and are amenable to small molecule targeting as detailed in Section 1.2.3.1 then their therapeutic potential should not be limited to Cancer. Therapeutic G-quadruplex targeting could be extrapolated to genes involved in other diseases which would benefit from a switch to control the rate of transcription. In this chapter I expand the role of G-quadruplex from its implication in cancer into the field of neuroscience and explore the mechanism through which G-quadruplex may function at to regulate selected genes associated with neurological disorders.

3.2 Aims

1. Address the potential for repeat elements in the promoter of selected genes to form G-quadruplex using both experimental and in-silico analysis.
    i. ssDNA mapping
    ii. Potential G-quadruplex sequence analysis: Quadparser programme
2. Determine whether ssDNA binding protein hnRNP K and CTCF, both regulators of chromatin architecture and transcription bind to the promoter regions of selected genes.
3. Explore G-quadruplex associated transcription factor binding and methylation status at the promoters of selected genes.

3.3 Results

3.3.1 Mapping repeat elements with the potential to form G-quadruplex in the promoters of selected genes associated with cancer and neurological disorders

In order to begin exploring the hypothesis that repetitive sequences of DNA within 1500bp of transcriptional start sites can form secondary DNA structures termed G-quadruplex I performed both experimental and in-silico bioinformatic analysis to identify such regions of DNA. The following search criteria were used:

i. Regions within 1500bp of the transcriptional start site.

ii. Regions of high ssDNA

iii. Sequences with the potential to form G-quadruplex

As detailed in Section 1.2.2.2 although DNA is a highly dynamic molecule its overall conformation is stable in duplex form despite continuous transient transactions such as transcription and DNA replication. Duplex DNA's conformational stability is a result of its hydrophobic and electrostatic base-stacking and base-pairing (Yakovchuk et al., 2006). It is therefore energetically favourable for the molecule to maintain a duplex form and will only alter its structural conformation if significant mechanical force is applied to it. Transcription induced dynamic supercoiling is one such force and is able to spread ~1.5 kilobases upstream of the start sites of active genes (Kouzine et al., 2013a). This data provides the 1st search criteria; analyse sequences located within 1500bp of the transcriptional start site as these are susceptible to transcription induced dynamic supercoiling.

For duplex DNA to adopt alternative conformations including G-quadruplex, the B-DNA must first undergo unwinding (negative supercoiling) as detailed in Section 1.2.2.2. This unwinding of the double helix forces the DNA to become single stranded in part. The ssDNA protocol described in (Kouzine et al., 2013b)

provides a unique method to capture these regions and is therefore a very informative tool when analysing regions of DNA which may form G-quadruplex. *In vivo* ssDNA-seq data generated by the Levens Lab (Kouzine et al., 2013b) has been used to generate a ssDNA mapping track on UCSC genome browser which identifies regions of the Raji cell genome that adopt ssDNA. This ssDNA mapping track forms the basis of the second search criteria; regions which have high areas of ssDNA are likely to form G-quadruplex. It should be noted that the ssDNA mapping assay has a bias against G-quadruplex as the potassium permanganate (KMnO4) treatment modifies ssDNA by oxidizing pyrimidine bases with preference for thymidine so although the assay will locate ssDNA guanine (G) rich areas will not peak as highly as the surrounding thymine (T) rich areas.

The potential for a sequence of DNA to form G-quadruplex is determined by the arrangement of G bases, specifically x3 G's, repeated four times with between 1 and 7 residues between each run of GGG as defined by the 'Quadparser' algorithm: (G3+ N1-7 G3+ N1-7 G3+ N1-7 G3). The Quadparser algorithm was used to create a potential G-quadruplex mapping track on UCSC by Dr. Ashutosh Gupta a colleague in the Levens Lab. This track highlights any regions which have a sequence which falls into the range G3+ N1-7 G3+ N1-7 G3+ N1-7 G3 and thus forms the 3rd search criteria.

The Levens Lab's focus is on cancer biology (where most of the literature on G-quadruplex has previously been focused) and the Quinn Lab's primary research interest is in neurological disorders. However both labs are interested in the underlying mechanisms of transcriptional regulation common to both sets of disease. With this in mind a number of cancer and neuroscience genes were screened using University California Santa Cruz (UCSC) genome browser and the tracks outlined

above to identify 3 genes associated with cancer and 3 genes associated with neurological disorders that met the search criteria previously specified. Figure 3.2 a – g. details C-MYC, hTERT and VEGFA as the selected cancer target genes, having been previously identified as containing potential G-quadruplex forming regions (Brooks and Hurley, 2009). 5-HTT, MAOA and DAT1 are the focus of ongoing studies in the Quinn Lab and were selected as the neurolocical target genes. These genes have no literature describing them as having potential G-quadruplex sequence but my analysis has identified repetitive regions of DNA overlapping potential G-quadruplex (pG4) sequence within 1.5 kb of the transcriptional start site within regions of ssDNA**.**

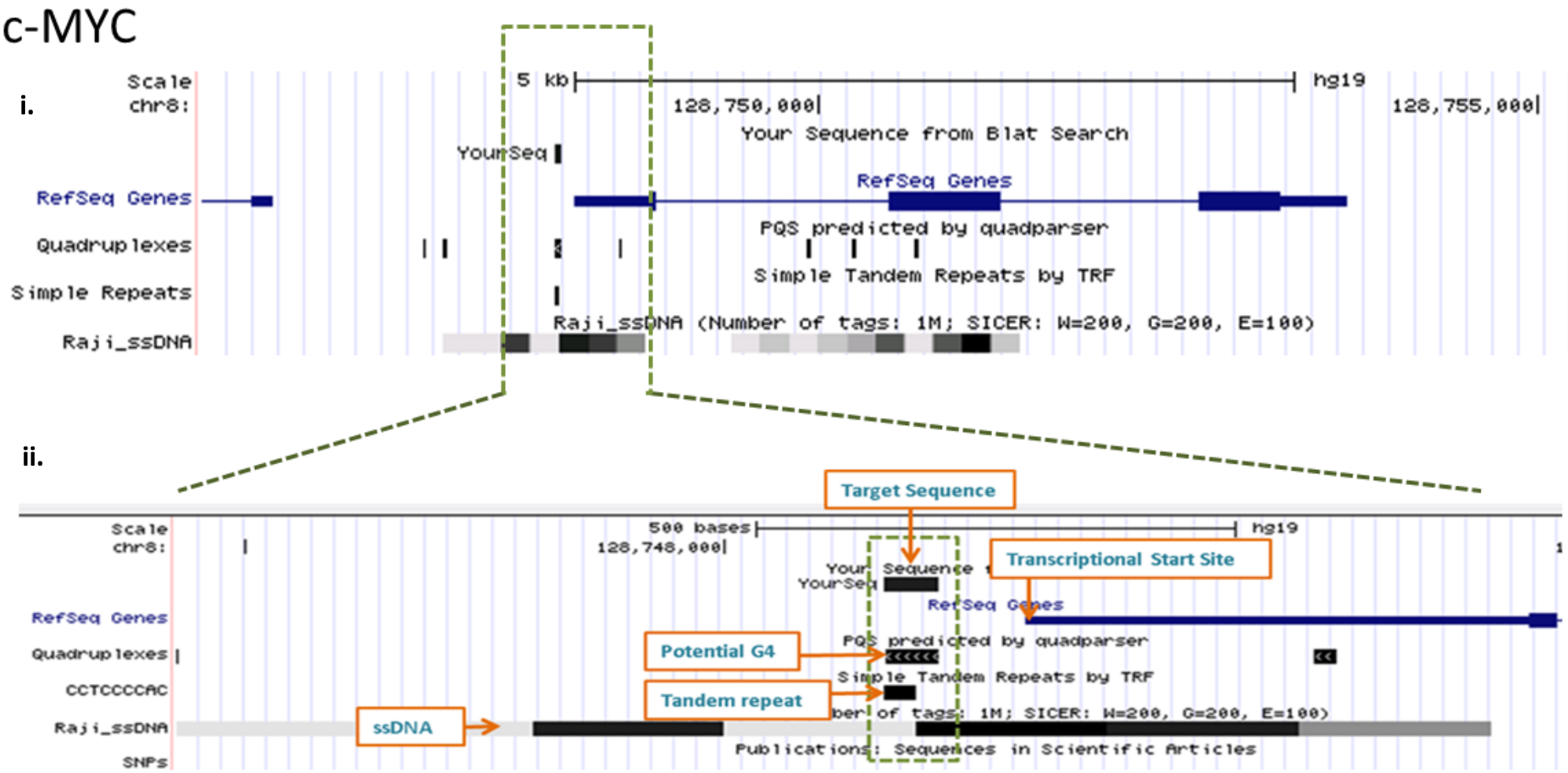


Figure 3.2a c-MYC UCSC in-silico analysis

i: UCSC genome browser screenshot identifying c-MYC gene and the distribution of pG4 sites, repetitive sequences and ssDNA across the entire gene.

ii. A magnified view of the target sequence selected due to its repetitive nature, pG4 and location within 1.5kb of the transcriptional start site

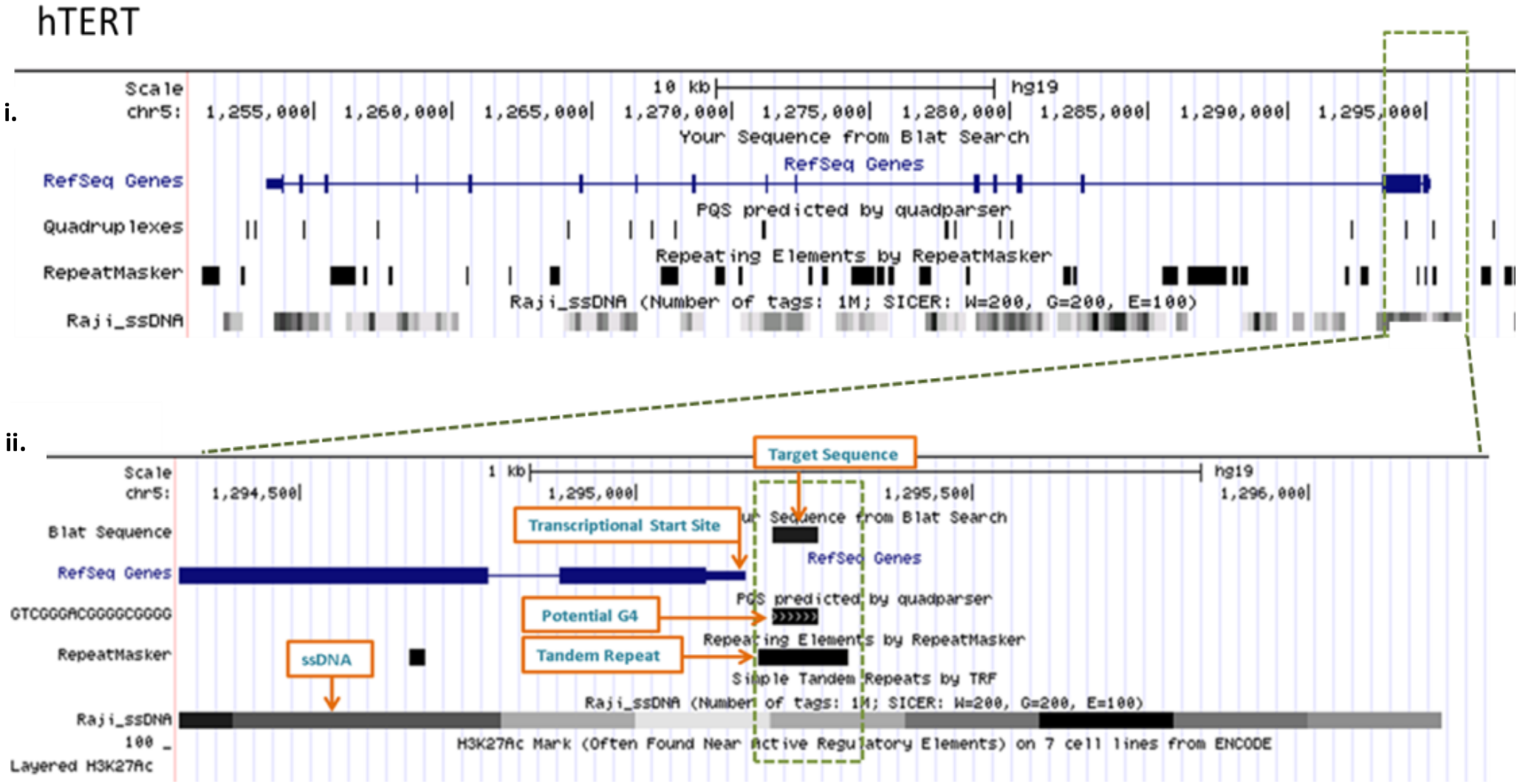


Figure 3.2b hTERT UCSC in-silico analysis

i. UCSC genome browser screenshot identifying hTERT gene and the distribution of pG4 sites, repetitive sequences and ssDNA across the entire gene.

ii. A magnified view of the target sequence selected due to its repetitive nature, pG4 and location within 1.5kb of the transcriptional start site

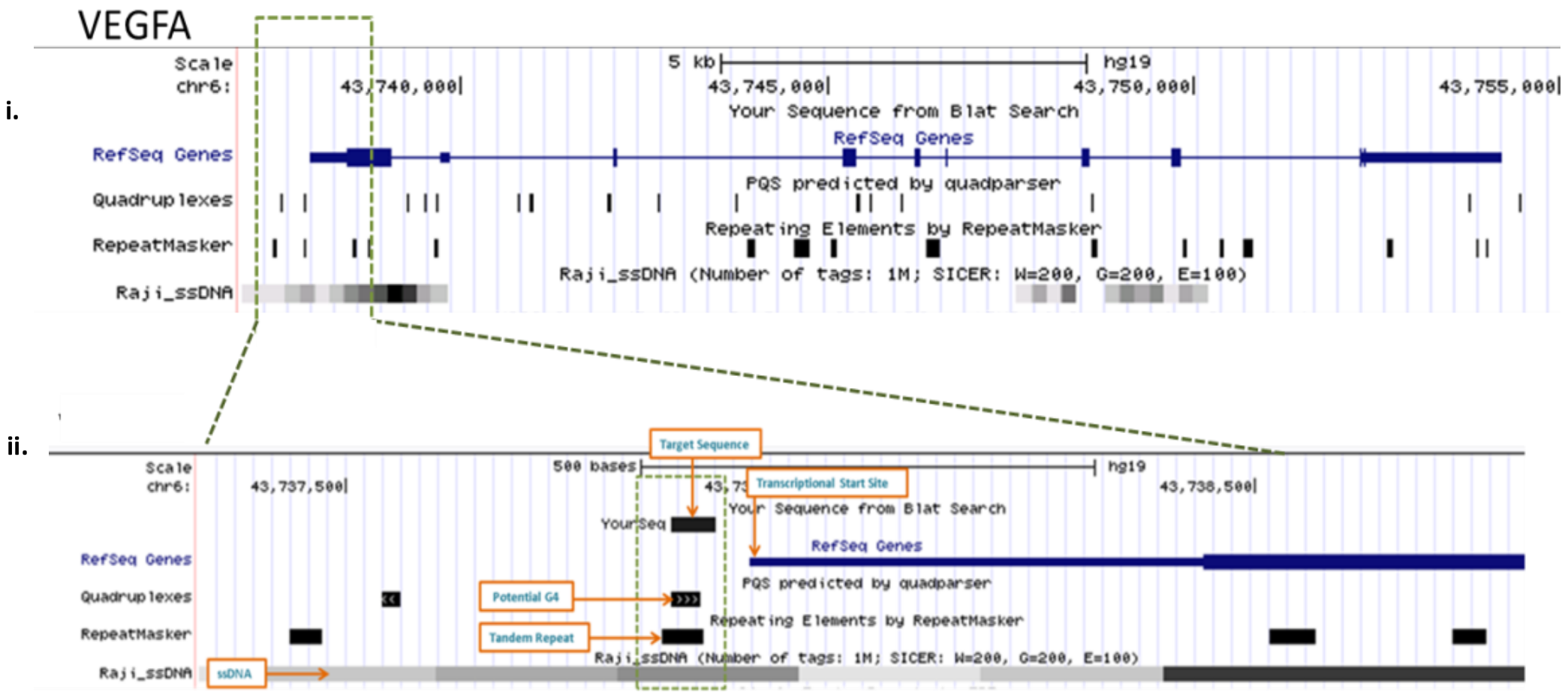


Figure 3.2c VEGFA UCSC in-silico analysis

i. UCSC genome browser screenshot identifying VEGFA gene and the distribution of pG4 sites, repetitive sequences and ssDNA across the entire gene.

ii. A magnified view of the target sequence selected due to its repetitive nature, pG4 and location within 1.5kb of the transcriptional start site

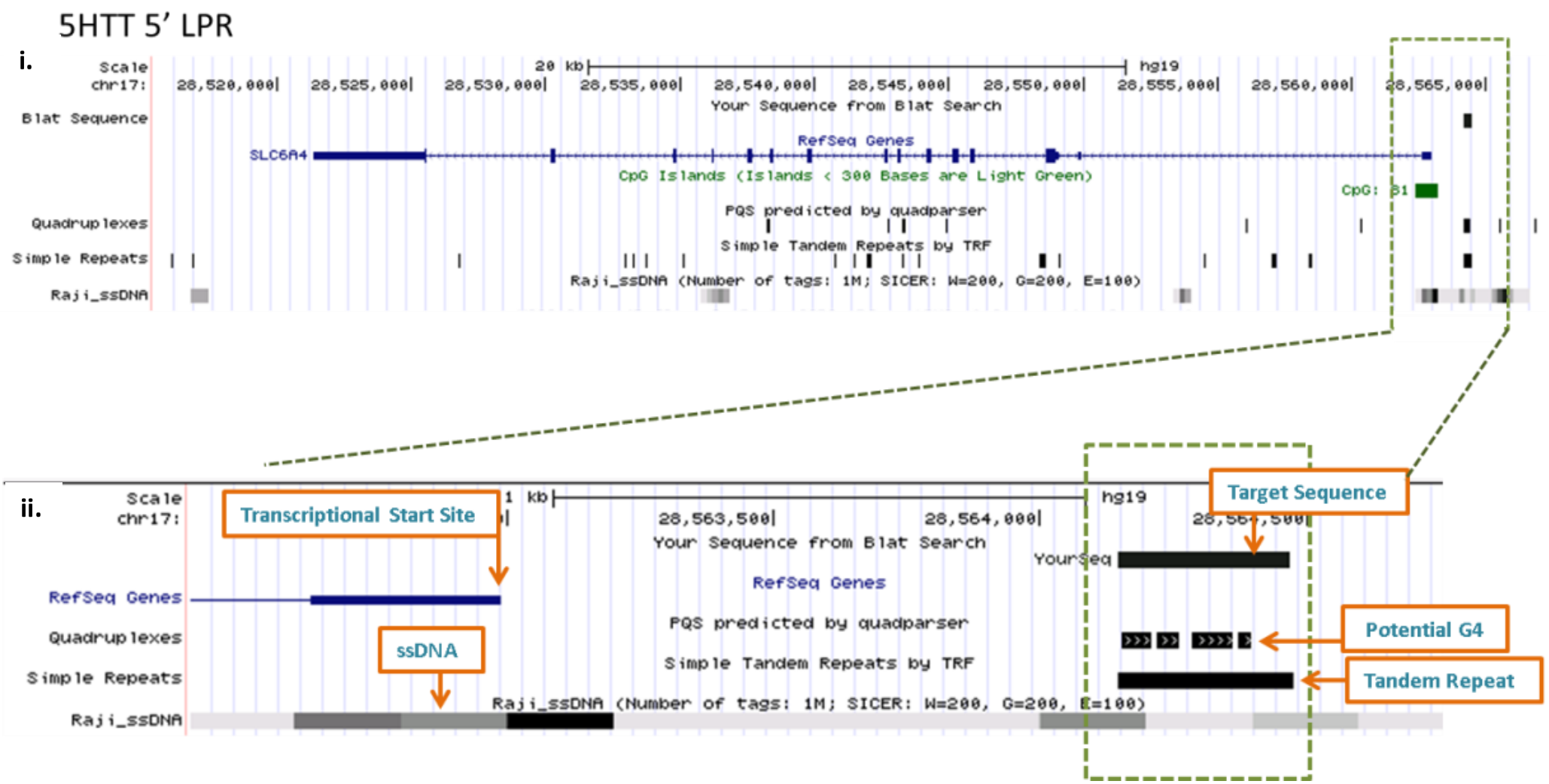


Figure 3.2d 5-HTT 5'LPR UCSC in-silico analysis

i. UCSC genome browser screenshot identifying 5-HTTgene and the distribution of pG4 sites, repetitive sequences and ssDNA across the entire gene.

ii. A magnified view of the target sequence selected due to its repetitive nature, pG4 and location within 1.5kb of the transcriptional start site

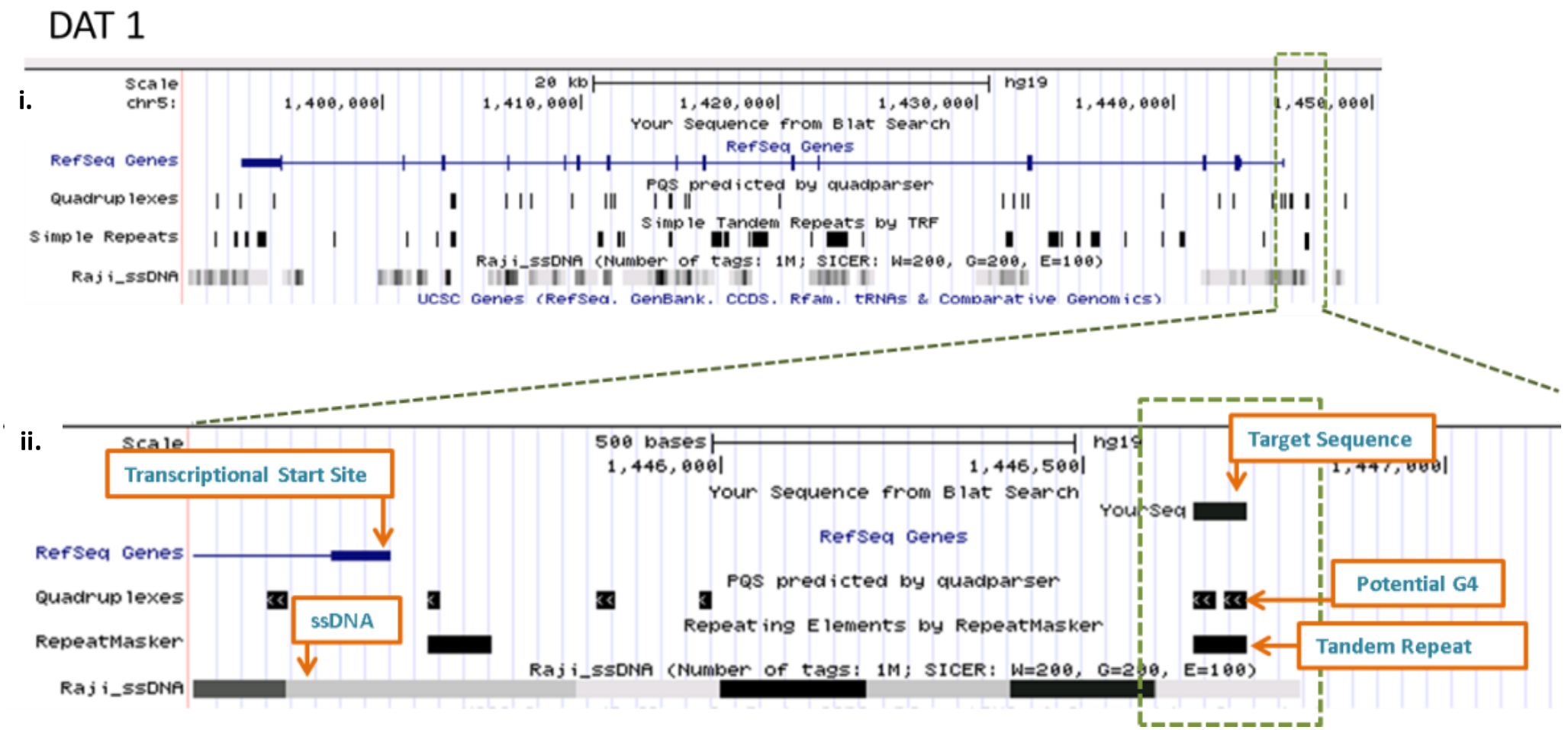


Figure 3.2e DAT1 UCSC in-silico analysis

i. UCSC genome browser screenshot identifying DAT1 gene and the distribution of pG4 sites, repetitive sequences and ssDNA across the entire gene.

ii. A magnified view of the target sequence selected due to its repetitive nature, pG4 and location within 1.5kb of the transcriptional start site

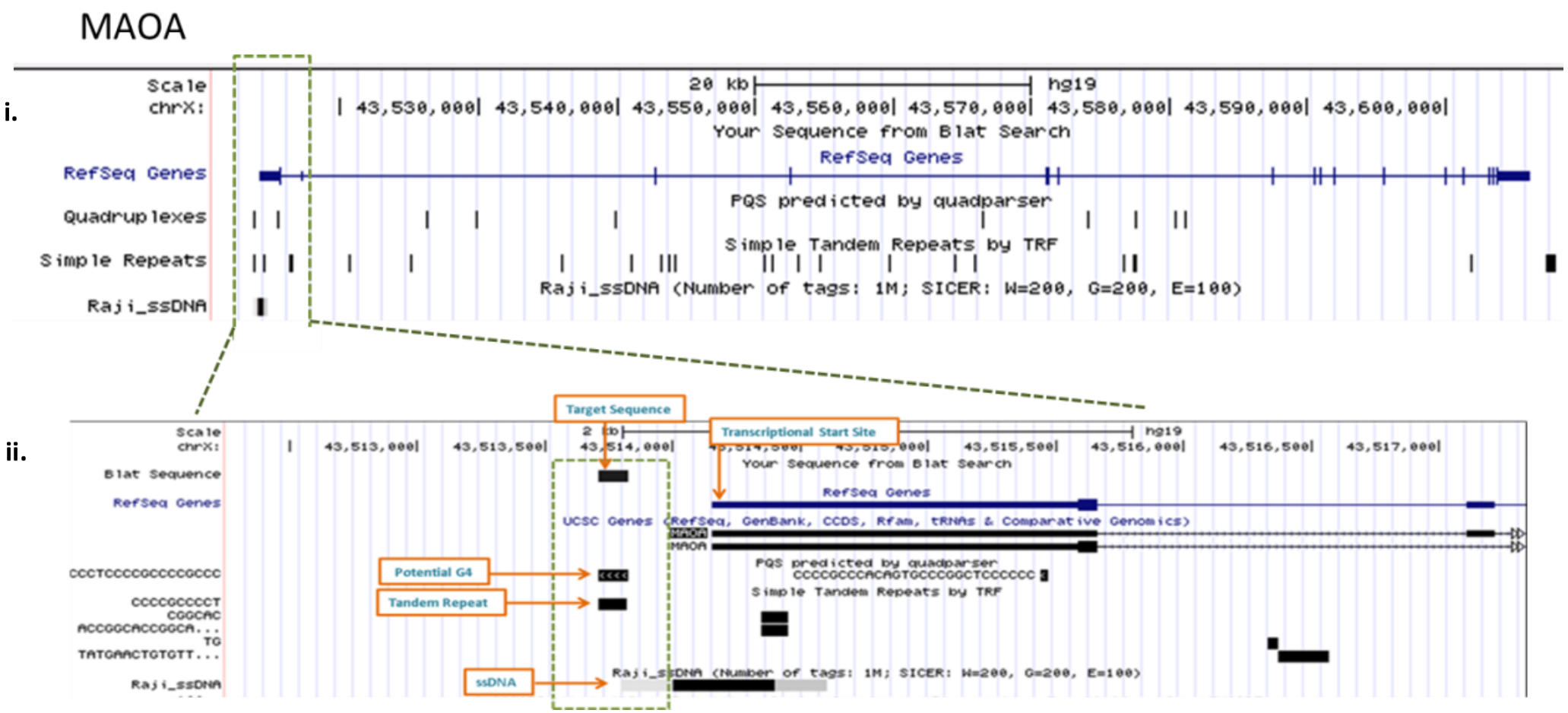


Figure 3.2f MAOA UCSC in-silico analysis

i. UCSC genome browser screenshot identifying MAOA gene and the distribution of pG4 sites, repetitive sequences and ssDNA across the entire gene.

ii. A magnified view of the target sequence selected due to its repetitive nature, pG4 and location within 1.5kb of the transcriptional start site

3.3.2 hnRNP K and CTCF bind to the promoters of c-MYC and 5-HTT.

The ssDNA binding protein and transactivator hnRNP K is known to bind to the CT element and regulate c-MYC (Tomonaga and Levens, 1995) (Tomonaga and Levens, 1996, Michelotti et al., 1996a). The regulator of chromatin architecture & transcription; CTCF is known to bind to the CT element in the chicken c-MYC gene (Lobanenkov et al., 1990, Klenova et al., 1993) and has been shown to bind to the 5-HTT 5' LPR and influence expression of 5-HTT (Ali et al., 2010, Vasiliou et al., 2012). I decided to validate hnRNP K binding at the CT element in the human c-MYC gene and examine its binding at the 5-HTT 5' LPR in HCT-116 cells through ChIP. ChIP analysis of the CT element was performed under normal growth conditions. ChIP analysis of the 5-HTT 5' LPR was performed under normal growth conditions and following 1 hour treatment with 10μM cocaine. Cocaine is known to effect the monoamine pathway and has previously been shown to influence CTCF binding at the long and short alleles of the LPR VNTR at this concentration and time course (Ali et al., 2010). In HCT-116 cells, hnRNP K and CTCF bind to the CT element (Figure 3.3). hnRNP K and CTCF bind to the LPR region under normal growth conditions with binding of CTCF increasing in response to cocaine (1hr), (Figure 3.4). This preliminary biochemical analysis implicated hnRNP K and CTCF in the regulation of two of the target regions. Enabling me to move forward and begin to address the mechanism that could correlate with potential alternative DNA structure at these regions.

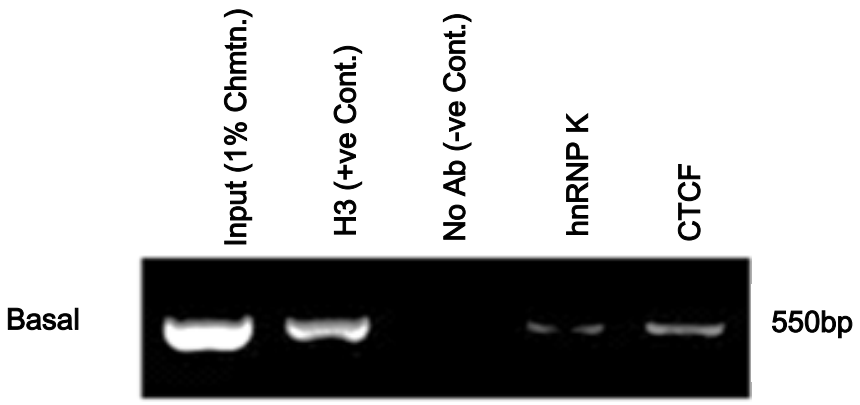


Figure 3.3. ChIP analyses of the CT element in c-MYC HCT-116 cells.

Agarose gel showing PCR products amplified following ChIP identifies hnRNP K and CTCF binding to the CT element c-MYC under normal growth conditions

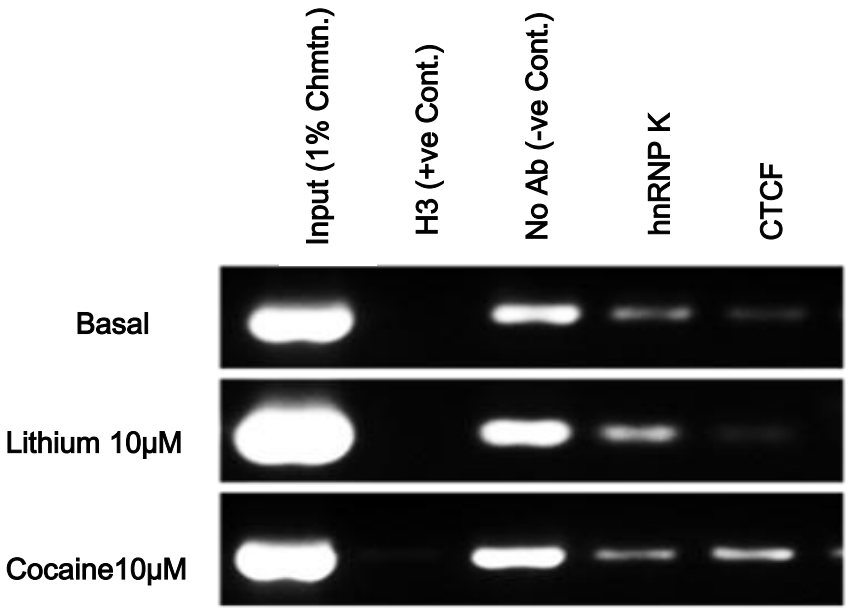


Figure 3.4. ChIP analyses over the LPR region, in HCT-116 cells.

Agarose gel showing 550bp PCR products amplified following ChIP identifies hnRNP K and CTCF binding to the LPR with CTCF binding appearing to increase in response to cocaine (1hr).

3.3.3 Allelic specific interactions of hnRNP K and CTCF at the LPR region

Owing to the dynamic nature of G-quadruplex formation discussed in section 1.3.1.2 along with the limitations of ChIP discussed in Section 4.4 it is difficult to gain a complete and accurate picture of transcription factor binding activity at the regions of interest. ChIP will capture a single frame across a population of cells, all of which may be at different points in the process of G-guadruplex formation. I therefore identified the JAr cell line which is heterozygous for the 14 copy (short) and 16 copy (long) LPR VNTR. Based on my Quadparser analysis (Figure 3.16) of this region the sequence of the short allele has the potential to form two G-quadruplex structures whereas the long allele has the potential to form four G-quadruplex structures (Figure 3.2e). This heterozygosity enabled comparison of factor binding between the two alleles within the same assay. Under normal growth conditions hnRNP K and CTCF both interact with the short allele where as in response to 1 hr exposure to 10μM cocaine both transcription factors are seen to interact with the long and short allele, with CTCF preferentially binding to the long allele (Figure 3.5).

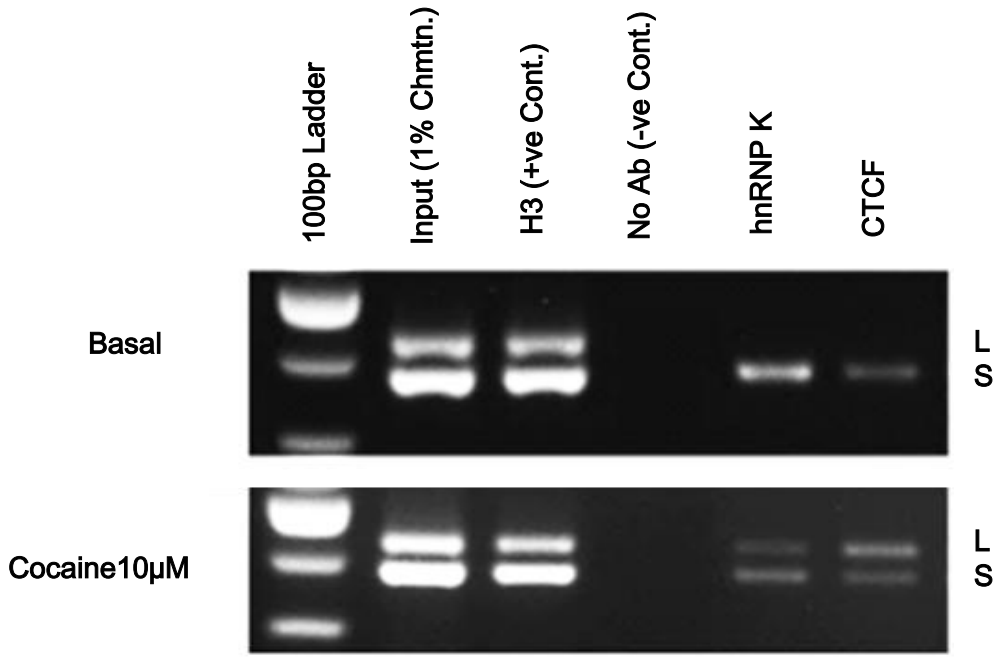
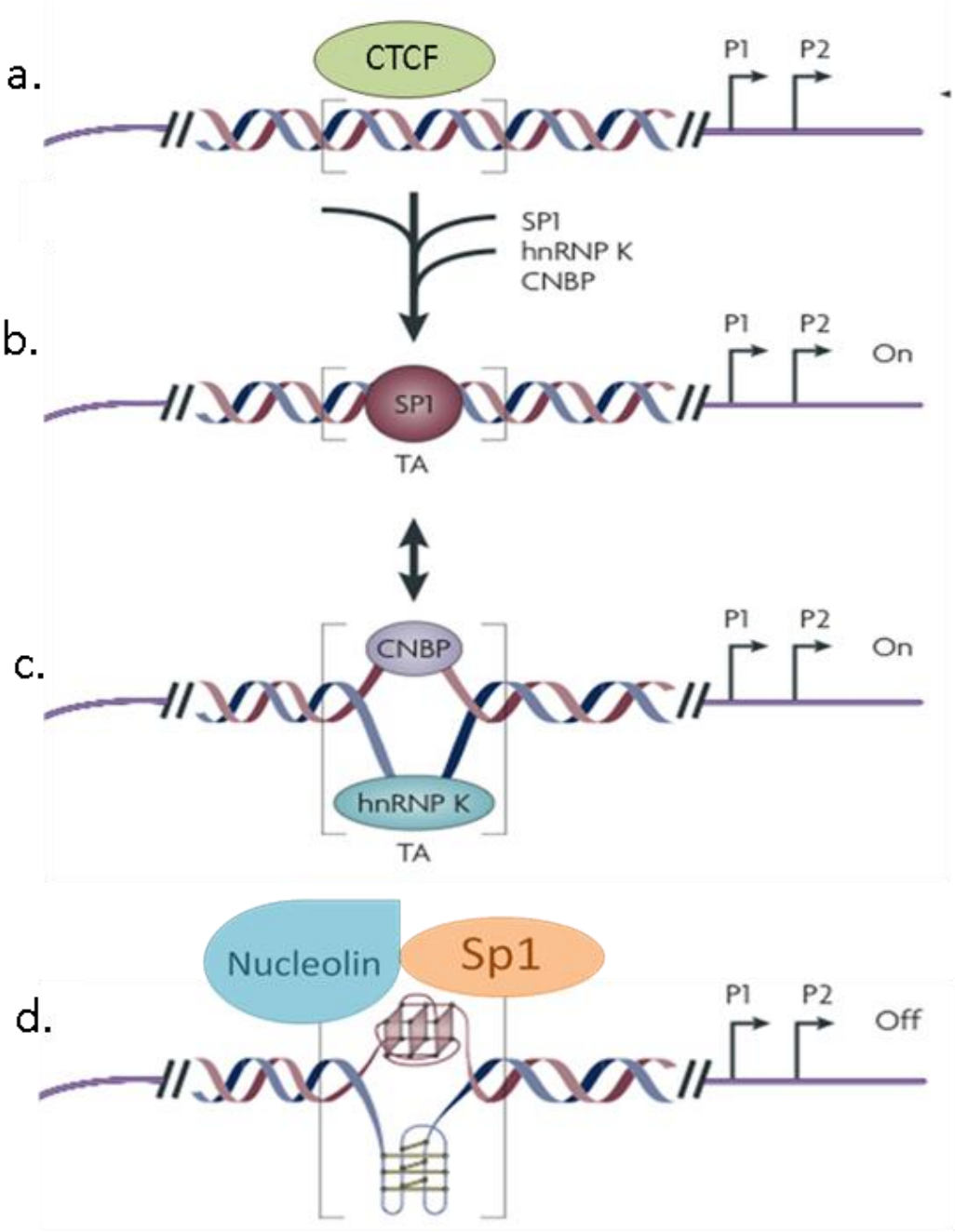


Figure 3.5 JAr cell line (heterozygous for long and short LPR) identifies allelic specific interactions at the region

Agarose gel showing PCR products amplified following ChIP of the 5HTT gene promoter. The LPR region is identified as two bands in 'input 1% chromatin' and confirmed by positive control Histone 3 (H3). hnRNP K and CTCF bind only to the short allele under basal conditions, in response to 10μM Cocaine (1hr) CTCF and hnRNP K bind to the long allele with CTCF showing preferential binding to the long over the short.

3.3.4 A model to explain the potential biophysical mechanism though which the CT element is able to form G-quadruplex and regulate c-MYC.

As previously mentioned in Section 1.3.1.2 the CT element has been shown to adopt non-B DNA structures in supercoiled DNA *in vitro* and in its endogenous location *in vivo* (Michelotti et al., 1996b, Kohwi and Kohwi-Shigematsu, 1991). When in B DNA form, CTCF can bind to the CT element (Klenova et al., 1993, Lobanenkov et al., 1990). The transcriptional activator Sp1 is also able to bind to the CT element in turn activating transcription (Hay et al., 1987, Geltinger et al., 1996). This activated transcription induces supercoiling which could accumulate and force the element into the single-stranded conformation. This new single stranded transcriptional landscape would enable alternative single strand specific transcriptions factor hnRNP K and CCHC-type zinc finger nucleic acid binding protein (CNBP) to bind the purine-rich and pyrimidine-rich strands, respectively, maintaining the active state (Brooks and Hurley, 2009, Tomonaga and Levens, 1996, Michelotti et al., 1995). Upon removal of the single stranded transcription factors, the G rich CT element is able to adopt stable non-B DNA structures; a G-quadruplex on the purine-rich strand and i-motif on the pyrimidine-rich strand (Sun and Hurley, 2009). These globular structures create a third transcriptional landscape across the same region of DNA by masking the standard B-DNA transcription factor binding sites within the structure. This potentially creates a new G-quadruplex specific transcription factor binding site to which nucleolin can bind (Gonzalez and Hurley, 2010b) and therefore silence transcription by stabilizing the G-quadruplex. It has recently be recently proposed that Sp1 itself can bind to the G-quadruplex (Raiber et al., 2012) with unknown effects (Figure 3.6).



3.6 A model to explain the potential biophysical mechanism though which the CT element is able to form G-quadruplex and regulate c-MYC

a. A duplex representation of the promoter with CTCF binding and silencing b. Binding of Sp1 to the duplex structure activates transcription leading to negative supercoiling. c. Binding of hnRNP K and CNBP to the single stranded C and G rich regions, respectively, could lead to the activation of c-MYC. d. The repression of c-MYC transcription when hnNRP K and CNBP are not bound leads to the formation of the G-quadruplex and i-motif, which create a new transcriptional landscape enabling G-quadruplex factors to bind.

3.3.5 Transcription factor binding at the CT element of c-MYC

The model in Section 3.3.3 which details the biophysical mechanism though which the CT element is able to form G-quadruplex implicates CTCF, Sp1, hnRNP K, CNBP and nucleolin in the process. I therefore performed further ChIP analysis for these factors across the CT element in the HCT-116 cell line. Under normal growth conditions hnRNP K, Nucleolin, Sp1 and the positive histone mark H3k9me3 bind at the CT element however CTCF does not appear to bind.

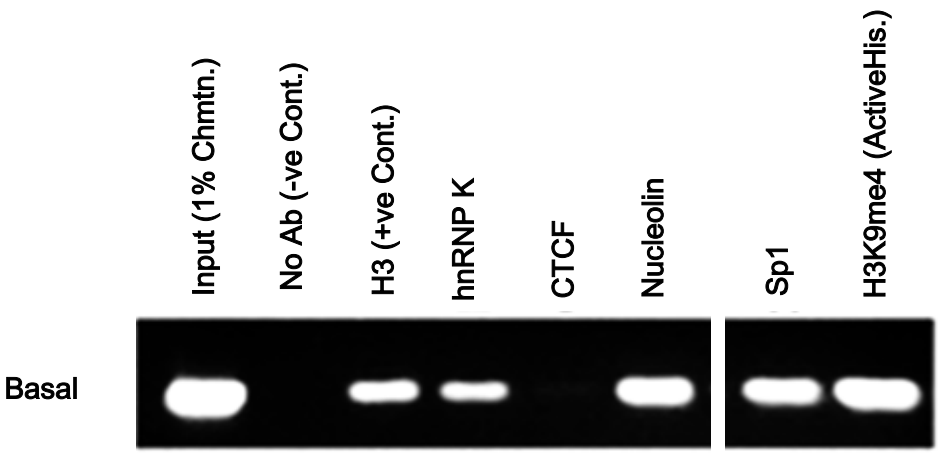


Figure 3.7 Transcription factor binding at the CT-element of C-Myc.

Agarose gel showing PCR products of (550bp) amplified following ChIP. Under basal conditions hnRNP K, Nucleolin, Sp1 and the positive histone mark H3k9me3 bind at the CT element, CTCF does not.

3.3.6 Transcription factors implicated in G-quadruplex formation bind in an allele specific manner at the 5'LPR of 5-HTT gene.

The ChIP data in Section 3.3.4 showed many of the factors in the G-quadruplex formation model (Section 3.4.3) binding to the CT region. Based on these findings I wanted to explore whether the same model may be relevant at other previously unidentified potential G-quadruplex sites. I therefore performed ChIP analysis for the same factors across the 5' LPR in the JAr cell line. This was in the hope that transcription factor binding profiles across this region may point to a common, dynamic structural regulatory mechanism. Such a mechanism could then be extrapolated to a number of promoter regions with potential G-quadruplex sequences within 1.5kb of their transcriptional start site.

In addition to normal growth conditions (basal) and 10μM cocaine (1hr), analysis was performed in response to 1mM Lithium (1hr). Lithium is a mood stabilising drug previously shown to regulate the 5-HTT gene in part, by repressing the binding of CTCF to the alleles of the 5' LPR (Ali et al., 2010). 1mM concentration was used during this study as this concentration is known to be clinically relevant (Wang et al., 2011), and further at higher concentrations was found to have non-specific metabolic effects (Roberts et al., 2007).

Under basal conditions CNBP, hnRNP K and CTCF were observed to bind at both alleles but preferentially to the short allele under the same conditions. Nucleolin and Sp1 bound only to the short allele. In response to 1mM Lithium; CNBP and hnRNP K bound preferentially to the short allele, CTCF binding was lost at the long allele and nucleolin and Sp1 remained bound to the short allele. After 1hr treatment with 10μM cocaine, binding of CNBP and hnRNP K was lost at the long allele. CTCF was observed to bind equally to both alleles and nucleolin and Sp1 were both

seen to bind preferentially at the short allele but cocaine resulted in binding of nucleolin and Sp1 at the long allele.

In essence a switch is observed at the long allele in response to cocaine: Single stranded binding proteins CNBP and hnRNP K bind to the long allele and G-quadruplex binding proteins do not bind to the long allele under normal growth conditions and in response to lithium. However in response to cocaine the reverse is observed; binding of CNBP and hnRNPK is lost and binding of nucleolin and Sp1 is gained.

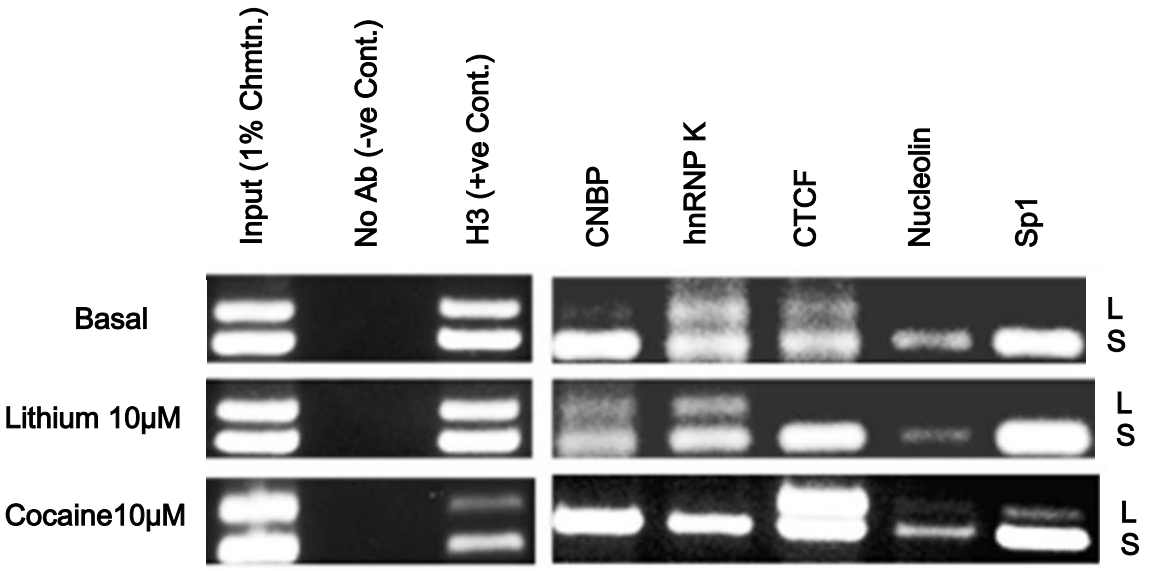


Figure 3.8 G-quadruplex associacted transcription factors bind in an allele specific manner at the 5' LPR.

Agarose gel showing PCR products of the 5'LPR amplified following ChIP analysis in the JAr cell line for CNBP, hnRNPK, CTCF, nucleolin and Sp1, identifies allelic specific interactions at the 5' LPR region identified as two bands in 'input 1% chromatin' top band is the L allele , S allele is the bottom band.

3.3.7 Transcription factors bind in an allele specific manner at the MAOA promoter.

Given the interesting and informative allele specific binding patterns across the 5-HTT 5' LPR in Section 3.3.6 I wanted to interrogate binding across another polymorphic region in a cell line heterozygous for the region. The neuroblastoma cell line SH-SY5Y is both female and heterozygous for the 3 and 4 copies of the proximal μVNTR (Figure 3.9b) and 9 and 10 copies of the distal dVNTR (Figure 3.9c) in the promoter of the MAOA gene found on the X chromosome. gDNA extracted from SH-SY5Y (Section 2.2.12.1) was amplified by PCR using primer sets MAOA distal (Table 1 of the appendix) and MAOA proximal (Table 1 of the appendix) using the PCR protocol detailed in Section 2.2.1.2. When amplifying repetitive sequence as in the case of VNTR's a PCR amplification bias is often observed between the two alleles (Ali et al., 2010, Vasiliou et al., 2012).

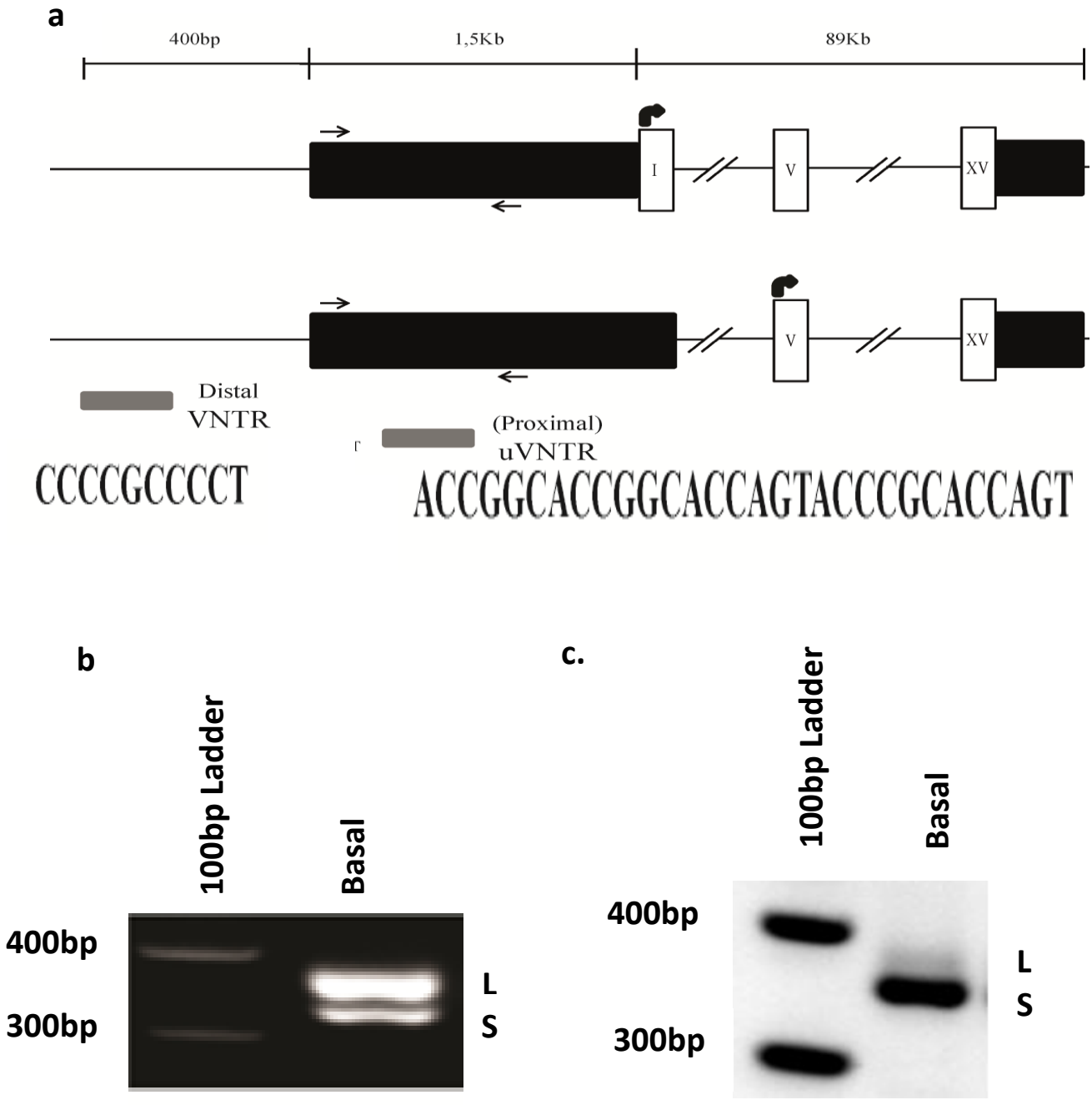
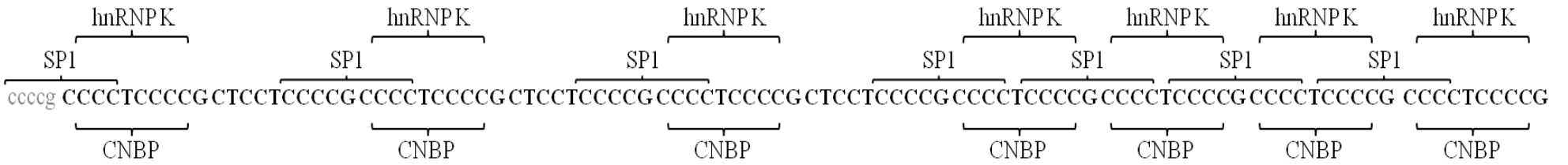


Figure 3.9 MAOA promoter contains a proximal (μ) VNTR and a distal (d) VNTR.

(a) Schematic of MAOA VNTR repeats along with 2 transcriptional start sites (a). (b) Agarose gel showing PCR amplified μVNTR from gDNA extracted from SH-SY5Y cells, heterozygous for 3 and 4 copies of the repeat. (c) Agarose gel of PCR amplified dVNTR from gDNA extracted from SH-SY5Y cells, heterozygous for 9 and 10 copies of the repeat.

It is not possible to differentiate between the μVNTR and the dVNTR in ChIP analysis due to the limited sequence of 400bp between the two regions. The chromatin shearing range is 500 – 1500 bp therefore fragments undergoing subsequent immunoprecipitation with specific antibodies are likely to incorporate both VNTR sequences. As such PCR amplification of one VNTR would not be accurate in defining factors binding to a specific VNTR because both regions will be enriched for despite specific primer sets. PCR results of the ChIP material at this region must refer to the MAOA promoter in general. However when ChIP data is correlated with in-silico sequence analysis detailing predicted specific transcription factor binding sites (Figure3.10), more specific inferences may be made. Consensus sequences for CNBP, hnRNPK and Sp1 can be found in the distal VNTR but not the proximal μVNTR.



Distal

gcccgccctcaacctagtgagggctggaggctgcgcagacctcgacgggccctacatgacgtcacaaaggggccagaccaagtggggcagcaccctgcgaccctgcgatcctgcctggctcagccgccttcatatatctgcttccttaagtccactcttgcccagat
agctttcagttaaaactaaagaatgaaagcactaggttgagagcccacgcggctacacccacgtctactcccccactctcgccaggcaaccgcgcccccgcctgcagtggcatcgtccggccacgcccagtggcagggtttccagcgcgagcctgcaggcaggc
cgggaag**GGC**GGAGCCAGGCCGGCCTAGAGTCACTTCTCCCCGCCCCTGACTGGGCCGGAGCCCGGGGCTGGTCTCTAAGAGTGGGTACCGAGAACAGCCTGACCG
TGGAGAAGGGCTGCGGGAAGCAGAACACCGCCCCCAGCGCCCAGCGTGCTCCAGAAACATGAGCACAAACGCCTCAGCCTCCTTCCCCGGCGGC

Proximal

ACCGGCACCGGCACCAGTACCCGCACCAGT ACCGGCACCGGCACCAGTACCCGCACCAGT ACCGGCACCGGCACCAGTACCCGCACCAGT

ACCGGCACCGGCACCGAGCGCAAGGCGGAGGGCCCGCCCGAAGCCGGGGGCACAACTGCCCAGGTCCCGAACCCGGACTCCAGCTTGGACGACACCTCCTACAG
CCTGTCCGAATGGAGCGTCCGTTCTGAGTGGCGGTCCGTCTCGGATCCGCTAGCCAGTTCCCAGTGGAGCACGTCCTCAACTGCCGAGGCCGCCTCCTGGAGCTCC
AGCATACACTCCCCAATCAGCACTACCGGTCTTAGCGAGAGTACTGACTCCGACTCCAAGAGTGGCCTCCGGGGTTTCAGCGCTTACAACCCGAGCAGTCGGATCC
CCAAGTCTACCACCAGCTCGAACTCCTCCGATGGGGCCGTCACAGCCTCCAATCAGGACACCGGCATTCCCTGGGTATTAGTAACAGGACCTACCCCGCCCGTAAAC
TCCCCCGTAGAGTCATTGCAAGGGTCTGCCTTCTCCTCAGGGTTCAGCACCCCACGGGGGTTTGGTAAAAGGACCGACCCTGCCCCCGGATTCCAACCTGACCTCAGT
GTCCGACTACACTTGGATATTTGTACGGGGACCTCCTATACCCAATGACCTTTCGCAAGTGTCAATACAAGCACCTCCTACACCCAGTAACACCCCCGAGTGTCAGTA
CAAGGGTCTGCCGCATCCTCAGTGTCCAGCTTCCCCTGGGGTTTGGTACCAGGACCACCTCTACCCAATAACATTTCCCCAGTGTCGCCACAAGCACCTCCTGCACC
CCATAACATCCCCCCAGTGTCAAGGCAGGCGTCTACCCCCACCTCAGTGCCTGACACTCCGCGGGGTTCAATACAAGAACCTCCTGCACCCAGTAATCCTTTCCAGC
TGCCGACACAAGGACATTCTAAACCTAATAACTCTCGCCGAGTGTCAGTACAAGGGTCCGCCCCGCTCTCAGTGCCCAGCTCCCCCCGGGTATCAGCTGAAACATCA
GCTCCGCCCCTGGGCGCTCCCGGAGTATCAGCAAAAGGGTTCGCCCCGCCCACAGTGCCCGGCTCCCCCCGGGTATCAAAAGAAGGATCGGCTCCGCCCCCGGGCT
CCCCGGGGGAGTTGATAGAAGGGTCCTTCCCACCCTTTGCCGTCCCCACTCCTGTGCCTACGACCCAGGAGCGTGTCAGCCAAAG

Figure 3.10 Bioinformatic analysis of the transcription factor binding sites across the MAOA Promoter.

Distal and proximal VNTRs highlighted along with consensus sequence for hnRNP K, CNBP and Sp1.Identifying transcription factor consensus sequences present only in the Distal VNTR and not the proximal VNTR.

ChIP analyses over the MAOA promoter region was performed under basal growth conditions and in response to 10 µM cocaine as in previous experiments, with the addition of the mood stabilizer sodium valproate at a physiologically relevant concentration of 2µM (Pan et al., 2005).

The ChIP analysis identified allelic specific interactions in the heterozygous SH-SY5Y cell line (identified as two bands in 'input 1% chromatin'). Histones are only bound to the 3 copy variant and the active H3K9me3 mark is increased in response to cocaine (1hr). Binding of transcription factors is only observed on the 3 copy allele until exposure to valproate and cocaine when we can see binding of hnRNP K and CNBP to both alleles and CTCF after cocaine exposure.

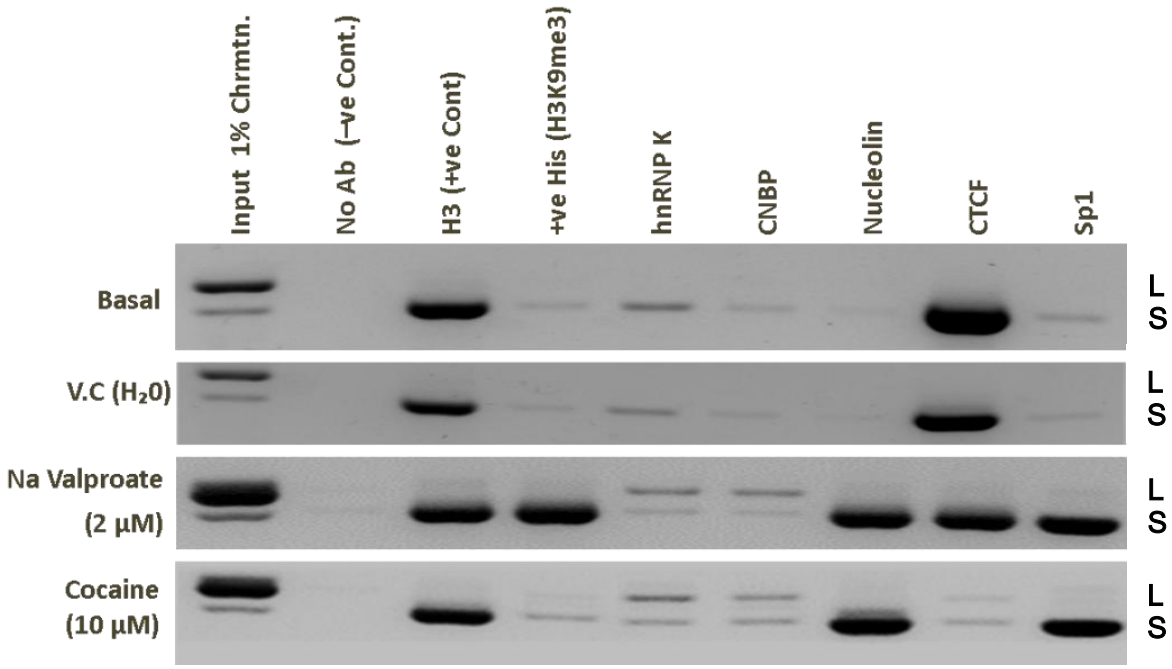


Figure 3.11 G-quadruplex associacted transcription factors bind in an allele specific manner at the MAOA promoter.

Agarose gel showing PCR amplified fragments across the MAOA promoter. Binding of transcription factors is only observed on the 3 copy allele until exposure to valproate and cocaine when we can see binding of hnRNPK and CNBP to both alleles.

3.3.8 Both 3 and 4 copy alleles of MAOA µVNTR are actively transcribing.

The unusual histone binding patterns observed inFigure 3.11; specifically no histones including "ubiquitous" H3 positive control antibody observed to bind the 4 copy allele would suggest unusual chromatin architecture. I performed ChIP for an active RNA polymerase (anti-RNA pol II CTD phospho Ser5 antibody) to determine if active Pol II was loaded on one allele but not the other. Active Pol II binds to both alleles of the MAOA µVNTR (Figure 3.12a) which would suggest both alleles are actively transcribing. I confirmed this by analysing cDNA from SH-SY5Y cells by PCR over the µVNTR demonstrating that both alleles are expressed in fact they are differentially expressed in response to sodium valproate (Figure 3.12b).

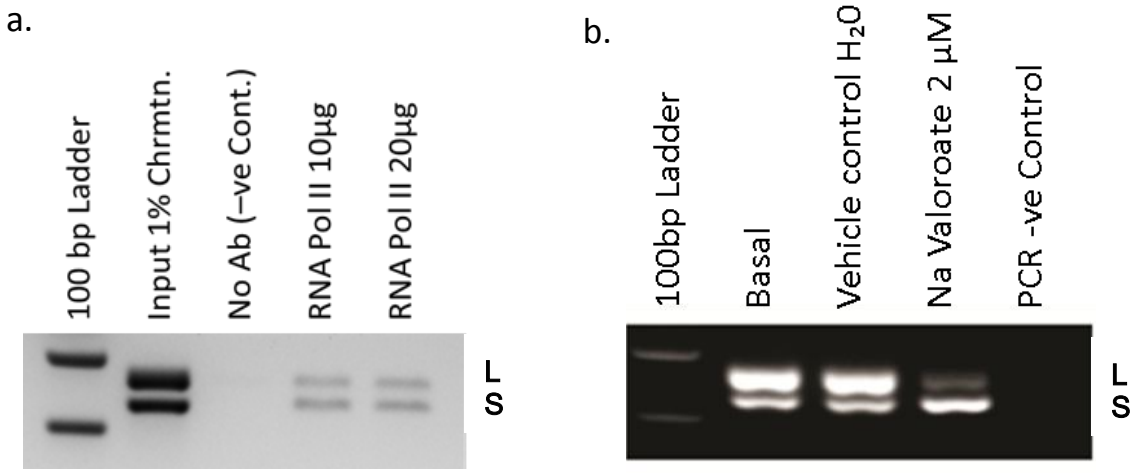


Figure 3.12 Both 3 and 4 copy alleles of MAOA µVNTR are actively transcribing in SH-SY5Y cells.

a. Agarose gel showing PCR amplified fragments across the MAOA promoter, ChIP analysis identifies active RNA polymerase II binding to both 3 and 4 copy µVNTR alleles under normal growth condidtions. b. Agarose gel showing PCR amplified cDNA reveals both alleles are expressed.

3.3.9 Nucleosome exclusion regions overlap with MAOA distal VNTR

A recent publication (Shumay and Fowler, 2010) gave a comprehensive bioinformatic analysis of the MAOA promoter region interrogating genomic features and attributes that might modulate its epigenetic sensitivity. Of particular interest were the identified nucleosome exclusion regions, highlighted in Figure 3.13. The most striking nucleosome exclusion peak (N.Ex) was highlighted in the publication but not described in any detail as the analysis focused primarily on the well characterized μVNTR. When combined with my own bioinformatic analysis, this peak overlaps with the potential G-quadruplex site identified in Figure 3.2g, termed MAOA distal VNTR. This nucleosome exclusion data is in support of the unusual histone binding I identify at the MAOA promoter detailed in Figure 3.11.

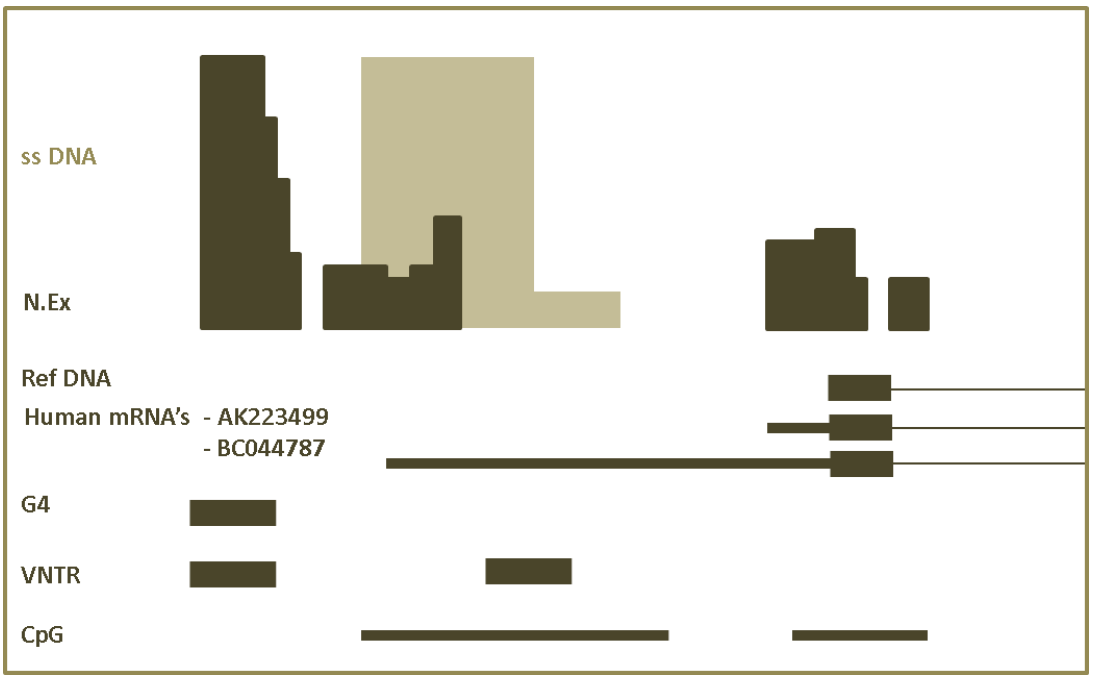


Figure 3.13 Bioinformatic analysis of MAOA promoter.

Top to bottom: ssDNA peak represents UCSC genome browser track of experimentally verified ssDNA at this region, nucleosome exclusion (N.Ex) peaks represent regions of Nucleosome exclusion generated using UCSC custom track. Ref DNA – gene as detailed in hg-19. Human mRNA x2 identified MAOA transcripts. G4 DNA; regions of potential G-quadruplex formation as predicted by Quadparser algorithm, VNTR Variable Number Tandem Repeats as identified by UCSC, CpG islands as identified by UCSC.

3.3.10 Methylation status of potential G-quadruplex target regions in SH-SY5Y cells.

The Balasubramanian group have suggested Sp1 could itself bind G-quadruplex (Raiber et al., 2012) and highlight the fact that that Sp1 elements have been found to protect CpG islands from *de novo* methylation (Brandeis et al., 1994). The group also detail a genome-wide study showing G-quadruplex structures can restrict methylation of CpG dinucleotides (Halder et al., 2010). The Balasubramanian group therefore imply that Sp1 might bind to G-quadruplex regions and prevent subsequent methylation of the sequence. Many promoter regions of housekeeping genes, which are regulated by Sp1, contain methylation free islands suggesting an epigenetic role for Sp1 (Bird, 1986, Holler et al., 1988). In light of this information I performed methylation analysis using a high affinity recombinant methyl-binding domain (MBD) protein pre-bound to magnetic beads (Section 2.2.12.2) across the potential G-quadruplex target regions. The CT element had MBDP bound under normal growth conditions, under vehicle control and after 1 hr 10μM cocaine treatment implying this region is methylated under all conditions (Figure 3.14a.). The 5-HTT 5' LPR had MDBP bound to the short allele only under the same conditions as above implying only the short allele is methylated (Figure 3.14 b). MAOA promoter (μVNTR primer set) had MBDP bound to the 3 copy allele under normal growth conditions (Figure 3.15a,b), after 1 hr 2μM Valproate MDBP was seen to bind to both 3 and 4 copy alleles (Figure 3.15 a) and in response to 10μM cocaine MDBP was lost at both alleles (Figure 3.15 b) implying the 3 copy allele is methylated under normal conditions, both 3 and 4 are methylated in response to valproate and methylation is lost in response to cocaine.

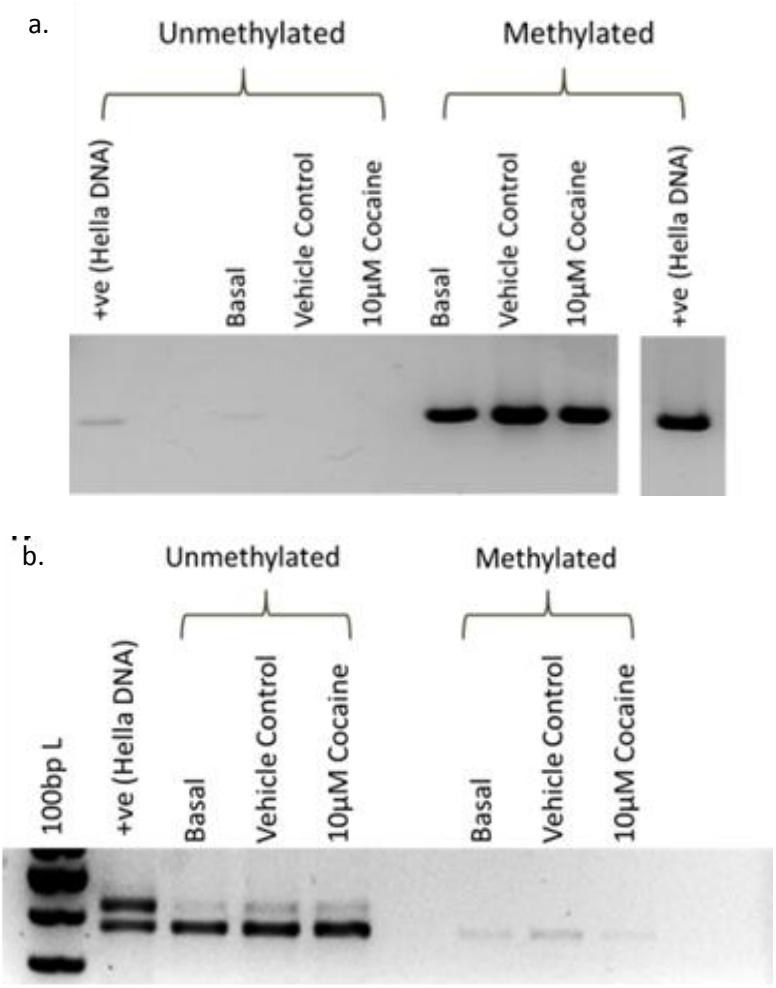


Figure 3.14 a,b Methylation at the CT element, 5-HTT LPR

Agarose gel showing PCR amplification of sheared gDNA post CpG methylQuest treatment: unmethylated DNA; not bound to methyl domain binding protein (CpG MethylQuest™ Protein) Methylated DNA bound to methyl domain binding protein (CpG MethylQuest™ Protein). a. Agarose gel showing PCR amplification of CT element of c-MYC identifying the region is methylated under normal growth conditions and in response to cocaine. b. Agarose gel showing PCR amplification of 5-HTT 5' LPR short allele identifying it is methylated under normal growth conditions and in response to cocaine. (Sheared Hela DNA is used as a positive control)

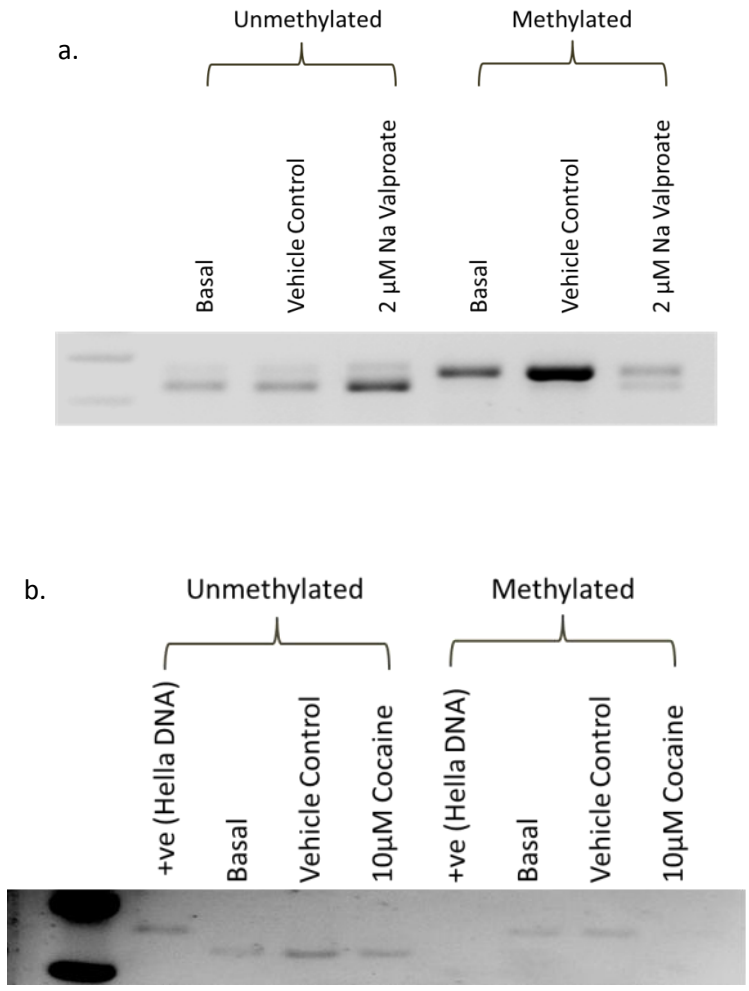


Figure 3.15 a,b Methylation at the MAOA promoter

Agarose gel of PCR amplification of sheared gDNA post CpG methylQuest treatment: unmethylated DNA; not bound to methyl domain binding protein (CpG MethylQuest™ Protein) Methylated DNA bound to methyl domain binding protein (CpG MethylQuest™ Protein) . a Agarose gel showing PCR amplification of MAOA, µVNTR 4 copy allele identifying it is methylated under basal conditions, in response Na valproate both 3 and 4 copy alleles are methylated. b. Agarose gel showing PCR amplification of MAOA, µVNTR identifying methylation is lost at the 4 copy allele in response to cocaine (Sheared Hela DNA is used as a positive control).

3.4. Discussion

An extensive bioinformatic analysis of the promoters of selected genes implicated in cancer (c-MYC, VEGFA and hTERT) and genes implicated in neurological disorders (5-HTT, MAOA and DAT1) was performed. Within the promoters of these genes I identified repetitive sequences of DNA within 1.5kb of the transcriptional start site, that also have the potential to form G-quadruplex as determined by the Quadparser algorithm. These sequences are located within regions of DNA known to adopt ssDNA conformation and have previously been implicated as regulatory regions, however not through their ability to adopt the non B-DNA conformation; G-quadruplex. When considered in the context of the following findings:

i. Transcription has the capacity to modify mechanically DNA topology, DNA structure, and nucleosome arrangement (Kouzine et al., 2013a)
ii. G-quadrpulex structures form in the genome of mammalian cells (Biffi et al., 2013)

This in-silico analysis becomes an informative data set to support further biochemical analysis of the regions. I therefore proceeded to interrogate the transcription factor binding profiles and methylation status of the selected regions.

The model of G-quadruplex formation at the CT element described in Figure 3.6 is adapted from (Brooks and Hurley, 2009) which is in turn based on data from the Levens group (Sun and Hurley, 2009, Michelotti et al., 1995, Tomonaga and Levens, 1996, Tomonaga and Levens, 1995). Although this model is indeed plausible based on the aforementioned studies compelling *in vivo* data detailing the function of G-quadruplex in cells is lacking.

One way to interrogate these regions for the adoption of G-quadruplex is to examine the specific factors known to bind to the three alternate DNA conformations: CTCF and Sp1 to duplex DNA, hnRNP K and CNBP to ssDNA and nucleolin and potentially Sp1 to bind G-quadruplex DNA. Initial ChIP analysis for the G-quadruplex related factors CTCF, hnRNPK, Sp1, and nucleolin at the CT element in the HCT-116 cell line identified all factors except CTCF as binding under normal growth conditions. It was encouraging to see the factors implicated in the dynamic transcription dependent model of G-quadruplex formation present at the CT element however no real interpretation can be made or conclusions drawn due to the complex and dynamic nature of the proposed model. ChIP will capture a single frame across a population of cells, all of which may be at different points in the process of G-quadrplex formation and therefore no clear signal is distinguishable.

In an effort to overcome the limitations of ChIP I moved to the identified 5' LPR target region which is polymorphic and as such heterozygous for the 14 copy (s) allele and 16 copy (l) VNTR the in the JAr cell line. This is of interest as the *short* allele has sequence with the potential to form 2 discrete G-quadruplexs, whereas the long allele has the potential to form 4 discrete G-quadruplexs (Figure 3.16)

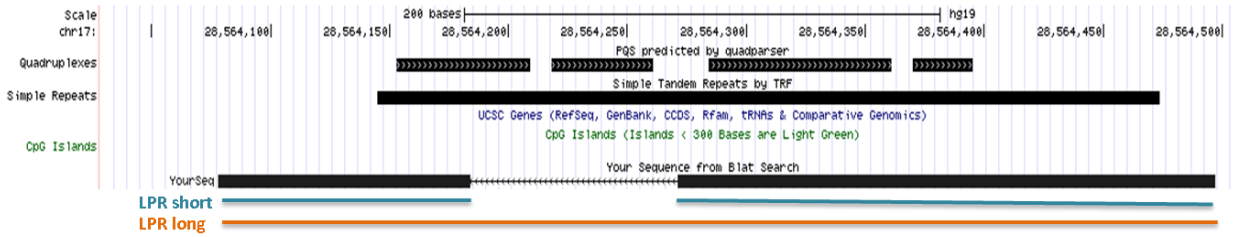


Figure 3.16 G-Quadruplex within the 5-HTT 5' LPR (long and short).

UCSC display magnified to show the 4 potential G-quadruplex sequences within the 5' LPR long (orange line). A BLAT search for the 5' LPR short sequence identifies 2 potential G-quadruplex within it.

The heterozygosity for the polymorphic VNTR's gives an additional layer of information which may provide further clues as to the mechanism behind G-quadruplex formation. The heterozygosity also enables comparison of factor binding between the two alleles in the same assay which is a unique internal control. When coupled with environmental challenges in the form of lithium and cocaine additional layers of information are available to draw upon when assessing a role for factors and alternative structure in the regulation of these regions. These challenges which are known to impact on the expression of 5-HTT through the 5'LPR (Ali et al., 2010, Vasiliou et al., 2012, Roberts et al., 2007) should trigger the regulatory mechanism at play and therefore present a different transcriptional landscape possibly though the structural perturbations as identified in the G-quadruplex formation model.

Under normal growth conditions CNBP, hnRNPK and CTCF bind to both alleles, (preferentially to the short allele in all cases). Whereas Nucleolin and Sp1 bind only to the short allele, the same binding pattern is observed in response to lithium with the exception of CTCF which is lost at the long allele. Binding in response to cocaine is of interest as it appears there is a direct switch between ssDNA factors and G-quadruplex factors on the long allele. CNBP and hnRNPK

binding is lost and nucleolin and Sp1 binding is gained at the long allele, if the model is accurate this could be representative of a change in the conformation of DNA from ssDNA to G-quadruplex at the longer (x4 discrete G-quadruplex sites).

The factor binding patterns of the short allele does not change in response to either challenge which would imply the long VNTR is more susceptible to modulation in response to challenge potentially through G-quadruplex formation. This may correlate to the methylation status observed in Section 3.3.10 as the short allele appears to be methylated under normal growth conditions and in response to cocaine whereas no methylation is observed at the long allele perhaps implying that the long allele is protected by G-quadruplex as suggested by (Raiber et al., 2012)

As described in Section 3.3.6 the neuroblastoma cell line SH-SY5Y is heterozygous for two VNTR's within the MAOA promoter. The 3 and 4 copy proximal μVNTR (Figure 3.9b) and 9 and 10 copy distal dVNTR (Figure 3.9c). The MAOA μVNTR is one of the most well characterised GxE interactions in the literature and the accepted hypothesis states that it regulates the differential expression of MAOA in response to the environment. The μVNTR was originally described as being 1.2kb 5' of the transcriptional start site of the MAOA gene however bioinformatic analysis of data in the Hg19 sequence of the human genome demonstrates there is a 2nd transcriptional start site which is 5' of the μVNTR , now locating this domain in the 5'UTR of the MAOA gene, Figure 3.9a. The distally located VNTR has largely been ignored, possibly due to the complexity of the PCR required to amplify the region due to its highly repetitive GC rich sequence. However in light of the newly described transcriptional start site it is possible that the two VNTR's each regulate an individual transcript.

As described in Section 3.3.6 PCR amplification of ChIP material with a primer set specific for one of the VNTR's would not be accurate in defining factors binding to the specific VNTR as both regions will be enriched for; despite specific primer sets, rather results must refer to the MAOA promoter region in general. However when ChIP data is correlated to in-silico sequence analysis detailing specific transcription factor binding sites (figure 3.10), more specific inferences may be made. Consensus sequences for CNBP, hnRNPK and Sp1 can be found in the distal VNTR but not the proximal µVNTR. The primer set used to amplify the MAOA promoter (sequences can be found in table 1 of the appendix) are for the µVNTR as amplification of the distal VNTR proved inconsistent and generally unsuccessful after many attempts under a wide range of conditions. Interestingly the µVNTR sequence does not contain consensus sequence for the factors identified as binding there so it is reasonable to assume these factors are in fact binding to the distal VNTR.

ChIP across the MAOA promoter in the heterozygous SH-SY5Y cell line offers another opportunity to assess allele specific G-quadruplex related factor binding to identify potential patterns between the two target regions in response to cocaine and the mood stabiliser sodium valproate. Of particular interest was the fact that proteins including the ubiquitous Histone 3 positive control only appeared to bind to the short allele under normal growth conditions and in response to valproate the ssDNA binding factors bind to both the short and the long allele and in response to cocaine ssDNA binding factors and CTCF bind to both alleles. This data in conjunction with the overlapping nucleosome exclusion regions (Shumay et al., 2012) detailed in Figure 3.10 would suggest there is unusual chromatin structure at this region. This potential G-quadrpulex area could be nucleosome free as in the case

of the CT-element, this raises the question are the nucleosomes shifted at the MAOA promoter and if so what is the cause? It is know that alternative DNA structure such as G-Quadruplex is unable to bind to histones (Wong et al., 2007). To add further complexity; when ChIP was performed with an antibody for active RNA polymerase (anti-RNA pol II CTD phospho Ser5 antibody) binding was observed at both alleles suggesting that both alleles are active despite the unusual chromatin architecture.

Taken collectively the bioinformatic interrogation, transcription factor binding analysis and assessment of methylation at the target regions, provide some evidence to support the model of G-quadruplex formation outlined in section 3.3.4. However there is no robust or compelling evidence that has captured and fully described such a mechanism or G-qudaruplex function. Owing in part to limitations of the approach, specifically the standard ChIP assay is not kinetically sensitive therefore is unable to able capture the dynamic nature of the model.

The ChIP assay can accurately determine the location at which proteins bind to DNA *in vivo* (Rhee and Pugh, 2011) and as such has provided fundamental insight into where and how gene regulatory processes occur in cells (Poorey et al., 2013). However recent publications have drawn attention to the limitations of the assay and advised caution when analysing ChIP data (Poorey et al., 2013, Teytelman et al., 2013, Schmiedeberg et al., 2009). Poorey et al., 2013 discuss the fact that ChIP signals fail to differentiate between high-occupancy stable binding or low occupancy dynamic binding. As discussed in Section 1.1.2 many chromatin interactions are short lived with function likely to relate to binding kinetics (Hager et al., 2009). Therefore the time it takes formaldehyde to crosslink the transcription factor to the DNA and the nature of this cross linking is critically important. It is clear from the aforementioned studies that a complete understanding of this process itself is

lacking. This is illustrated in Figure 3.16 which describes four possible cases in which crosslinking kinetics and transcription factor binding dynamics contribute to the increase in ChIP signal observed with increasing formaldehyde incubation time (Poorey et al., 2013).

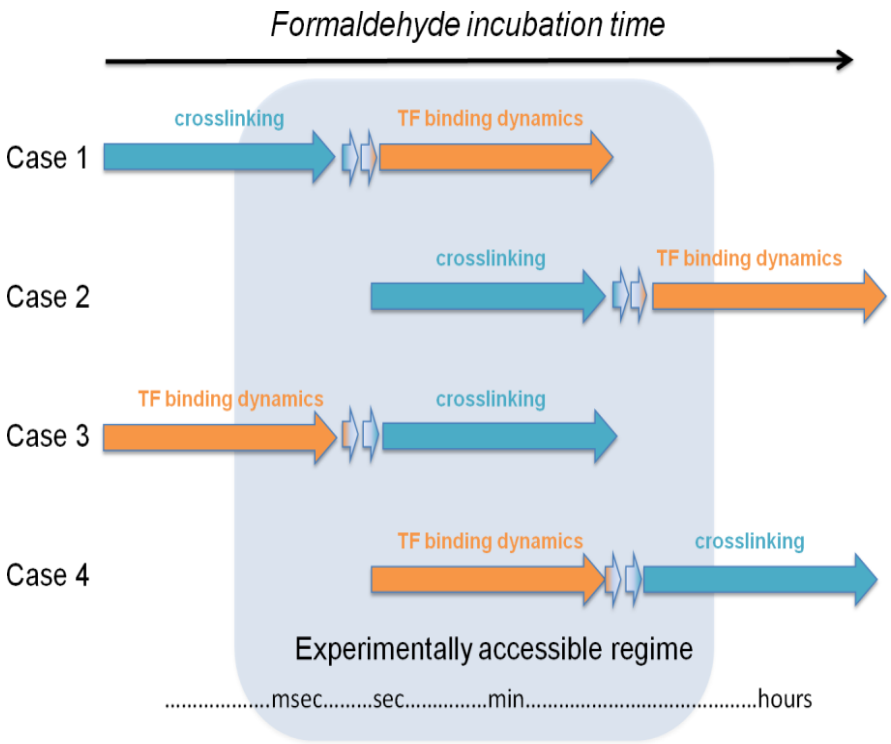


Figure 3.17 Formaldehyde crosslinking kinetics and TF binding dynamics

In cases 1 and 2, crosslinking kinetics is assumed to be much faster than transcription factor (TF) - chromatin binding dynamics. In Cases 3 and 4, TF-chromatin binding dynamics is much faster than crosslinking kinetics (Adapted from (Poorey et al., 2013)

As shown in Figure 3.17, it has not yet been resolved which set of crosslinking kinetics and transcription factor binding dynamics actually occur. Further a complete description of the time it takes formaldehyde to crosslink the protein to the DNA is lacking. *In vitro* reaction rates of ~10^4 $M^{-1}s^{-1}$ have been reported for the covalent modification of DNA bases with formaldehyde (McGhee and von Hippel, 1975,

Siomin et al., 1973). However these do not describe the time taken to crosslink a complete protein and it should be noted that not all proteins are equal with crosslinking times likely to vary between proteins and further variation is likely between chromatin sites to. Perhaps this is explained by recent analysis of crosslinking of MeCP2 to chromatin with studies indicating complete crosslinking time of ~5 seconds with MeCP2 mutants residing in heterochromatin for less than 2.5 seconds on average failing to be captured by crosslink chemistry (Schmiedeberg et al., 2009). These data would support case 3 of Figure 3.17. These findings are striking when we consider the fact that if a protein – DNA interaction occurs and dissociates before the crosslink has time to form, the interaction will not be captured. Schmiedeberg et al., 2009 also proposed formaldehyde-ChIP may depend upon the presence of histones and associated proteins to topologically trap the DNA via cross linking to them, rather than covalent linkages to DNA itself. Further consideration to be noted are the range of possible cross linking events that can occur when formaldehyde enters the cell, as detailed in Figure 3.18.

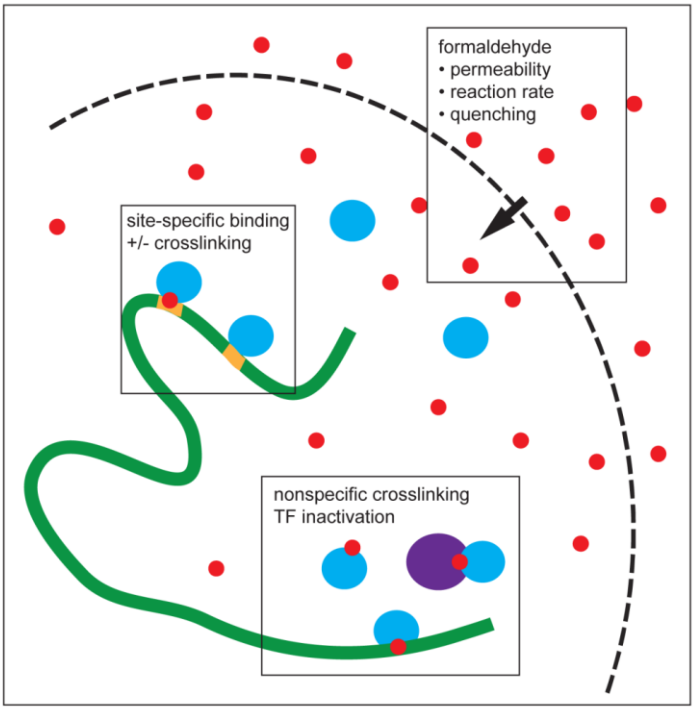


Figure 3.18: Schematic describing possible events that can occur when formaldehyde enters a cell.

Membrane represented by the dotted line, transcription factors shown as the light blue circles, chromatin shown as the thick green line, and formaldehyde molecules as red circles. The orange chromatin segments are specific transcription factor binding sites and red circles on the transcription factor or chromatin represent chemical crosslinking events.

In addition to considerations relating to binding dynamics, a recent study has described "hyper-ChIPable" loci. These represent an artefact in ChIP that results in reproducible but biologically irrelevant enrichment of proteins at highly expressed genes caused by high levels of RNA Pol II and RNA Pol III transcription possibly due to non-specific interactions between some antibodies and polymerase. (Teytelman et al., 2013).

Despite a lack of conclusive evidence describing the G-quadruplex formation, studies at the MAOA promoter revealed other interesting findings. These findings call into question the accepted dogma that X chromosome inactivation by

methylation of one allele normalises this gene for behaviour between males and females.

Data on the MAOA μVNTR in the SH-SY5Y cell line indicated distinct epigenetic structures on each female allele (Figure 3.15) consistent with the view that one allele is methylated. However using the heterozygosity of the μVNTR , I was able to observe active RNA Pol II binding and expression from both alleles. This indicates they are in fact both transcriptionally active (Figure 3.12) and differentially bind transcription factors consistent with their modulation by drug action (Figure 3.11).

Recent work on the X chromosome indicates methylation is not necessarily correlated with gene repression. This may be related more generally to the model in which local genomic regulation allows 15% of the X chromosome genes to escape X inactivation (Mugford et al., 2014, Joo et al., 2014). Complex methylation at this locus would be consistent with data on the epigenetic status of the MAOA locus in a longitudinal study of twins (Wong et al., 2010).

Chapter 4

A strategy to destabilize G-quadruplex in the promoter of c-MYC via homologous recombination

4.1 Introduction

Despite aforementioned studies (Section 3.3.1) detailing the ability of G-quadruplex to form *in vitro* and Section 3.4.3 detailing the mechanism presented as to how they may form *in vivo* and in turn describing how they may function to regulate downstream targets, much skepticism remains in the scientific community. Whether these structures exist under physiological conditions is still a topic of debate and even more controversy exists when considering G-quadruplex biological relevance. It is assumed that any proposed regulatory function of these elements is a result of transcription factor binding; specifically Sp1 with the consensus sequence (GGGCGGGG) which is in essence the main body of a G-quadruplex sequence. The doubt surrounding G-quadruplex formation and function is understandable as a compelling *in vivo* demonstration of G-quadruplex function is lacking. As discussed in Section 3.5, ChIP analysis of the regions we analysed to determine transcription factor binding in the hope that factors would correlate to alternative DNA topology is variable and inconclusive; preventing the assertion of any definitive conclusions.

Here I present a strategy to assess directly the capacity of and mechanism through which these secondary structures influence gene expression. AAV- targeting vectors have been used to re-engineer the G-quadruplex forming elements in c-MYC by homologous recombination. I have designed strategic mutations that destabilize secondary structures, while maintaining Sp1 binding sites and attempted to insert them into the promoter of c-MYC in the HCT-116 cell line via homologous recombination. In addition I have designed mutant sequences to destroy Sp1 binding sites while preserving G-Quadruplex and reengineered these back into HCT-116. QPCR and ChIP analysis of the relative expression and factor binding profiles of G-quadruplex, Sp1 targeted and wild-type alleles will enable a more definitive

structure function analysis of basal or induced expression; determining whether these secondary structures have a functional role in regulating gene expression and the factors involved.

4.2 Aims

Assess directly the capacity of, and mechanism through which G-quadruplex influences gene expression, initially validating the approach in the c-MYC promoter with the well characterized CT element

1. Design mutant sequences to destabilize G-quadruplex while maintaining Sp1 binding sites in the 6 target regions of selected genes.

2. Design mutant sequences to destroy Sp1 binding sites while preserving G-quadruplex in the 6 target regions of selected genes.

3. Construct AAV targeting vectors with strategic mutations that destabilize G-quadruplex while preserving Sp1 binding sites at the promoter of c-MYC.
4. Construct AAV targeting vectors with strategic mutations that destroy Sp1 binding sites while preserving G-quadruplex at the promoter of c-MYC.

5. Use AAV targeting vectors to introduce strategic mutations that destroy Sp1 binding sites while preserving G-quadruplex into the promoter of c-MYC in the HCT-116 cell line via homologous recombination.

6. Use AAV- targeting vectors to introduce strategic mutations that destabilize G-quadruplex while preserving Sp1 binding sites, into the promoter of c-MYC in the HCT-116 cell line via homologous recombination.

7. Use qPCR and ChIP analysis of the relative expression and factor binding profiles of targeted and wild-type alleles to provide definitive structure function analysis of basal or induced expression; determining whether these secondary structures have a functional role in regulating gene expression and the factors involved.

4.3 Results

4.3.1 Mutant sequence design

The potential for a sequence of DNA to form G-quadruplex is determined by the arrangement of Guanine (G) bases, specifically X3 G's, repeated four times with between 1 and 7 residues between each run of GGG as defined by the 'Quadparser' algorithm: (G3+ N1-7 G3+ N1-7 G3+ N1-7 G3+) (Huppert and Balasubramanian, 2007) used in the preliminary in-silico analysis (Figure 3.2). The full Sp1 consensus sequence is defined as: GGA**GGG**A**GGG**ATCGCG, however as with many zinc finger transcription factors, this is degenerate sequence and Sp1 is known to bind a core sequence of GGNGGG (DesJardins and Hay, 1993) this was the reference benchmark core sequence when designing mutations to destroying Sp1 binding sites.

The G-quadruplex and Sp1 sites within each of the WT target sequences have been identified (Figure 4.1) and multiple base pair substitutions made to create G-quadruplex knock out (G4-KO) sequences. These base pair substitutions ensured G-quadruplex structure was unable to form and Sp1 core binding sequence was preserved. A separate mutant sequence was designed; Sp1 KO with base pair substitutions to ensure Sp1 core binding sequence was destroyed while G-quadruplex structure was able to form.

a. **Key:**

Mutated base pairs

Sp1 Binding Site

Potential G-quadruplex Sequence (pG4)

b. c-MYC CT element: Chromosome 8, strand -, position: 128748166 - 128748222 57bp

WT – Sp1 sites (X4) pG4 (X2)

GGGAACCCGGGAGGGGCGCTTATGGGGAGGGTGGGGAGGGTGGGGAAGGTGGGGAGG

Sp1-KO – x 2 pG4 maintained

GGGAACCCGCGAGGGGCGCTTATGGGCAGTGTGGGGAGTGTGGGGAAGGTGAGGAGG

G4-KO – x 4 Sp1 sites maintained

GCGAACCCGGGAGGGGCGCTTATCGGGACGGTGGGGAGGGTGGCGAAGGTGGGGAGG

c. hTERT: Chromosome 5, strand +, position: 1295204-1295271 68bp

WT: Sp1 sites (x4) pG4 sequences (x1):

GGGGAGGGGCTGGGAGGGCCCGGAGGGGGCTGGGCCGGGGACCCGGGAGGGGTCGGGACGGGCGGGG

SP1–KO: (x1) pG4 site maintained:

GGGGATGGGCTGGCAGGGCCCGCAGGGGGCTGGGCCGGGGACCCGGGAGGGGTCGGGACGGCGCGGGG

G4-KO: (x3)Sp1 sites maintained:

GGGGAGGCGCTCGGAGGGCCCGGAGGGGGCTGCGCCGCGGACCCGAGAGGCGTCGTGACGGGCGGGG

d. VEGFA: Chromosome 6, strand +, position: 43737860 -43737909 50bp

WT: Sp1 sites (x3) potential pG4 sequences (x1):

GGGGCGGGCCGGGGGCGGGGTCCCGGCGGGGCGGAGCCATGCGCCCCCCC

SP1–KO (x1 potential pG4 site maintained):

GGCGCGGGCCGGGGGCCGGGTCCCGGCGGGGCCGGAGCCATGCGCCCCCCC

G4-KO (x3 Sp1 sites maintained):

GCGGCGGGCCGGCGGCGGGGTCCCGGCGGGGCGGAGCCATGCGCCCCCCC

e. 5-HTT 5' LPR: Chromosome 17 + strand, position 28564146 – 28564362, 217bp

WT – Sp1 sites (X1) pG4 (X3)

ATGCTGGAGGGGTGCAGGGGGGATGCCGGGGGTGCATGGGGGGATGCTGGGGGGTGCAGGGGGGATACTGCGAGGGG
TGCAGGGGGGATAATGGGGGTTGCAGGGGAGATCCTGGGAGAGGTGCAGGGGGATGCTGGAAGGGCTGCAGGGGGA
TGCTGGGGGTGCAGGGGAGATGCTGGGGGGGCTGCAGGGGGGATGCTGGGGGTGCAGGGGGA

Sp1-KO – x 3 pG4 maintained

ATGCTGGAGGGGTGCAGGGGGATGCCGGGGGTGCATGGGGGGATGCTGGGGGGTGCAGGGGGGATACTGCGAGGGG
TGCAGGGGGGATAATGGGGGTTGCAGGGGAGATCCTGGGAGAGGTGCAGGGGGATGCTGGAAGGGCTGCAGGGGGA
TGCTGGGGGTGCAGGGGAGATGCTGGGGGGGCTGCAGGGGGGATGCTGGGGGTGCAGGGGGA

G4-KO – x 1 Sp1 site maintained

ATGCTGGAGGGGTGCAGGGGCGATGCCGGGCGTGCATGGGGGGATGCTGGGGGGTGCAGGGGGGATACTGCGAGGGG
TGCAGGGGGGATAATGGGGGTTGCAGGGGAGATCCTGGGAGAGGTGCAGGGGGATGCTGGAAGGGCTGCAGGGGGA

f. DAT 1: Chromosome 5, strand - , position : 1446653 - 1446726, 74bp

WT – Sp1 sites (X4) pG4 (X2)

GGGGAGCGGGGGAGCGGGGGGCGGGGAGGGGAGTGGTGGTGTGCGGGGAGTGCGGGGCGGGCGCAGGGGGTGGG

Sp1-KO – x 2 pG4 maintained

GGGGAGCGGGGGAGCGGGGCGCGGGCAGGGGAGTGGTGGTGTGCGGGGAGTGCGGGACGGGCGCAGGGGCTGGG

G4-KO – x 4 Sp1 sites maintained

GGGGAGCGGCGGAGCGGGCGGCGGGGAGGCGAGTGGTGGTGTGCGGGGAGTGCGGGGCGGCCGCAGGGGGTGGG

g. MAOA Distal VNTR: Chromosome X, strand -, position: 43513713 -43513824, 112bp

WT – Sp1 sites (X11) pG4 (X5)

GGGCGGGGCGGGGAGGGGCGGGGAGGGGCGGGGAGGGGCGGGGAGGGGCGGGGAGGAGCGGGGAGGGGCGGGG
AGGAGCGGGGAGGGGCGGGGAGGAGCGGGGAGGGGCGGG

Sp1-KO – x 4 pG4 maintained

GCGCGGGGCCGGGACGGGCGGCGAGGGGCCGGGACGGGCGGGGACGGGCCGGGAGCAGCGGGCAGGGGCAGGGAG
GAGCGGGGACGGGCGGGCAGGAGCGGGGACGGGCAGG

G4-KO – x 10 Sp1 sites maintained

CGGCGGGGCGGGGAGGCGCGAGGAGGGGCGCGGAGGGGCGCGGAGGGGCGCGGAGGAGCGGGGAGGCGCCGGGAG
GAGCCGGGAGGGGCGCGGAGGAGCCGGGAGGGGCGGG

Figure 4.1 Mutated sequence design.

a. Key to identify, mutated base pairs (Green bold underlined), Sp1 core binding sites (highlighted yellow) and potential G-quadruplex forming sequence (Red Bold underlined). b-g. Target genes c-MYC, hTERT, VEGFA, 5-HTT 5' LPR, MAOA and DAT1 detailing location, bp length, wild type sequence and base pair substitutions made to create G4-KO and Sp1-KO sequences.

4.3.2 Assembly of AAV targeting vectors

The CT element 115bp upstream of c-MYC is the most widely studied G-quadruplex target (Brooks and Hurley, 2010) making it the most compelling target to pursue. Owing to the time and cost of AAV targeting vectors as a strategy to generate isogenic cell lines, I decided to pursue one target in isolation and the success of c-MYC editing would inform how to proceed with the remaining 5 targets.

4.3.2.1 AAV targeting vector design

AAV targeting vectors used to introduce the mutant sequence into the HCT-116 genome rely on a neomycin selectable marker gene to provide resistance to cells having undergone homologous recombination to incorporate the mutant sequence. This selection cassette relies on the Cre-Lox system to remove the selectable marker post selection leaving a scar of 34bp (LoxP site) once the selectable marker is removed. It was essential to design the targeting vector in such a way that the remaining 34bpb had no effect on the regulation of transcription as this could mask effects of G-quadruplex destabilising mutations. This was achieved by designing the vector to leave the LoxP scar downstream of the transcriptional start site. To control for potential transcriptional modulation by the LoxP scar, an internal control target cell line was engineered alongside the G4-KO target. The internal control target was termed AAV-LoxP-CT (Figure 4.11), this had no G-quadruplex destabilizing substitutions within the CT element but left the 34bp LoxP insertion at the same location as the AAV-G4-CT (Figure 4.10) LoxP scar (a full list of vectors is shown in table 2 of the appendix). Comparison between wild type, AAV-G4-CT and AAV-LoxP-CT

would enable me to determine any possible impact of the LoxP insertion on downstream analysis assays. The LoxP insertion also acts as a tag to identify the targeted alleles, distinguishing them from the respective unmodified alleles. Further design considerations include maximum homology of the AAV targeting vector to ensure highest efficiency of homologous recombination events (Deng and Capecchi, 1992). This is limited by the 4.7 kb packaging capacity of AAV, the rAAV TK accepter has 3kb available for homology.

4.3.2.2 AAV targeting vector assembly.

Left and right homology arms was constructed from strands of synthetic dsDNA, this simplified the mutagenesis procedure considerably as base pair substitutions designed in Section 4.3.1 were incorporated into the synthetic sequence ordered as gBlocks from Integrated DNA Technologies (IDT). The gBlocks and pAAV-TK-accepter vector DNA were assembled by Gibson isothermal assembly (Gibson et al., 2009) a recently developed cloning strategy which enables up to 20 fragments of linear DNA to be assembled in a single 1 hour reaction. Fragments of double stranded DNA with homologous ends (12-20bp overlap) are incubated with T5 exonuclease a mesophilic digester with 5′→3′ single-strand-specific endonuclease activity which chews back the 5' strand exposing a 3' ssDNA overhang (T5 exonuclease is inactivated at 50 ˚C. The ssDNA overhangs then align with the complementary strand, hydrogen bonds form and proof reading polymerase phusion now active at 50 ˚C fills in the missing bp, the remaining nick in the phosphodiester backbone is sealed with Taq ligase (Figure 4.2).

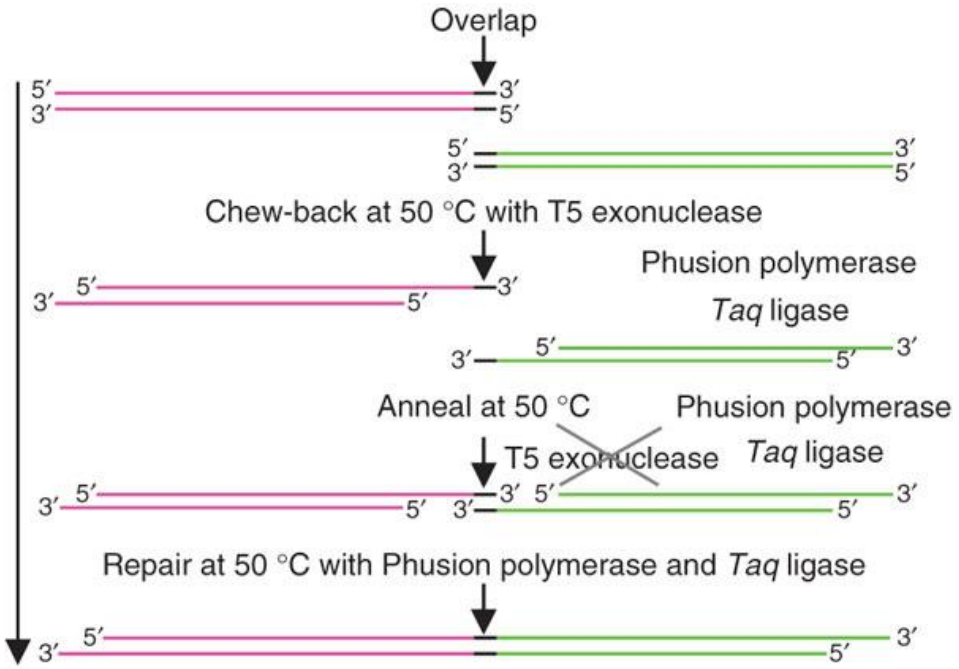


Figure 4.2 Gibson Isothermal Assembly

Schematic detailing the cloning strategy used to join multiple strands of linear DNA together in a single reaction detailed in section 4.3.2.2.

4.3.2.2.1 Left homology arm assembly

The left homology arm was assembled from x3 gBlocks including the mutated sequence within gBlock 3 in a single isothermal assembly reaction as follows.

1. gBlock 1 (Figure 4.3) begins with 20bp homologous to the multiple cloning site of pAAV-TK-accepter; specifically the 20bp prior to the BamH1 restriction site (Figure 4.4a). The remaining 460bp are homologous to the endogenous c-MYC promoter region.
2. gBlock 2 (Figure 4.3) begins with 20bp homologous to the last 20bp of gBlock 1 (overlap sequence) with the remaining 480bp also homologous to endogenous DNA, with the exception of the 4bp substitutions which prevent G-quadruplex formation making this the mutant.
3. gBlock 3 (Figure 4.3) begins with a 20bp overlap of the last 20bp of gBlock 2 (Figure 4.3) the next 460bp are once again homologous to endogenous DNA and the last 20bp are homologous to the 1st 20bp after the BamH1 restriction site.
4. gBlock 1,2,3 (Figure 4.3) are assembled via Gibson Isothermal assembly outlines in Section 4.3.2.2 to create a 1.5kb fragment of dsDNA.

Figure 4.3 Left homology arm assembly

gBlock 1,2,3 to be assembled via Gibson isothermal assembly to make the Left Homology arm.

4.3.2.2.2 Right homology arm assembly

The right homology arm is gBlock 4 which starts with 20bp homologous to the 2nd multiple cloning site of pAAV-TK-accepter; specifically the 20bp prior to the SacII restriction site, the next 460bp are homologous to endogenous c-MYC promoter region, continuing from the endogenous sequence in gblock3 , prior to its 20bp vector overlap. The right homology (gBlock 4) is inserted into the pAAV-TK-accepter via isothermal assembly detailed in Section 4.3.3.5.

4.3.2.2.3 Construction of the pAAV-G4-CT targeting vector

The construction of the pAAV-G4-CT vector occurred in two phases (two individual isothermal assembly reactions). Although it is possible to assemble 6 fragments in one reaction a double digestion of pAAV-TK-accepter at SacII and BamHI would have exposed x2 identical overlap regions; the LoxP sites. This would have resulting in erroneous assembly due to lack of specificity in the form of repeated overlapping sequence.

4.3.2.2.3.1 Linearization of pAAV-TK-accepter and insertion of RHA (gBlock 4)

pAAV-TK-accepter plasmid was linearized by digestion with SacII (Section 2.2.5), once digested the linearized vector was treated with T4 DNA polymerase (Section 2.2.2.2) to blunt end the 3' overhangs created by SacII digestion. The linearized, blunt ended pAAV-TK accepter was isolated by gel electrophoresis (Section 2.2.1.4), (Figure 4.4) and purified using gel wizard gel clean up (Section 2.2.1.5).

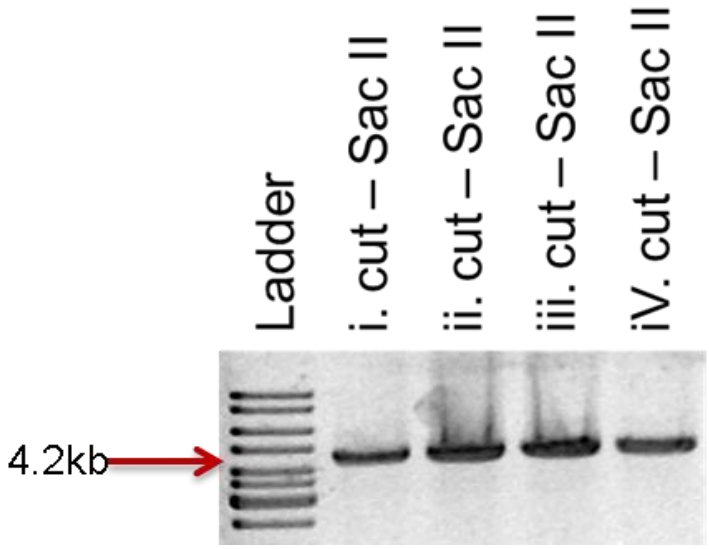


Figure 4.4 Linearized pAAV-TK-accepter

Agarose gel showing pAAV-TK-accepter linearized by digestion with sacII, x4 samples were analysed using gel electrophoresis to reveal a band of 4.2kb.

Purified pAAV-TK-accepter was combined with the RHA (gBlock 4) via Gibson Isothermal Assembly (Section 2.2.2), (Figure 4.5)

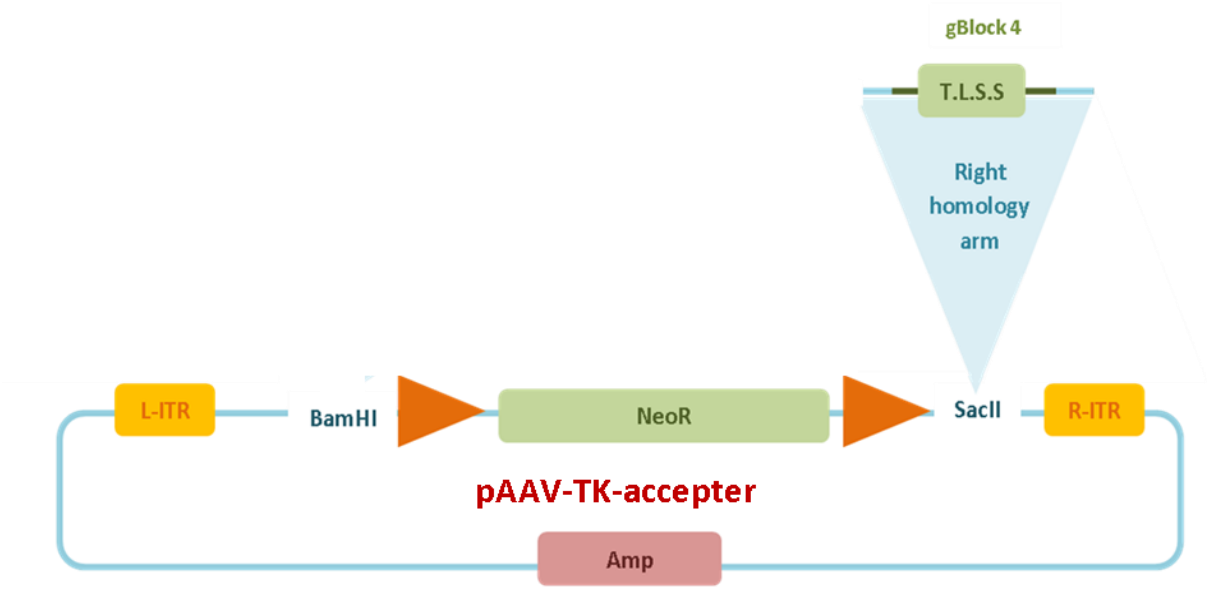


Figure 4.5 Insertion of RHA (gBlock 4) into the pAAV-TK-accepter.

pAAV-TK-accepter was linearized by digestion with SacII, the RHA (gBlock 4) was inserted at the SacII site via Gibson Isothermal Assembly.

`The pAAV-TK-accepter + RHA (rAAV-Gb4) plasmid was transformed using XL Gold ultracompotent cells (Section 2.2.3.3). 20 colonies were picked, expanded and purified (Section 2.2.4.1) then screened by restriction enzyme digestion (section2.2.5). pAAV-Gb4 was digested with restriction enzyme AatII to cut once and linearize the plasmid to enable comparison against uncut. pAAV-Gb4 was also digested with restriction enzyme puvII which cut in four locations, twice in the insert and twice in the backbone. Therefore 4 fragments (500bp, 700bp, 815bp and 2.7kb) will be present on the agarose gel analysis (Figure 4.6) if successful assembly had occurred.

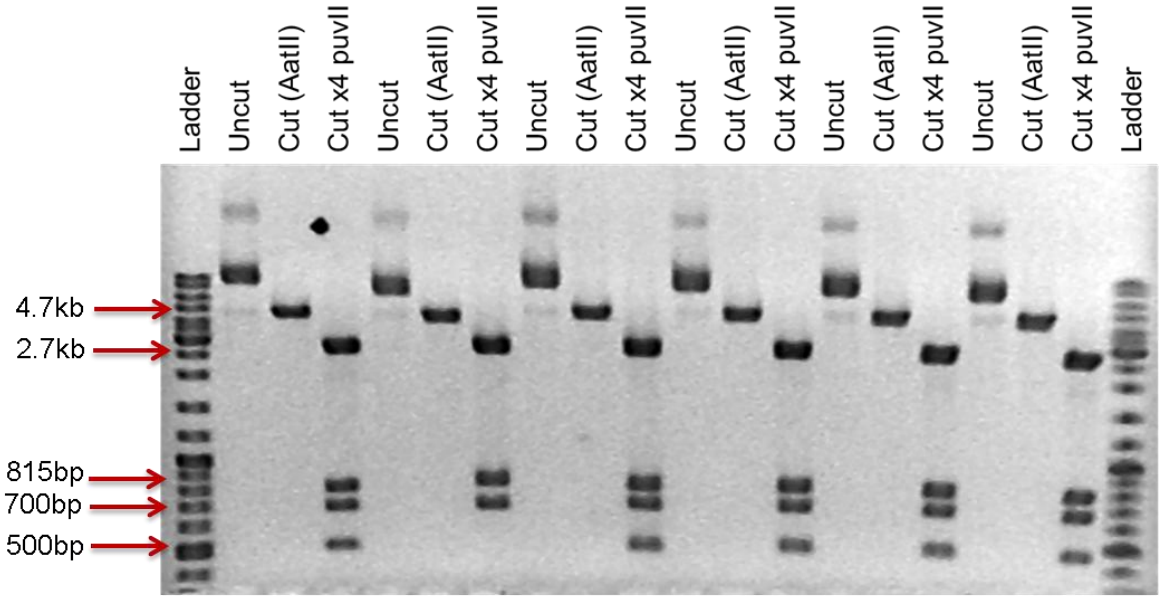


Figure 4.6 Restriction enzyme digestion analysis of rAAV-Gb4.

Gel electrophoresis of uncut purified plasmid, cut x1 (digested with AatII) and cut x4 digested puvII, 4 fragments identify successful assembly.

4.3.2.2.3.2 Linearization of pAAV-Gb4 and insertion of LHA (gBlock 1,2,3)

pAAV-Gb4 was re-linearized with restriction enzyme BamHI (Section 2.2.5) and isolated by gel electrophoresis (Section 2.2.1.4). The linearized band was purified using gel wizard gel clean up (Section 2.2.1.5). The Isolated, linearized pAAV-Gb4 was combined with gBlocks 1,2,3 (LHA) and fragments were once again assembled via Gibson Isothermal assembly (Section 2.2.2) to create a circular plasmid; pAAV-Gb4 + LHA inserted in the BamHI site giving rise to rAAV-G4-CT (figure4.7), this was verified by restriction enzyme digestion (figure4.8), The enzyme restriction enzyme AccI cut once in the insert and once in the vector therefore 2 bands of the correct size (3.7 and 2.5) identify a successful Isothermal Assembly event.

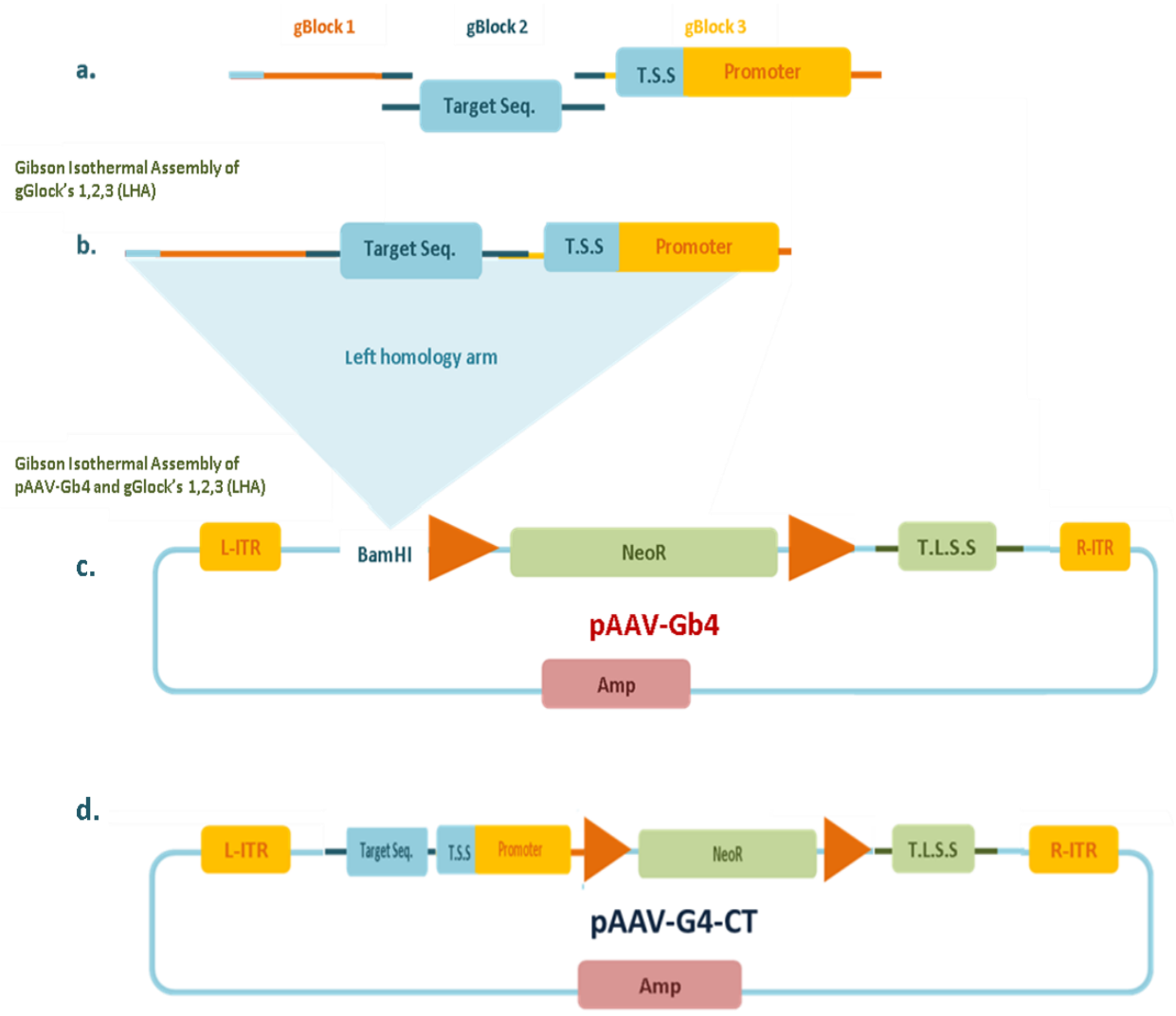


Figure 4.7 Construction of the pAAV-G4-CT targeting vector.

a. gBlocks 1,2,3,4 each 500bp, colour coded to identify overlapping regions, designed for Gibson Isothermal Assembly, to each other and into the vector backbone. b.LHA consisting of gBlock1+2+3 c. pAAV-Gb4 detailing restriction site BamHI for insertion of LHA via Gibson Isothermal Assembly. d. LHA, and RHA inserted to form rAAV-G4-CT.

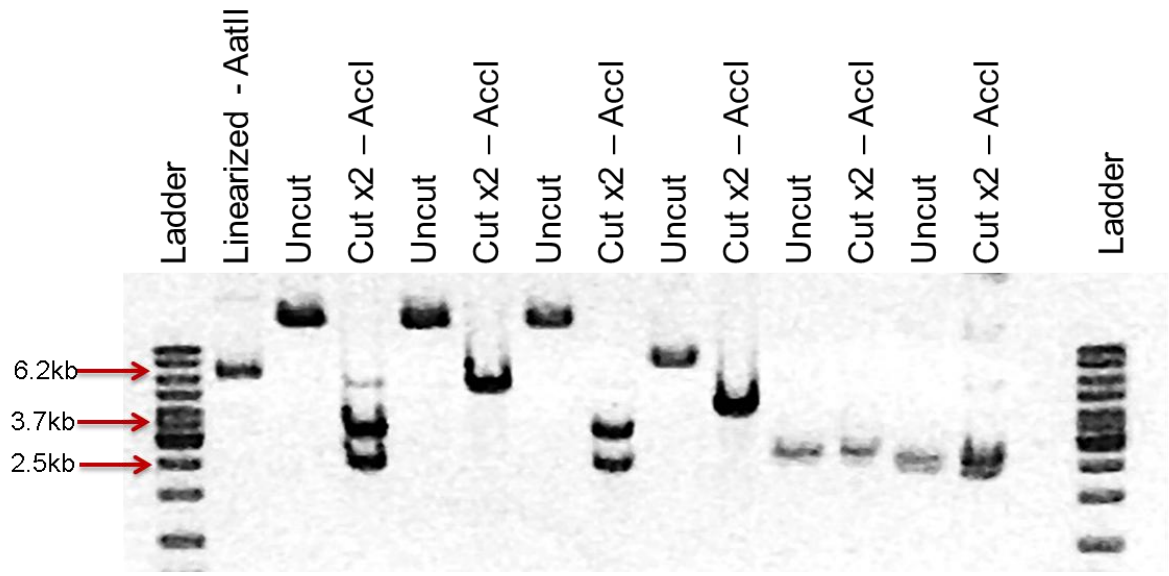


Figure 4.8 Restriction enzyme digestion analysis of rAAV-G4-CT targeting vector

Gel electrophoresis of uncut purified plasmid, cut x1 (digested with AatII) and cut x2 (digested AccI) 2 fragments identify successful assembly.

4.3.2.3 Production of infectious AAV particles

In order to obtain the infectious AAV particles the recombinant expression plasmid (pAAV-G4-CT) was co-transfected into the HEK-293 cells (Section 2.2.9.1) with pHelper (carrying adenovirus-derived genes) and pAAV-RC (carrying AAV-2 replication and capsid genes). Together these vectors supplied all of the *trans*-acting factors required for AAV replication and packaging in the HEK-293 cells.

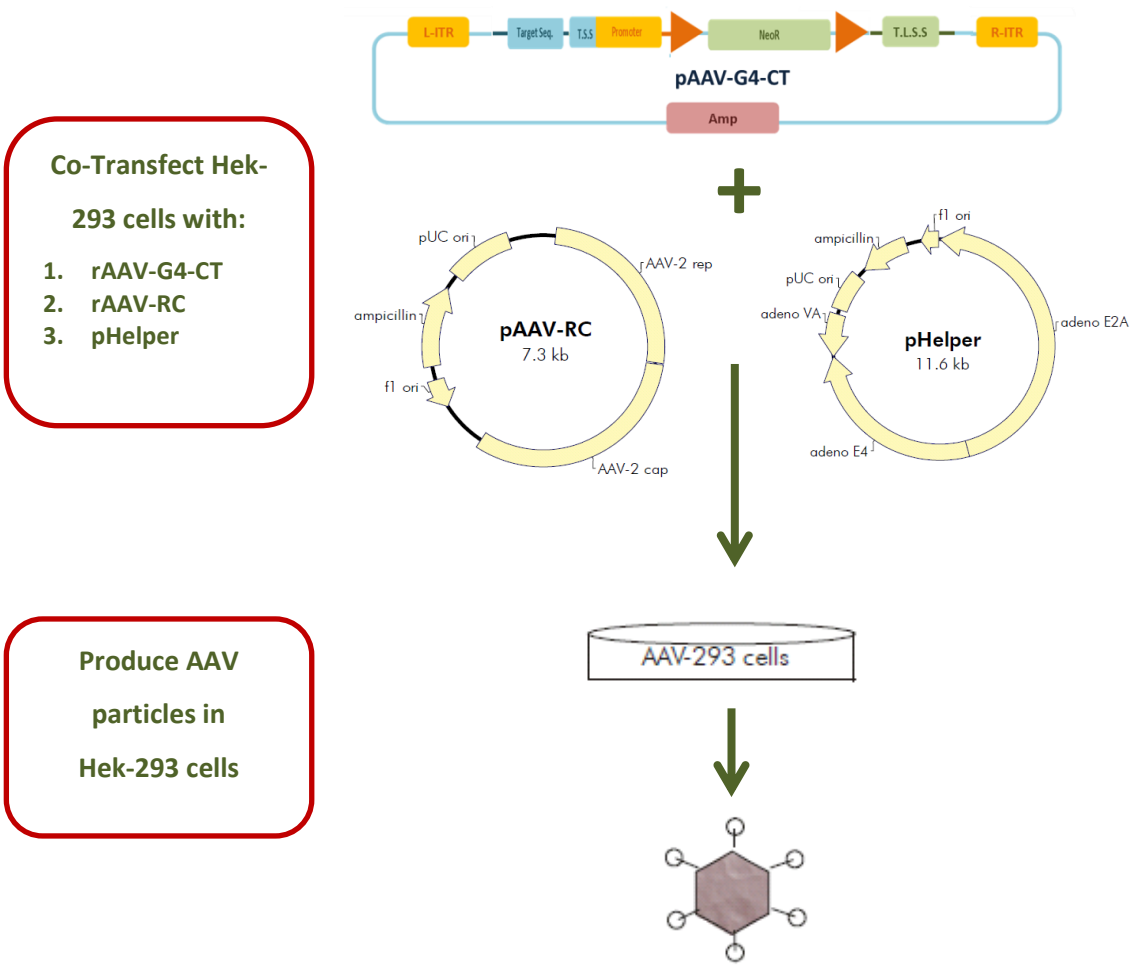


Figure 4.9 AAV virus particle production

Co-transfection of pAAV-G4-CT, pAAV-RC and pHelper together supplying all *trans*-acting factors required for AAV replication and packaging in the Hek-293 cells (Adapted from AAV-helper free system Instruction manual; Agilent technologies)

4.3.2.4 Overview of AAV targeting approach

Recombinant AAV-2 viral particles prepared from infected HEK-293 cells were then used to infect HCT-116 cells. AAV vector genomes are efficiently uncoated and transported to the nucleus where single stranded forms are available for pairing with homologous genomic sequence, homologous recombination occurs at higher frequencies than are observed with conventional approaches (Hirata and Russell, 2000)

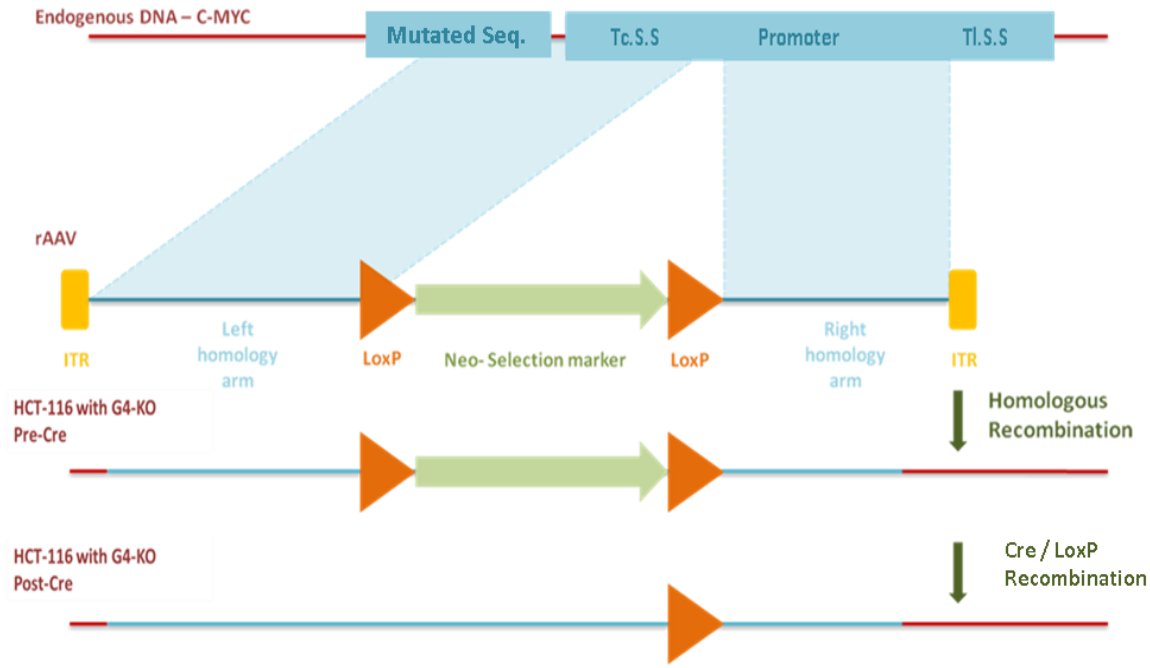


Figure 4.10 Overview of AAV targeting approach to KO G-quadruplex at the CT element.

Overview of the AAV targeting strategy used to introduce 4 base pair substitutions to prevent G-quadruplex formation at the CT element 115bp upstream of the c-MYC promoter, in the HCT-116 cell line via homologous recombination.

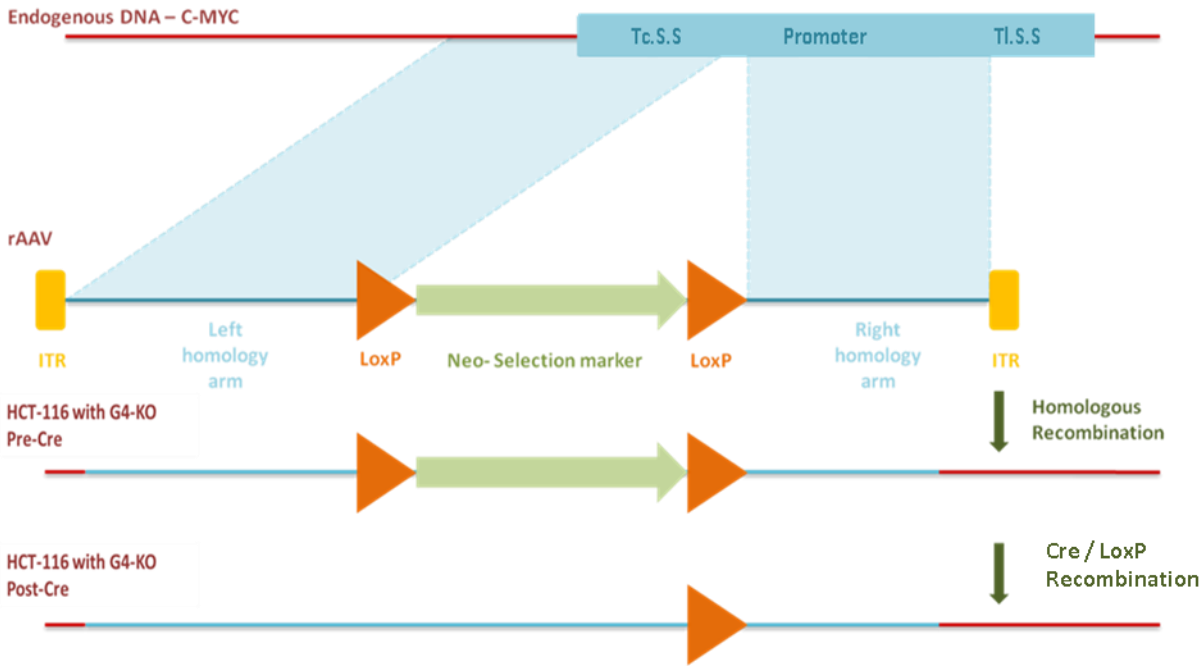


Figure 4.11 Overview of AAV targeting approach to introduce LoxP internal control at the CT element.

Overview of the AAV targeting strategy used to introduce neutral base changes (LoxP scar) to the 5´-UTRs of the targeted alleles, distinguishing them from their respective unmodified homologous partners whilst acting as an internal negative control for remaining Lox P site in the G4-KO cell line (a. above).

4.3.3 AAV targeting strategy

4.3.3.1 AAV targeting strategy 1

Viral particles were generated, HCT-116 cells infected and subsequently plated out based on the protocol outlined in (Kim et al., 2008) , an overview of the approach can be seen in Figure 4.12. Once cells were plated out in selection media, unfortunately no colonies grew, suggesting no recombination events had occurred, a detailed examination as to the possible causes can be found in the discussion Section of this chapter (Section 4.5).

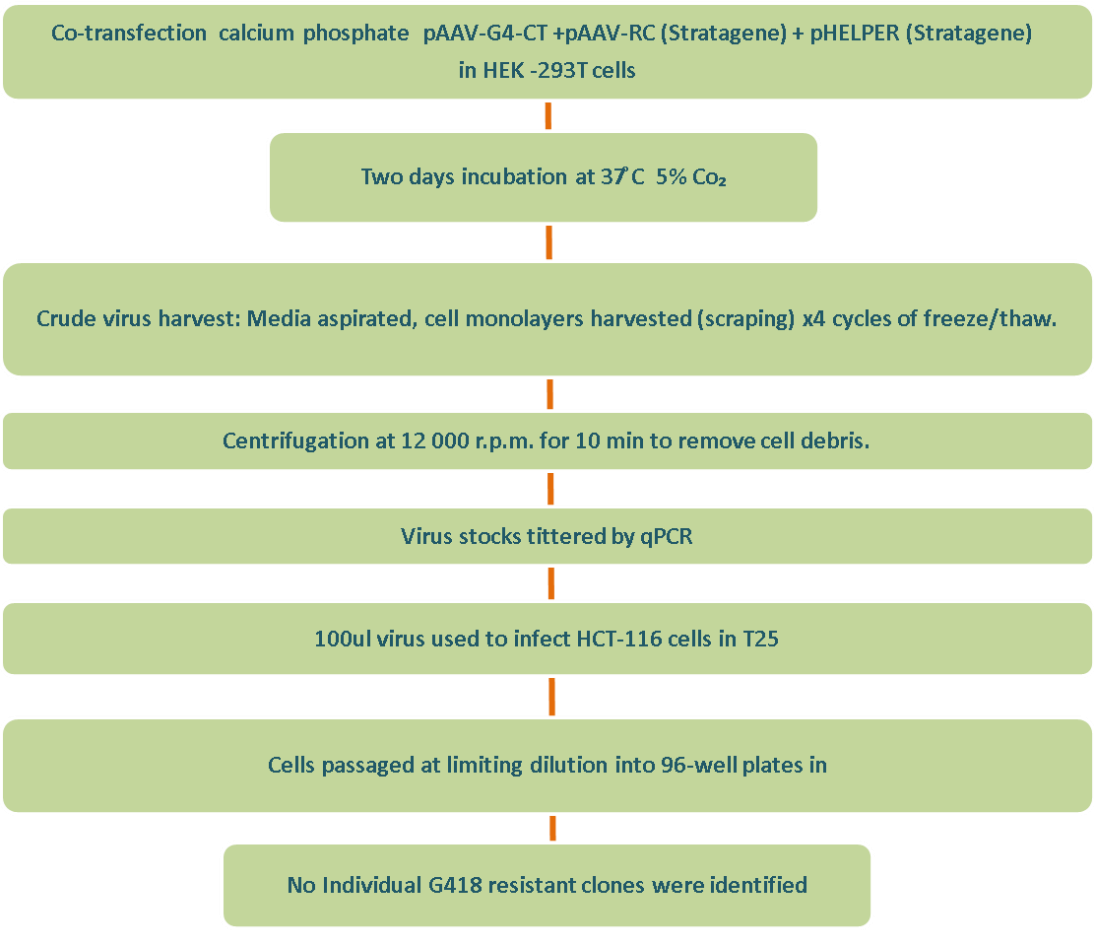


Figure 4.12 AAV targeting strategy 1:

Details the 1st strategy used to generate live AAV viral targeting vectors, infect HCT-116 cells and select clone.

4.3.3.2 AAV targeting strategy 2

Based on analysis and troubleshooting of the previous approach to AAV targeting, a modified strategy was designed. Modifications of strategy 2 included:

1. Altered co-transfection protocol using Lipofectamine LTX (Section 2.2.9.1)
2. Alternative helper plasmid system -pDP2rs (Plasmid Factory)
3. Increased incubation periods up from 48hrs to 72 hrs
4. Extensive virus purification (VIRAPUR AAV purification Kit) detailed in section2.2.13.3.
5. Increased range of virus to cell ratio; MOI.
6. Increased range at with infected cells were plated into 0.5ml/mg G418 selection media (Section 2.2.13.5)

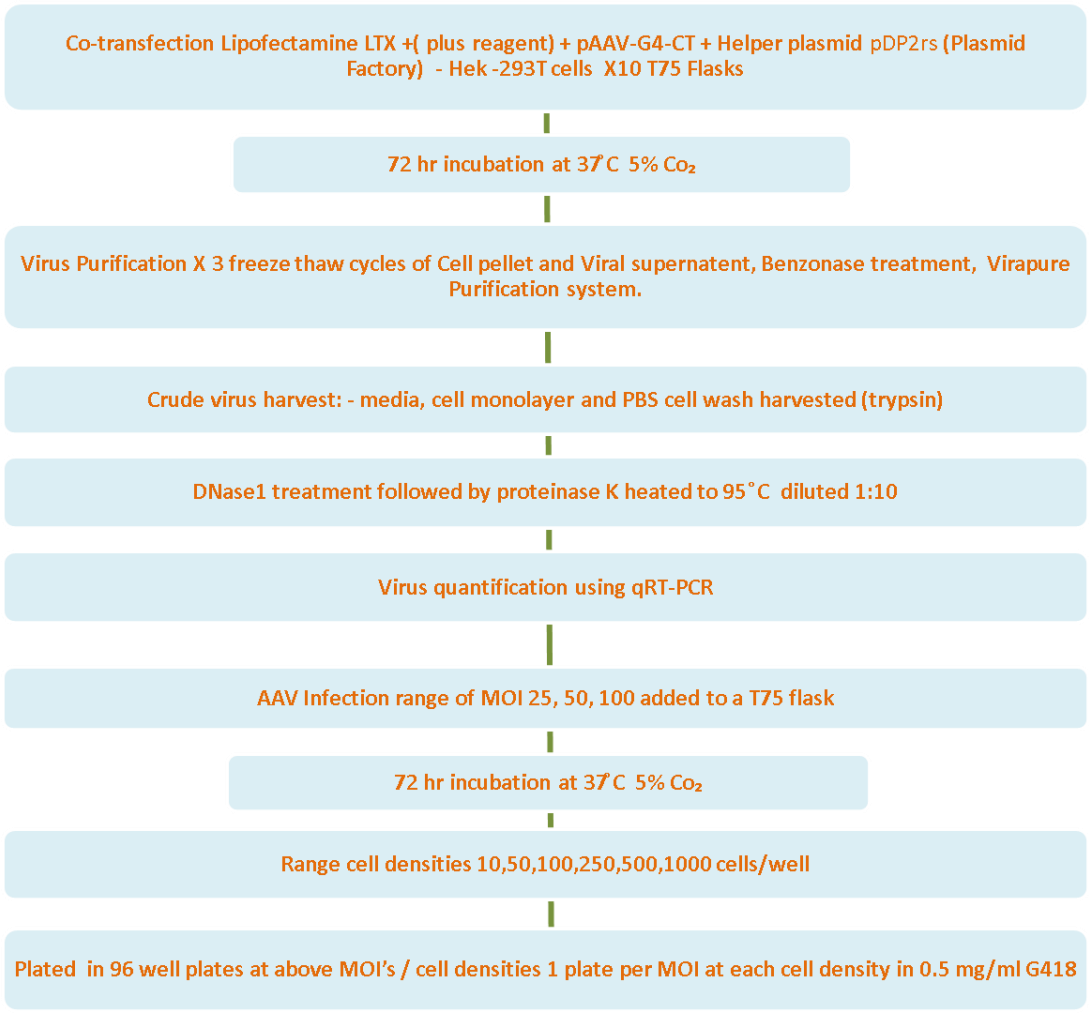


Figure 4.13 AAV targeting strategy 2

Details the modified strategy used to generate live AAV viral targeting vectors, infect HCT-116 cells and select clones.

4.3.3.2.1 Co-transfection of pDP2rs pAAV-G4-CT / pAAV-LoxP-CT.

As part of the modified strategy a helper plasmid tagged with red fluorescent protein (RFP) was chosen to allow confirmation of successful Co-transfection. In order to asses transfection efficiency of pDP2rs pAAV-G4-CT or pAAV-LoxP-CT into HCT-16 cells, 24 hours post transfection, cells were fixed and imaged using an Evos digital inverted microscope. RFP expression from helper plasmid pDP2rs was observed (Figure 4.14) when co-transfected. Transfection efficiency was estimated at 30%.

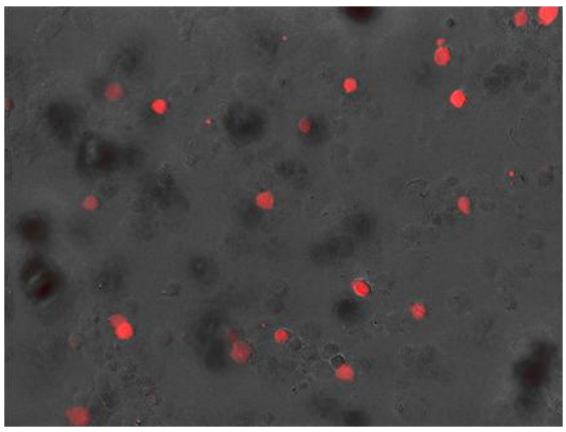


Figure 4.14 Co-transfection efficiency - pDP2rs rAAV-G4-CT

Lipofectamine LTX transfection reagent used to transfect pDP2rs pAAV-G4-CT into HCT-116 cells at 70% confluency estimated transfection efficiency is 30%.

4.3.3.2.2 Range of cell to virus ratio (MOI) and range at which infected cells were plated (cell / well).

A further amendment in the revised strategy was the need to increase the range of MOI and the range at which infected cells were plated as having these variables at the optimum ratio will allow homologous recombination to occur at highest frequencies (Miller, 2011). Cells were plated into 96 well plates at desities of 10, 50, 100, 250, 500, 100 cells per well, each cell desity was infected at an MOI of 25, 50, and 100, as summarized in Table 4.1.

25 MOI	50 MOI	100 MOI
10 cells / well	10 cells / well	10 cells / well
50 cells / well	50 cells / well	50 cells / well
100 cells / well	100 cells / well	100 cells / well
250 cells / well	250 cells / well	250 cells / well
500 cells / well	500 cells / well	500 cells / well
100 cells / well	100 cells / well	100 cells / well

Table 4.1 Range of cell to virus ratio (MOI) and range at which infected cells were plated (cell / well).

4.3.3.2.3 PCR screen: AAV-G4-CT and AAV-LoxP-CT

Once cells were infected and plated at the ranges detailed in Table 4.1 positive recombination events were identified through PCR. Pools of clones were PCR screened using a primer set with the reverse primer nested within the Neomycin selection cassette and the forward primed outside the LHA producing an 1810pb positive product. In parallel a WT primer set was used as a negative control for recombination event and a positive control for the HCT-116 gDNA and PCR reaction conditions amplifying a product of 1141bp.

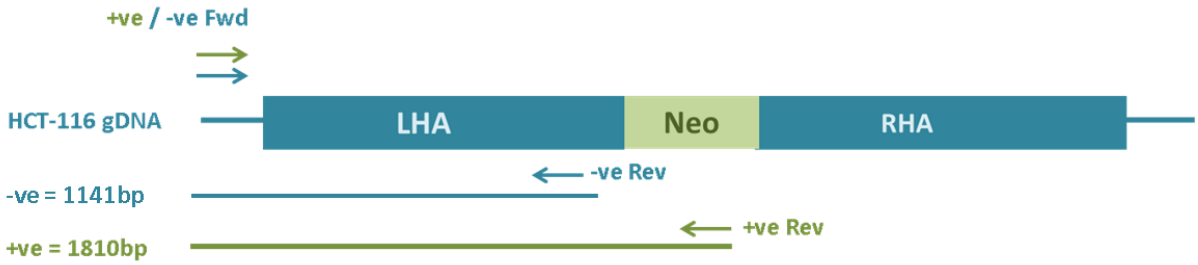


Figure 4.15 PCR screen primer design and strategy

Schematic detailing primer set used to identify positive recombination events amplifies a PCR product 1810bp (green) along with a primer set to identify negative recombination events. This primer set also provides a positive control for gDNA extracted from the pools of targeted HCT-116 cells and a positive control for the PCR reaction conditions amplifying a product of 1141bp (blue).

4.3.3.2.4 rAAV-G4-CT/ rAAV-LoxP-CT control PCR.

HCT-116 gDNA and PCR screen reaction conditions were verified as correct by screening the 204 colony containing wells from rAAV-G4-CT plate out and 526 colony containing wells from rAAV-LoxP-CT plate out with the control primer set identified in Figure 4.15 amplifying a product of 1141bp, full primer sequence can be found in Table 1 of the Appendix.

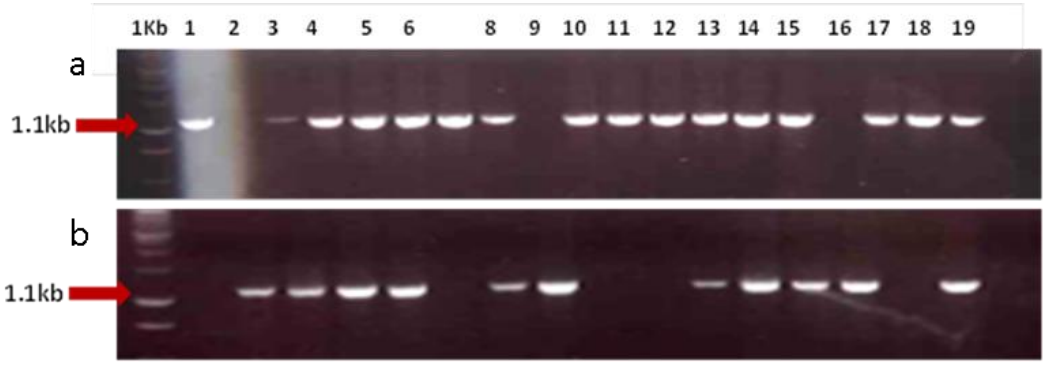


Figure 4.16 rAAV-G4-CT/ rAAV-LoxP-CT control PCR

a. Agarose gel showing a selection of the 204 rAAV-G4-CT wells containing colonies post infection b. agarose gel showing a selection of the 526 wells containing colonies post infection PCR screened to verify HCT-116 gDNA integrity and suitable PCR reaction conditions PCR product 1141.

4.3.3.2.5 rAAV-G4-CT PCR screen.

rAAV-G4-CT recombination events were verified by PCR screening gDNA extracted from plated HCT-116 cells. Of the 1728 wells plated with infected HCT-116 cells (96 wells at each range detailed in Section 4.3.3.2); 204 wells grew colonies. Positive primer set identified in table 1 of the appendix was used to amplify a PCR product of 1141bp, of the 204 screened two positive clones were identified.

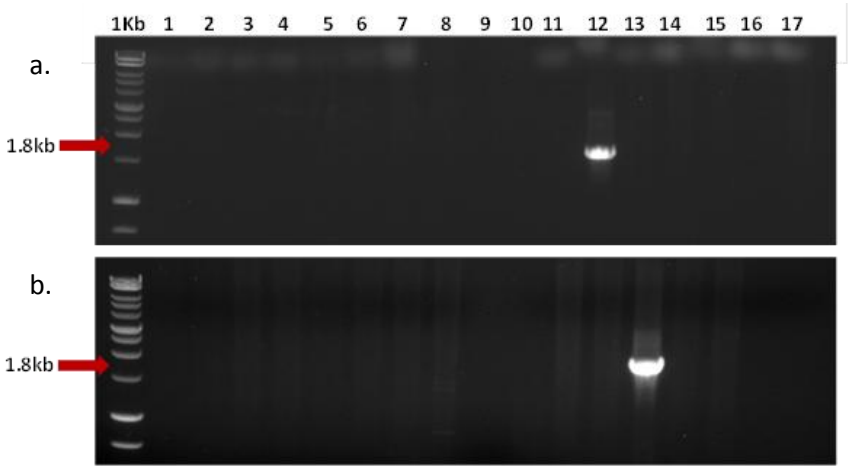


Figure 4.17 rAAV-G4-CT PCR screen

Agarose gel showing two positive recombination events were identified by amplification of a 1810bp fragment in gel a; lane 12, and gel b; lane 13.

4.3.3.2.6 rAAV-LoxP-CT PCR screen

rAAV-LoxP-CT recombination events were verified by PCR screening gDNA extracted from plated HCT-116 cells. Of the 1728 wells plated with infected HCT-116 cells (96 wells at each range detailed in Section 4.3.3.2) 526 wells grew colonies. The positive primer (details of which can be found in table 1 of the appendix) was used amplifying a PCR product of 1810b, 8 positive clones were identified.

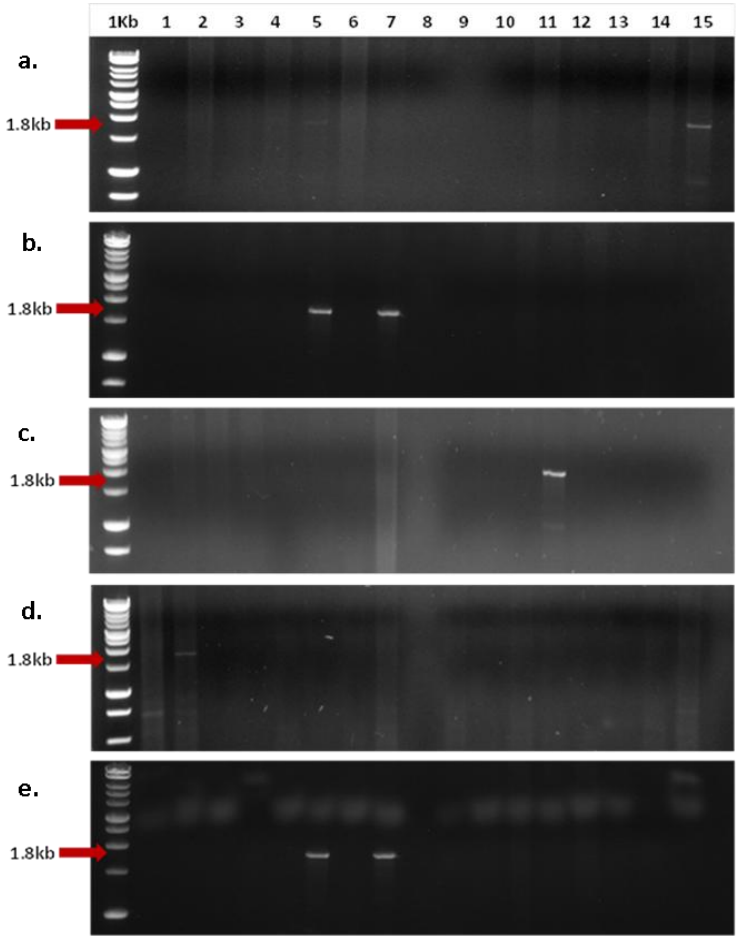


Figure 4.18 rAAV-LoxP-CT PCR screen

Amplification of the 1810bp fragment identified 7 positive recombination events in gel a; lane 5, 15, gel b Lane 5,7 gel c lane 11, gel d lane b and gel e lane 6 and 8.

4.3.3.2.6 Positive rAAV-LoxP-CT and rAAV-G4-CT PCR re-screen.

Positive pools identified in Section 4.3.3.5 and 4.3.3.6 were expanded up to the 24 well stage to ensure robust growth prior to single cell dilution (SCD) to isolate clonal populations. At the time of the SCD the pools of positive cells not expanded up to 24 well plate were re-PCR screened with the same two primer sets outlined in Section 4.3.3.3. Upon rescreening only one of the 6 samples screen was identified as positive Figure 4.19.

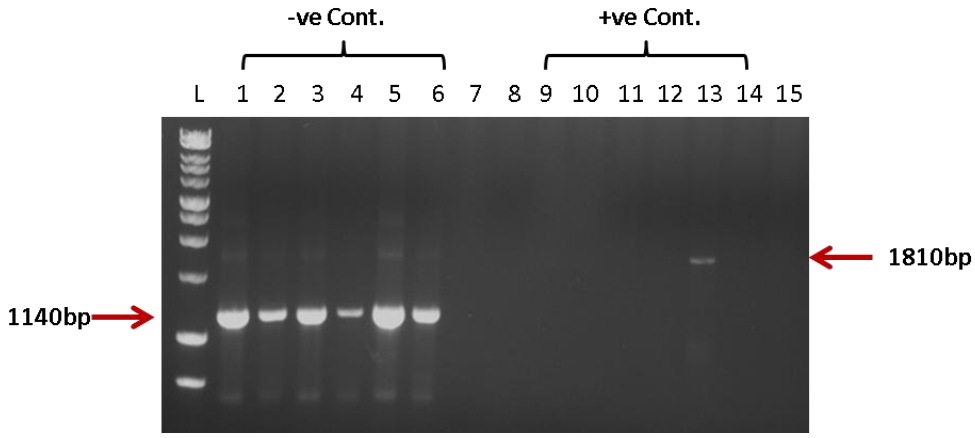


Figure 4.19 Positive rAAV-LoxP-CT and rAAV-G4-CT PCR re-screen.

gDNA extracted from targeted HCT-116 cells, previously identified as positive (at the 96 well stage) Lane 1-6 –ve control primer set to verify gDNA and reaction conditions are suitable Lane 9-14 +ve primer set for inserted Neomycin cassette outline in Section 4.3.3.3.

4.3.3.2.7 Sequence verification of positive rAAV-LoxP-CT.

All 6 samples re-PCR screened in Section 4.3.3.7 were sent for sequencing (Source Bioscience Life Sciences). Only 1 of the 6 were sequence verified as positive, this correlated with the 1 positive sample identified in the PCR re-screen Section 4.3.3.3, sequencing data was verified using the sequence alignment tool clustal omega (EMBL EBI) http://www.ebi.ac.uk/Tools/msa/clustalo.

```
H12_seq_        ----------------------------------------------------CTGG-AAGAC 9
Expected_       TTGTCACTCCCACTTGTGCCCAAAAAGCCTTGCCACATCTTCCCTCCTCACTGGAAAGAC 60
                                                                   **** *****

H12_seq_        AGCACTCTTCTGTGTTAAGTATTTGATTTTGTGATTTGTCTTTCAGAATTGGAAATAGAG 69
Expected_       AGCACTCTTCTGTGTTAAGTATTTGATTTTGTGATTTGTCTTTCAGAATTGGAAATAGAG 120
                ************************************************************

H12_seq_        CGGCCATTTGGATTTGGGCAGGAAGCAGCCGAGCACAGCTTTGGATCCTTCTTTAGGGAA 129
Expected_       CGGCCATTTGGATTTGGGCAGGAAGCAGCCGAGCACAGCTTTGGATCCTTCTTTAGGGAA 180
                ************************************************************

H12_seq_        ATCGAGTTATGGATTTATGGTCCCGGTCAAGCTCAGCCCATCCCCAGGCAGGGGCGGGCT 189
Expected_       ATCGAGTTATGGATTTATGGTCCCGGTCAAGCTCAGCCCATCCCCAGGCAGGGGCGGGCT 240
                ************************************************************

H12_seq_        CAGCGAGCAGCAAGAGTTCTGGTGGCGGCGGCGGCGGCAGTAGCAGCGGCAGCGGTAGCA 249
Expected_       CAGCGAGCAGCAAGAGTTCTGGTGGCGGCGGCGGCGGCAGTAGCAGCGGCAGCGGTAGCA 300
                ************************************************************

H12_seq_        GCGGCAGCGGTAGCAGCGGCAGCGGCAGCTTATAACTTCGTATAATGTATGCTATACGAA 309
Expected_       GCGGCAGCGGTAGCAGCGGCAGCGGCAGCTTATAACTTCGTATAATGTATGCTATACGAA 360
                ************************************************************

H12_seq_        GTTATGACCTGGAACCTTCCTGAAAGTCTAGAAATTCTACCGGGTAGGGGAGGCGCTTTT 369
Expected_       GTTATGACCTGGAACCTTCCTGAAAGTCTAGAAATTCTACCGGGTAGGGGAGGCGCTTTT 420
                ************************************************************

H12_seq_        CCCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGGCACTTGGCGCTACACAAG 429
Expected_       CCCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGGCACTTGGCGCTACACAAG 480
                ************************************************************

H12_seq_        TGGCCTCTGGCCTCGCACACATTCCACATCCAC 462
Expected_       TGGCCTCTGGCCTCGCACACATTCCACATCCAC 513
                ********************************
```

Figure 4.20 Sequencing data from positive pool of cells verified using the sequence alignment tool clustal omega (EMBL EBI).

4.4 Discussion

In order to assess directly the capacity of and mechanism through which G-quadruplex influences gene expression, I designed a set of mutant sequences to destabilize G-quadruplex while maintaining Sp1 binding sites. Alongside these I designed a set of mutant sequences to destroy Sp1 binding sites while preserving G-quadruplex in the 6 target regions of selected genes (Figure 4.1). The intended approach was to use AAV targeting vectors to reengineer the genome of HCT-116 cells to incorporate these mutant sequences via homologous recombination. This approach is novel, high risk and high cost in both time and resources. I therefore proceeded with one preliminary gene target; the well characterised G-quadruplex containing CT element of c-MYC in which to optimise and streamline the protocol before continuing with the remaining target genes.

The combination of synthetic gBlock fragments incorporating the mutant sequence and Gibson Isothermal Assembly was a quick and efficient methodology to engineer the pAAV targeting constructs. This technique was piloted in the Levens Lab NIH and brought back to the Quinn Lab where it is now being used routinely.

The initial attempt to produce infectious recombinant AAV particles was based on the methods outlined in (Kim et al., 2008) and proved ineffective in my hands. A comprehensive analysis of the protocol revealed a number of steps in that could be altered to improve efficiency and are discussed below.

The initial strategy relied upon a calcium phosphate co-transfection protocol, although this can be an effective method to deliver pAAV constructs into cells, I decided to use to the cationic liposome formulation method Lipofectamine LTX (Section 2.2.9.1). In addition, I used an alternative helper plasmid system -pDP2rs (Plasmid Factory), which is tagged with RFP. This enabled me to assess transfection

efficiency which I was unable to do using the previous pHelper system (Stratagene). By inserting internal controls throughout the protocol it enabled me to determine the success of the individual steps and therefore effectively troubleshoot the approach if I failed to get recombinant virus on the second approach. Based on findings in a recent review discussing AAV targeting efficiency (Miller, 2011). I also increased the incubation period post transfection and post infection from 48hrs to 72hrs. The initial protocol used a "crude" virus with limited purification, which was isolated by four cycles of freeze/thaw. The only purification step was centrifugation at 12,000 r.p.m. for 10 min to remove cell debris however no further viral purification was performed. After consulting with colleagues using AAV at the Liverpool CRUK I decided to purify the virus using VIRAPUR AAV purification kit. I also treated the viral DNA with DNaseI to digest residual and contaminating DNA prior to quantification via qPCR as this could lead to a more accurate viral titre. Based on the fact that one of the largest determinants of recombination frequency is MOI (Miller, 2011) I decided to increase the range of both MOI and the density at which the cells were plated out at. This was increased from 100µl of virus in a T25 tissue culture flask as recommended in (Kim et al., 2008) to the range identified in Section 4.3.3.2.2 in the hope that the optimum virus to cell ratio would be found.

The amended strategy did give improved results, as outlined in Section 4.3.3.2.3. From the 204 colonies screened 2 positive clones were identified for rAAV-G4-CT. From the 526 colonies screened for rAAV-LoxP-CT 8 positive clones were identified which is an improvement on no colony growth from strategy 1. Despite this improvement the efficiency was low, this was further decreased as 9 of the 10 positive pools of cells were lost through SCD leaving only 1 sequence

verified positive clone from 3456 wells plated with live virus giving a targeting frequency of 0.03%.

The initial targeting frequency was at 0.1% for rAAV-G4-CT and 0.5% for the rAAV-LoxP-CT in the pools of cells, prior to SCD. The frequency of G-quadruplex KO at the CT element was lower than the neutral base pair changes in the 5' UTR (LoxP insertion). This could be due to the fact that a key regulatory region in c-MYC which is essential for so many cellular processes as detailed in Section 1.3.1.was targeted. It is attractive to postulate that the mutation may have functioned to prevent potential G-quadruplex formation and as such has influenced c-MYC function. However this can not be confirmed because subsequent down stream analysis was prevented by further loss of positive pools through SCD. Although low and not sequence verified these initial targeting frequencies from the pools of cells are in line with estimated targeting efficiencies of AAV of 0.1–1.0% (Miller, 2011) and several orders of magnitude higher than those achieved with other gene targeting strategies (Thomas and Capecchi, 1987, Vasileva and Jessberger, 2005).

It was essential to decipher why the positive colonies in the pools of cells were lost through SCD, in order to move forward and improve efficiency of targeting. Deciphering the best cause of action at this stage in the protocol was difficult it is essential to gain single colonies in order to confirm a truly "isogenic" cell line. It is therefore preferable to get the pools of cells to single cell dilution rapidly but in contrast, colonies need to reach sufficient confluency to be robust enough to be PCR screened (withstanding direct PCR lysis) and SCD. For this reason it is beneficial to have single colonies in the initial plate out. This is most likely from plateing cells at low density i.e. 10 cells per well plates, however in these

experiments very few of the low density (cells/well) plates formed colonies. The 500 cells per well and 1000 cells per well plates formed the majority of colonies which in turn increases the number of colonies forming within the well, thus creating pools of colonies with a mix of correct and incorrect recombinants.

In this second attempt I had cultured the positive pools of clones to the 24 well stage, during which time I believe they were outcompeted by non-recombinant or erroneous recombinants and the positive clones were lost. Therefore when redesigning the strategy I would SCD directly after the confirmation of positives through PCR. Alongside this I would pick colonies using blotting paper from cells seeded in 10cm dishes to avoid SCD.

Chapter 5

CRISPR & AAV, an efficient approach to gene editing

5.1 Introduction

AAV targeting of the CT element in HCT-116 cells as detailed in chapter 4 resulted in a low targeting frequency of 0.03%, with only 1 successful recombination event being maintained out of a potential 3,456. This prevented further downstream analysis. Here I present a strategy combining the recent technology of the CRISPR-Cas system with traditional AAV targeting to KO the microRNA Mir137 in SH-SY5Y cells.

Alongside AAV targeting of the CT-element in c-MYC; a key gene for the Levens Lab, I supervised an MRes student Daniel Grifiths in applying AAV targeting vectors to KO the Mir-137; a key gene for the Quinn lab. Despite revised AAV protocols, both targeting events failed to produce a functional number of recombination events. Owing to funding requirements of my PhD, upon my return to the UK I moved my attention from the CT-element to the Mir-137 gene, in order develop an efficient gene editing approach combining the CRISPR-Cas system and rAAV. to be applied to the CT element and other previously identified G-quadruplex targets at a later date.

5.1.1 CRISPR-Cas

RNA-guided Cas9 nucleases derived from clustered regularly interspaced short palindromic repeats (CRISPR)-Cas systems have transformed the field of gene editing. It is believed that tools and techniques based on Cas9 which is able to colocalize RNA, DNA and protein, will grant unprecedented control over cellular organization, regulation and behaviour (Mali et al., 2013b).

Prokaryotes have evolved adaptive immune defences; CRISPR-CRISPR–associated (Cas) systems that use two short RNAs to direct a single polypeptide nuclease; Cas9 to target sites to degrade foreign nucleic acids (Bhaya et al., 2011, Wiedenheft et al., 2012, Horvath and Barrangou, 2010) (Figure 5.1).

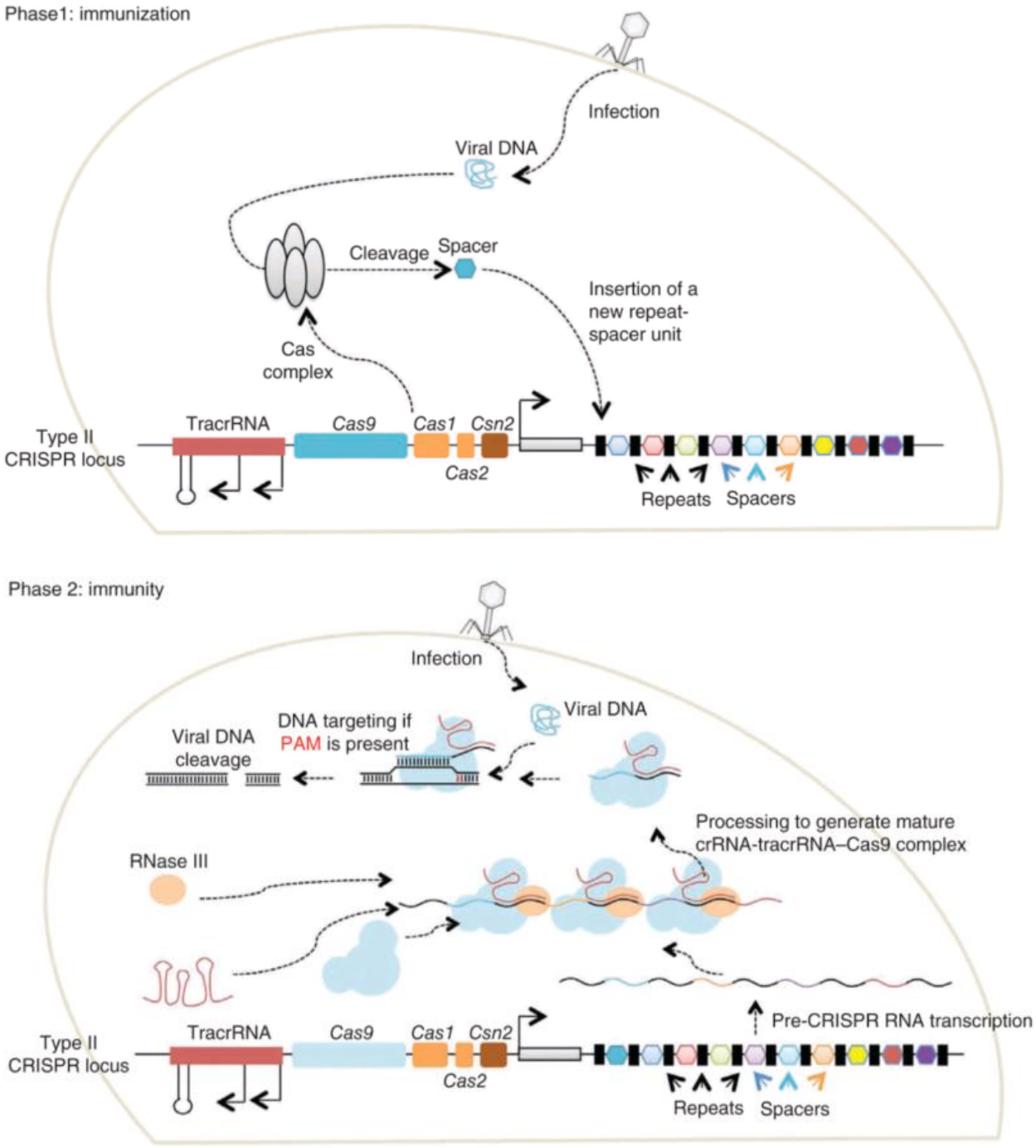


Figure 5.1 Type II CRISPR-Cas systems in bacteria.

Phase1 Immunization: CRISPR system stores the molecular finger print of previous infections by incorporating small fragments of invading phage or plasmid DNA into the CRISPR locus termed 'spacers'.

Phase 2 Immunity: The bacterium uses information from phase 1 to produce crRNAs, to guide nucleases to locate and cleave invading nucleic acids complementary to the spacer. (Adapted from (Mali et al., 2013b)).

The *Streptococcus pyogenes* system has been simplified for use in genome engineering. It is comprised of two components; the Cas9 nuclease and a chimeric guide RNA (gRNA). The gRNA recognizes a 20-nt target sequence next to the protospacer adjacent motif (PAM) trinucleotide; NGG in the genome. This directs Cas9-mediated cleavage of both DNA strands to generate a blunt-ended double-stranded break, 3 base pairs upstream of the 3′ end of the protospacer (Jinek et al., 2012) (Figure 5.2). This process has been shown to induce NHEJ and HR in eukaryotic systems (Cho et al., 2013, Mali et al., 2013c, Mali et al., 2013a). However, the specific details of this process remain poorly understood, as demonstrated by the wide range of efficiencies and specificities observed in many systems (Mali et al., 2013b).

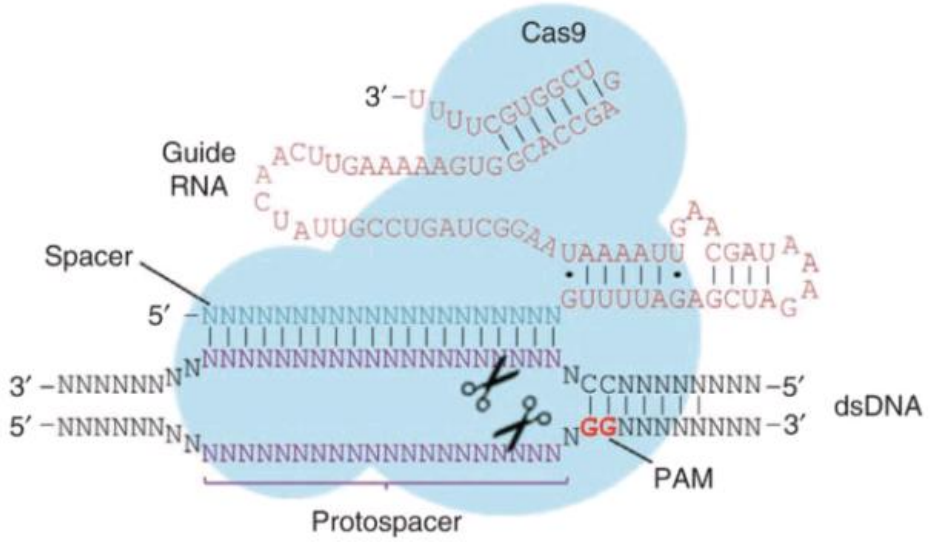


Figure 5.2 Cas9-sgRNA targeting complex.

The S.pyogenes Cas9-sgRNA RNA-guided nuclease complex detailing recognition and cleavage.

5.1.2 Gene targeting efficiency is increased by double strand breaks

The introduction of a DSB in target sequences has been shown to increase the frequency of gene targeting by 100- to several 1,000- fold in conventional gene targeting approaches. As such gene targeting with transfected, linear dsDNA constructs is believed to occur by DSB-induced HR (Brenneman et al., 1996, Smih et al., 1995, Sargent et al., 1997). It has since been shown that introducing DSBs into the chromosomal target could further improve rAAV targeting efficiencies (Miller et al., 2003, Porteus et al., 2003). These studies identify that DSBs result in up to 100- fold increased targeting frequencies. It should be noted that in both aforementioned studies, the frequency of random integration was not affected. It is therefore plausible to suggest a DSB might shift the ratio of targeted events from random-integration events towards targeting. Collectively, these studies indicate that the mechanism of rAAV-mediated gene targeting involves the repair of DSBs by HR, which is independent of random integration by NHEJ (Vasileva and Jessberger, 2005).

5.2 Aims

Use emerging technology CRISPR-Cas system to pump prime the AAV gene editing system to increase recombination efficiency in order to create a hemizygous Mir-137 KO.

1. Design and engineer CRISPR-Cas gene targeting system to introduce double stranded breaks at the Mir137 locus.
2. Target SH-SY5Y cells with Mir-137 CRISPR-Cas targeting system in tandem with rAAV-Mir-137-KO donor.

5.3 Results

5.3.1 A strategy to KO Mir-137 using CRISPR-Cas & AAV.

The CRISPR-Cas system uses RNA guided Cas9 nucleases to introduce double strand breaks at specific sites determined by the guide RNA (Terns and Terns, 2011, Mali et al., 2013c). In this case the Mir137 locus was the target site, a double strand break at this region should trigger the natural DNA repair pathways NHEJ and HR as described in Section 5.1.2. I have taken advantage of this mechanism and combined it with the AAV targeting strategy detailed in Chapter 4, which provides an rAAV donor to for integration into the Cas9 cleavage site, producing a gene editing system with increased targeting efficiency. The steps of this approach are which are outlined in Figure 5.3.

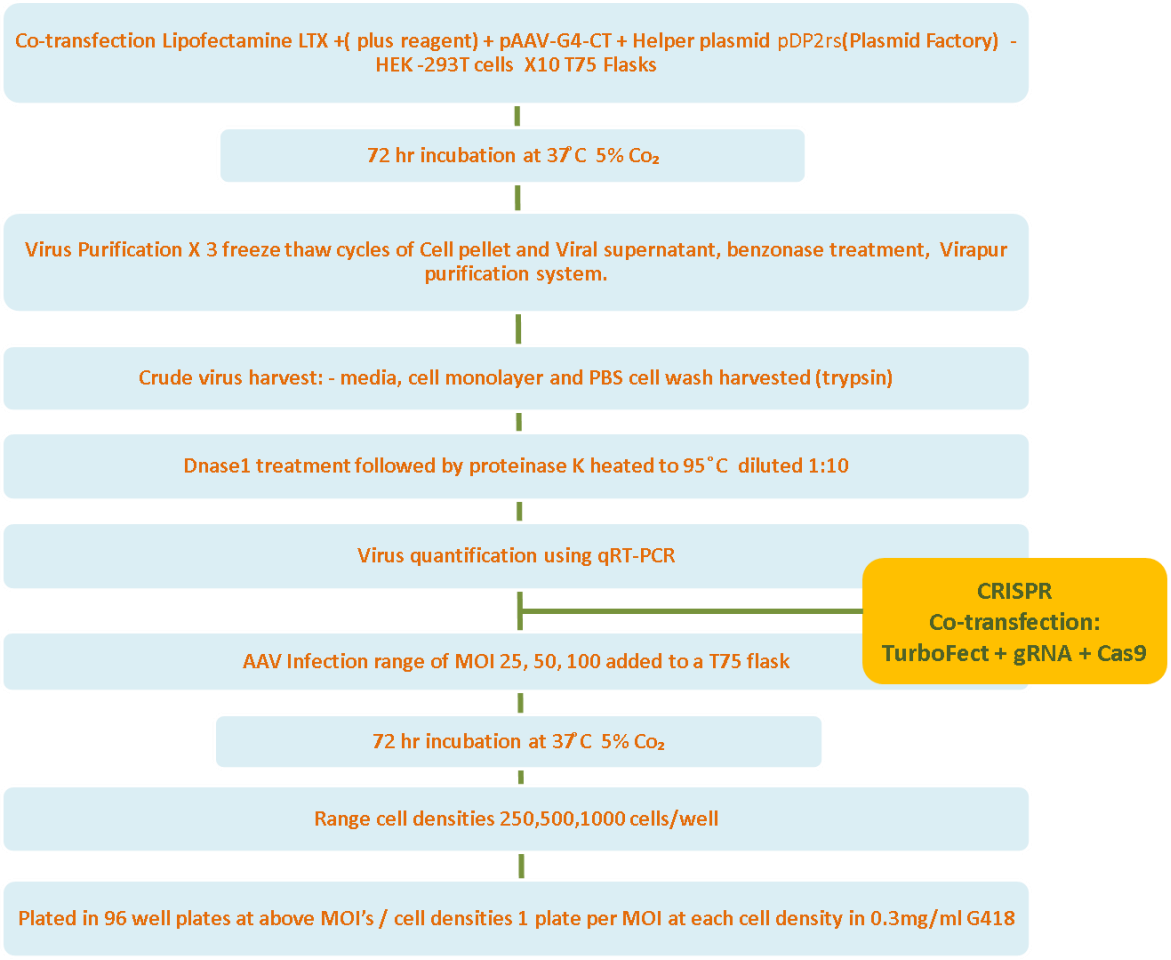


Figure 5.3 Modified gene editing strategy; CRISPR & AAV to KO Mir137

Outline of the stages involved in the CRISPR and AAV gene editing strategy, targeting SH-SY5Y cells with CRISPR-Cas plasmids prior to AAV infection.

5.3.2 Overview of CRISPR-Cas gene editing system to KO Mir137.

In order to effectively target a specific locus using the CRISPR-Cas system a number of design, implementation and verification steps are involved as outlined in Figure 5.4.

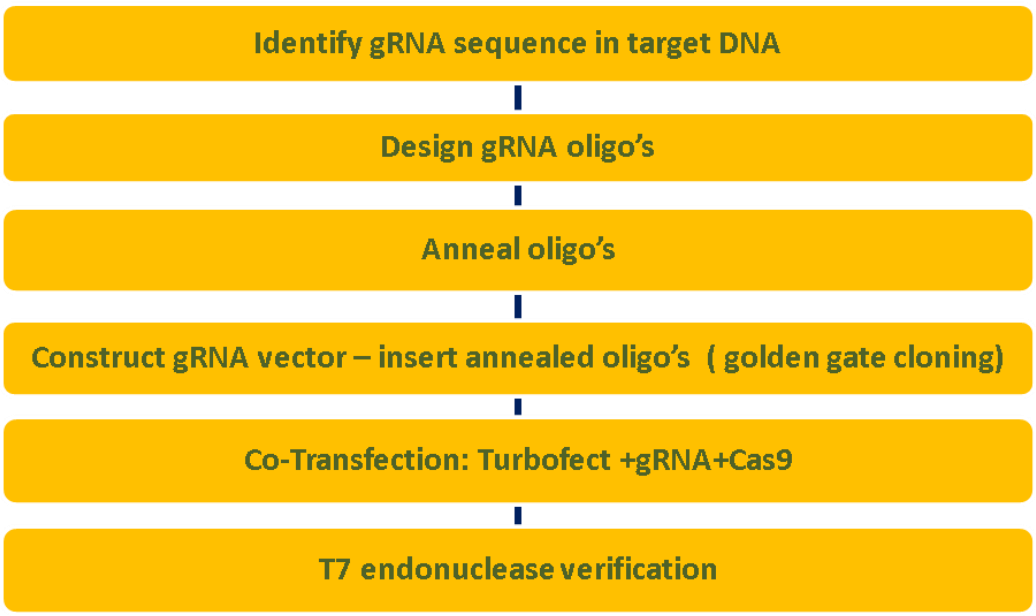


Figure 5.4 Stages involved in CRISPR-Cas mediated gene targeting.

5.3.3 Mir-137 gRNA oligonucleotide design.

As detailed in Section 5.1 CRISPR-Cas systems are readily retargeted by expressing or delivering appropriate gRNAs (Barrangou et al., 2007, Brouns et al., 2008, Hale et al., 2012, Jiang et al., 2013). I therefore identified an appropriate consensus target sequence G n^{20} GG within the 102 bp Mir-137 sequence (Figure 5.5). Once the consensus target sequence was identified I designed the appropriate sequence to be inserted into the gRNA vector to enable recognition of the Mir-137 by the Cas9 nuclease supplied in *trans* (Figure 5.5). This recognition sequence was constructed by annealing 2 complementary oligonucleotides also identified in Figure 5.5.

MiR-137: Chromosome 1, 98511626 98511727 - 102bp
GGTCCTCTGACTCTCTTCGGTGACGG***G***<u>*TATTCTTGGGTGGATAATA*</u>***CGG***AT
TA CGTTGTTATTGCTTAAGAATACGCGTAGTCGAGGAGAGTACCAGCGGCA

Consensus target sequence	5’ **G**nnnnnnnnnnnnnnnnnn**NGG** 3’	20 n’s
Required RNA sequence	5’ nnnnnnnnnnnnnnnnnn 3’	19 n’s
Forward oligo	5’ nnnnnnnnnnnnnnnnnnGT 3’	
Reverse oligo	3’ GCnnnnnnnnnnnnnnnnnn 5’	
*Forward oligo 5’->3’	5’ TATTCTTGGGTGGATAATAGT 3’	
Reverse oligo	3’ GCATAAGAACCCACCTATTAT 5’	
*Reverse oligo 5’->3’	5’ TATTATCCACCCAAGAATACG 3’	

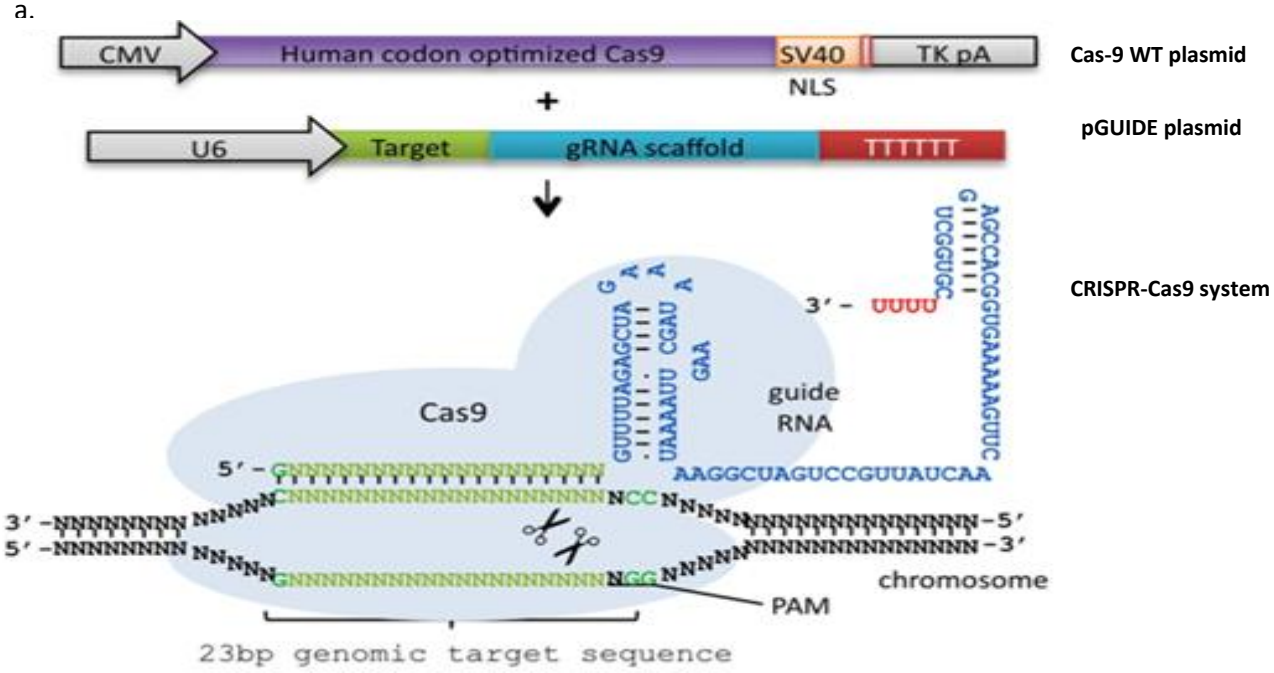
Figure 5.5 Identification of CRISPR-Cas gRNA recognition sequence in Mir-137

Mir-137 sequence chromosome location identified, gRNA recognition sequence italic and underlined. gRNA insert design identifying forward and complementary reverse oligonucleotides synthesized to be annealed.

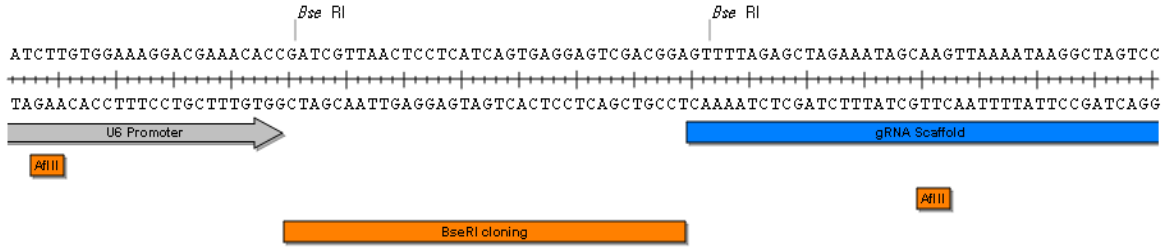
5.3.4 Inserting gRNA into pGUIDE

Golden Gate cloning (Section 2.2.3.2) was used as strategy to insert gRNA (annealed oligos) into the pGUIDE vector (kind gift from Dr. Patrick Harrison University College Cork) rather than gel purifying the backbone. Golden gate cloning makes use of two type II enzyme sites in this case *Bse*RI, treatment of pGUIDE plasmid with *Bse*RI liberates a short insert from the pGUIIDE vector (Figure 5.6b), which is outcompeted by the Mir-137 gRNA insert (annealed oligos) (Figure 5.6c).

a.



b.



c.

```
Consensus target sequence 5′  Gnnnnnnnnnnnnnnnnnnn NGG 3′

Required RNA sequence     5′   nnnnnnnnnnnnnnnnnnn    3′

Forward oligo             5′   nnnnnnnnnnnnnnnnnnnGT  3′
Reverse oligo             3′ GCnnnnnnnnnnnnnnnnnnn    5′
```

```
5′-TTGTGGAAAGGACGAAACACCG  ATCGTTAACTCCTCATCAGTGAGGAGTCGACGGAGT TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG-3′
3′-AACACCTTTCCTGCTTTGTG  GCTAGCAATTGAGGAGTAGTCACTCCTCAGCTGCCT CAAAATCTCGATCTTTATCGTTCAATTTTATTCCGATC-5′
```

```
5′-TTGTGGAAAGGACGAAACACCGnnnnnnnnnnnnnnnnnnnnGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG-3′
3′-AACACCTTTCCTGCTTTGTGGCnnnnnnnnnnnnnnnnnnnnCAAAATCTCGATCTTTATCGTTCAATTTTATTCCGATC-5′
```

Figure 5.6 Inserting gRNA into pGUIDE

schematic of the Cas9 WT plasmid plus pGUIDE and the complex they form once co-transfected.(adapted from (Mali et al., 2013c) b. pGUIDE plasmid with liberated segment (orange), upstream of the U6 promoter(grey) and downstream of the gRNA scaffold (blue) post *Bse*RI treatment. c. sequences describing the golden gated cloning approach to inserting required RNA (green) into the pGUIDE plasmid required for target recognition by pGUIDE colour coded to correlate with b.

5.3.5 CRISPR-Cas transfection optimisation

CRISPR-Cas targeting relies on successful co-transfection of the guide RNA plasmid (pGuide) and Cas9 plasmid (Cas-9 WT). In order to ensure maximum efficiency of CRISPR targeting, co-transfection efficiency of SH-SY5Y with TurboFect was optimised with a GFP plasmid. 12 well plates were seeded with cells and a range of transfection conditions were applied Figure 5.7. GFP transfections revealed 3µg DNA and 12µl TurboFect gave the optimum transfection efficiency of 40%.

GFP1 µg	GFP2 µg	GFP3 µg
400 µl DMEM	400 µl DMEM	400 µl DMEM
4µl GFP	8µl GFP	12µl GFP
4µl Turbofect	8µl Turbofect	12µl Turbofect

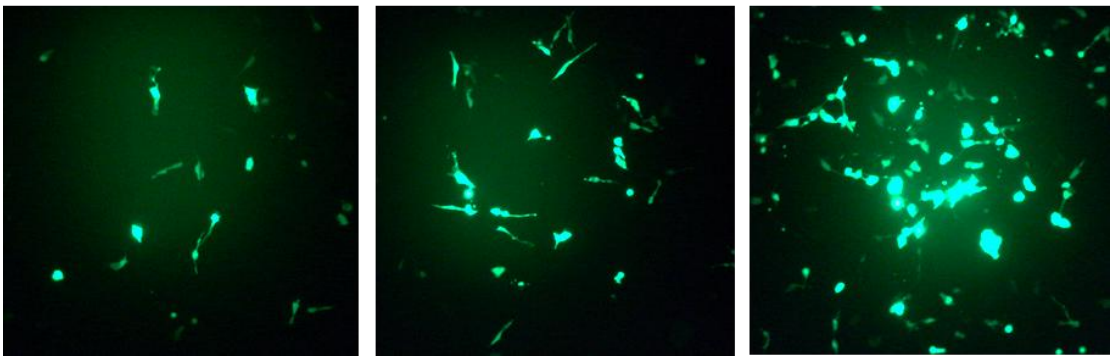


Figure 5.7 Transfection optimisation

SH-SY5Y cells transfected with a range of DNA concentrations and transfection reagent (TurboFect) amounts to determine the optimum transfection conditions to be applied to CRISPR-Cas co-transfcetion.

5.3.6 Verification of CRISPR-Cas targeting efficiency.

SH-SY5Y cells were co-transfected with equimolar Mir-137-pGuide and Cas9-WT, at a range of; 1 µg, 2 µg, 3µg (Cas9-WT plasmid). It was essential to determine efficiency of the CRISPR-Cas targeting before proceeding with AAV infection as not to waste time or resources. PCR and the T7 endonuclease 1 (T7E1) assay were used to determine the efficiency of CRISPR-Cas targeting (Sander and Joung, 2014).

CRISPR-Cas targeting introduces double strand breaks at or around the target region (Mir-137), the natural cellular repair process of NHEJ will repair these double strand breaks however it is an error prone ligation process that often results in small insertions and deletions at cleavage sites. PCR primers were designed to amplify a product of 538bp with the target sequence located within the amplicon (Figure 5.8a). The target region was amplified through PCR (Figure 5.8b) in each of the co-tranfection conditions, a % of the amplicons contained mutant sequences due to the NHEJ, these are the ampicons have been successfully targeted. To detect the % of targeted amplicons, the 538bp PCR amplicons (Figure 5.8b) were gel purified, denatured and allowed to re-anneal (Section 2.2.3.1). Due to the % of mutated sequences (NHEJ errors) the strands are not all complementary and are thus unable to form hydrogen bonds. Those where NHEJ has taken place will form heteroduplex DNA. The annealed amplicons are then treated with mismatch-sensitive T7E1 which will digest at the sites of flipped out base pairs (hetroduplex DNA). When analysed on a 2% agarose gel through electrophoresis the T7E1 treatment of heteroduplex DNA; CRISPR-Cas treated group gave rise to 320bp and 208 bp DNA fragments which when combined with the % of untargeted amplicons appear as 3 bands on the agarose gel (Figure 5.8c). The positive control Promega DNA isolated from a pool of

individuals with SNP's also appeared as 3 bands on the gel while the T7E1 treatment of heteroduplex WT SH-SY5Y gDNA gave rise to 1 band acting as the negative control. 1ug of DNA failed to show any targeting events , supported by the low GFP transfection efficiency (Figure 5.8c), 2µg gave moderate targeting 5% once again supported by the increased transfection efficiency, 3µg resulted in optimum targeting of 15% supporded by the GFP transfection (Figure 5.8c)

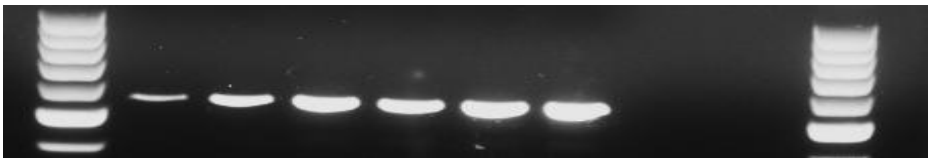
a.

MIR137 precursor sequence

>chr1:98511564-98512101 538bp CAGAGGAAAGCACTGGGAGA CTTGGCAACCACGGGCGTTTA

CAGAGGAAAGCACTGGGAGAgcaccaggtaaactgaaggttacttgtcactcccacttgtgcccaaaaagccttg
ccacatcttccctcctcactggaaagacagcactcttctgtgttaagtatttgattttgtgatttgtctttcaga
attggaaatagagcggccatttggatttgggcaggaagcagccgagcacagctttggatccttctttagggaaat
cgagttatggatttatggtcccggtcaagctcagcccatccccaggcaggggcgggctcagcgagcagcaagagt
tctggtggcggcggcggcggcagtagcagcggcagcggtagcagcggcagcggcagcttg
gtcctctgactctcttcggtgacggtattcttgggtggataatacggattacgttgttattgcttaagaatacg
cgtagtcgaggagagtaccagcggcaggggggcagcggccgccctccccagcccaccagctggccacTAAACGCC
CGTGGTTGCCAAG

b.



c.

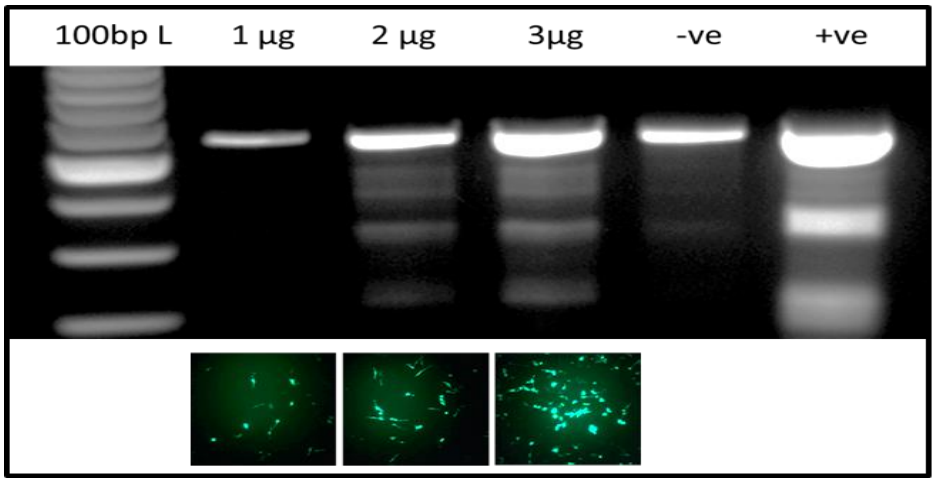


Figure 5.8 Verification of CRISPR-Cas targeting efficiency

PCR and mismatch sensitive enzyme T7E1 used to verify CRISPR-Cas targeting efficiency at the Mir-137 locus. a. PCR primers and chromosome location of Mir-137 along with amplicon sequence amplified via PCR containing Mir-137 target sequence (highlighted yellow) required consensus sequence (highlighted red) and PCR primers (highlighted blue). b. gDNA extracted from co-transfceted cells and PCR amplified with primers outlined in Figure a. verified through gel electrophoresis. c. PCR products shown in Figure b, gel purified, denatured, annealed, treated with mismatch sensitive T7E1 enzyme and analysed on a 2% agarose gel, the increased targeting efficiency identified through PCR reflects the increased GFP expression at the range of concentrations observed in Figure 5.7.

5.3.7 rAAV-Mir-137KO infection.

Following verification of successful CRISPR-Cas targeting to introduce to double strand breaks in turn triggering the natural NHEJ pathway, cells were infected with AAV viral particles generated by a former MRes student of the Quinn Lab Daniel Grifiths. Infection was carried out at MOI's of 25, 50 and 100 following the protocol outlined in Section 4.3.3.2. 72 hrs following infection cells were plated into 96 well plates at densities of 250, 500, 100 cells per well, each cell density was infected at an MOI of 25, 50, and 100, as summarized in Table 5.1

MOI 25	MOI 50	MOI 100
250 cells / well	250 cells / well	250 cells / well
500 cells / well	500 cells / well	500 cells / well
1000 cells / well	1000 cells / well	1000 cells / well

Table 5.1 Range of cell to virus ratio MOI and range at which infected cells were plated (cell / well)

5.3.8 PCR screen of AAV-Mir137-KO verifying positive recombination events

To identify positive recombination events, pools of clones were PCR screened using 2 separate primer sets. The first primer set has its forward primer in the LHA and its reverse primer in the neomycin resistance cassette amplifying a PCR product of 513bp (+ve1, green arrows Figure 5.9). This was a very efficient PCR reaction optimised prior to infection using the Mir137-AAV targeting construct as DNA template. It was designed as a 1st round screen to identify recombination events amongst the 2592 infected wells allowing us to considerably reduce tissue culture work, only maintaining those wells which contained recombinant DNA (Figure 5.10). Although this was a fast and efficient PCR the amplicon could have been generated from episomal DNA. Therefore positives identified in the 1st round were put through a 2nd round of screening using a separate set of primers (+ve2 red arrows Figure 5.9) with the forward primer outside the Left homology arm and the reverse nested in the neomycin cassete giving a PCR product of 2093 bp.

Figure 5.9 AAV-Mir137-KO PCR screen primer design and strategy.

Primer sets to identify positive recombination events, the 1st round primer set (green) amplified PCR products of 513bp and the 2nd round primer set amplified a product 2093bp (red).

5.3.9 rAAV-Mir-137KO PCR screen.

AAV-Mir-137-KO recombination events were verified by PCR screening gDNA extracted from plated SH-SY5Y cells. Of the 2592 wells plated with infected SH-SY5Y (96wells at each range detailed in Section 5.3.7) 1867 wells grew colonies in total. Positive primer set 1 detailed in Figure 5.9 (primer sequences can be found in table 2 of the Appendix) was used to amplify a PCR product of 513bp. The 96 well plate with an MOI of 100 and cell density of 1000 cells per well had 88 colonies form and was the 1st plate to reach a suitable confluency of 50% to withstand direct PCR liysis (Section 2.2.13.6.2.1). It was therefore selected to be screened as outlined in Section 2.2.13.6. with 64 clones positive for recombination identified from the 96 well plate (Figure 5.10).

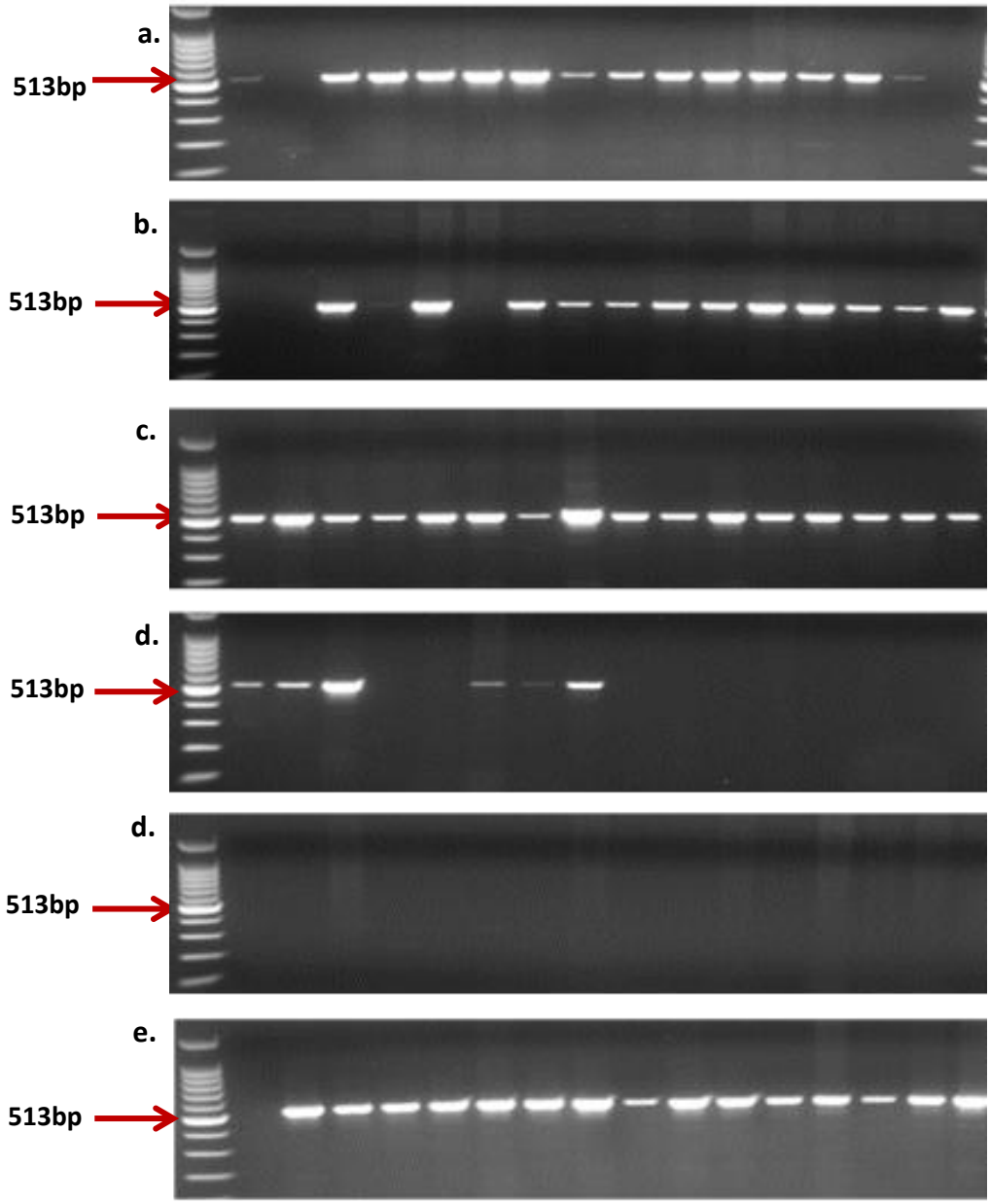


Figure 5.10 AAV-Mir-137-KO PCR screen

Agarose gel showing 64 positive recombination events by amplification of a 513bp fragment from gDNA extracted from targeted SH-SY5Y cells.

5.3.10 rAAV-Mir-137KO verification.

Of the 64 positive pools of clones identified, 9 were selected to be expanded into 48 well plates followed by SCD in order to obtain a single colony to ensure a truly isogenic cell line. At the time of SCD the 9 selected pools of clones were re screened to ensure they were still positive for recombination (Figure 5.11).

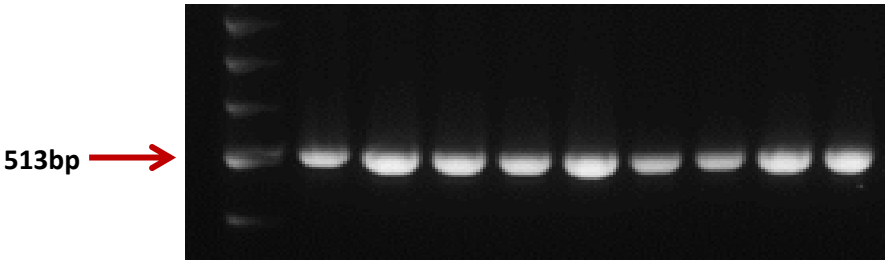


Figure5.11 rAAV-Mir-137 KO verification

Agarose gel showing positive clones selected for expansion and SCD were rescreened with original positive primer 1 amplifying a product of 513 bp

5.3.11 Positive PCR screen of single cell diluted Mir-137 KO clone to verify hemizygous KO

Following verification that all 9 selected pools were still positive for recombination after expansion up to the 48 cell stage, they were SCD (Section 2.2.13.6.5) and incubated for 8 weeks at 37 ˚C. Once single colonies formed and reached over 50% confluency in the 96 well plates, 3 colonies were selected to be re-screened with a new primer set (Figure 5.12a) identifying both alleles of the hemizyhgous KO (Figure 5 .12b). This primer set had a forward primer in the left homology arm and a reverse primer in the right homology arm. Therefore both the targeted and WT allele were amplified. Amplification of the targeted allele produced a 1911bp PCR product including the neomycin selection cassette and amplification of the WT allele produced a 545bp PCR product without the neomycin selection cassette (Figure 5.15).

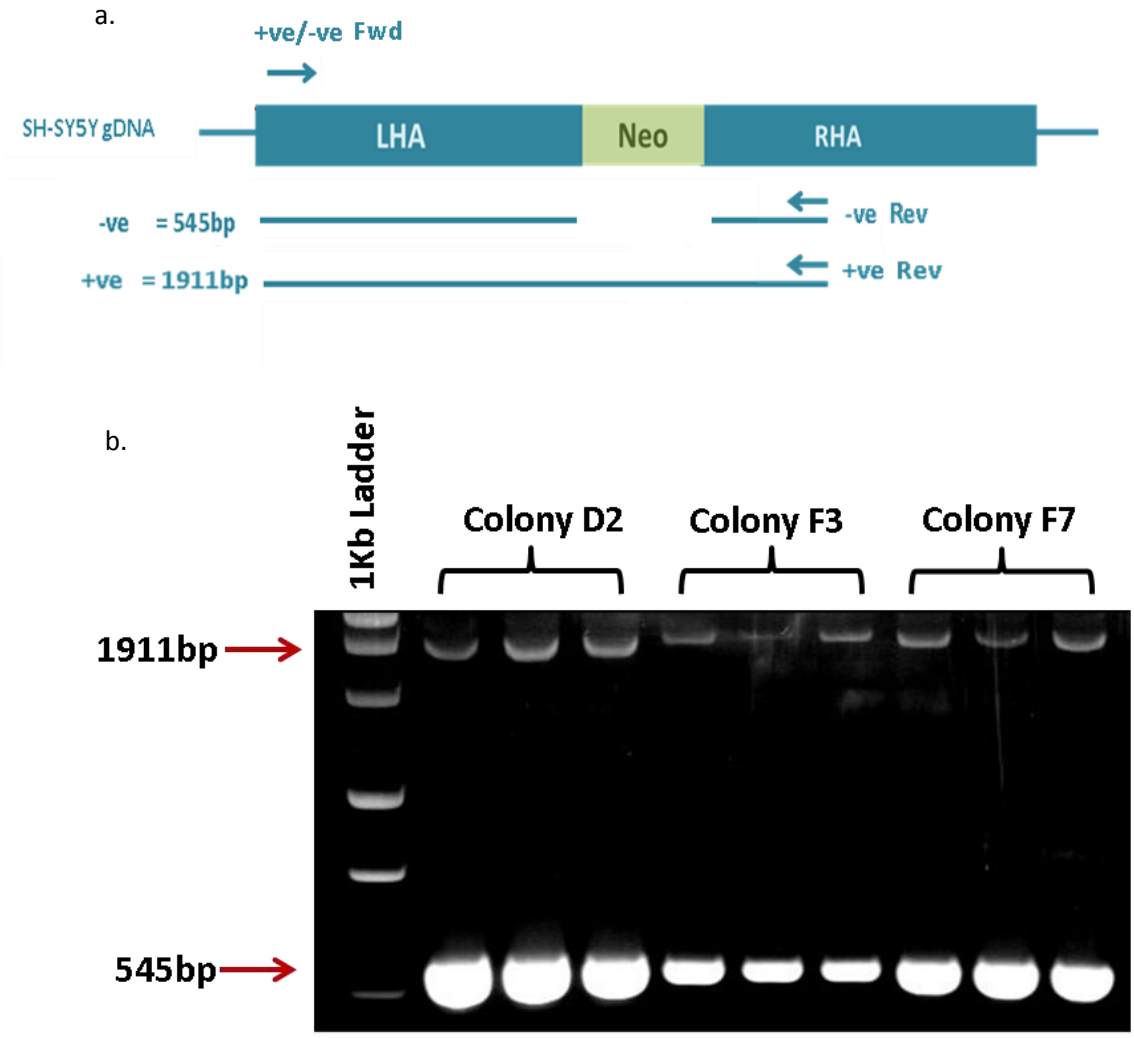


Figure 5.12 Positive PCR screen of single cell diluted Mir-137 KO clone

a. Schematic representation of primer set used to screen for hemizygous KO of Mir-137, 545 pb PCR product for the WT allele and 1911bp PCR product for the targeted allele. b Agarose gel showing products from a PCR reaction in triplicate for each of the 3 selected SCD colonies, using primer set identified in a.

5.3.12 Confirmation of positive PCR screen for single cell diluted Mir-137 KO clone.

SCD clones were screened with primer set 2 (Figure 5.9) to ensure the Mir-137 KO recombination event was not episomal.

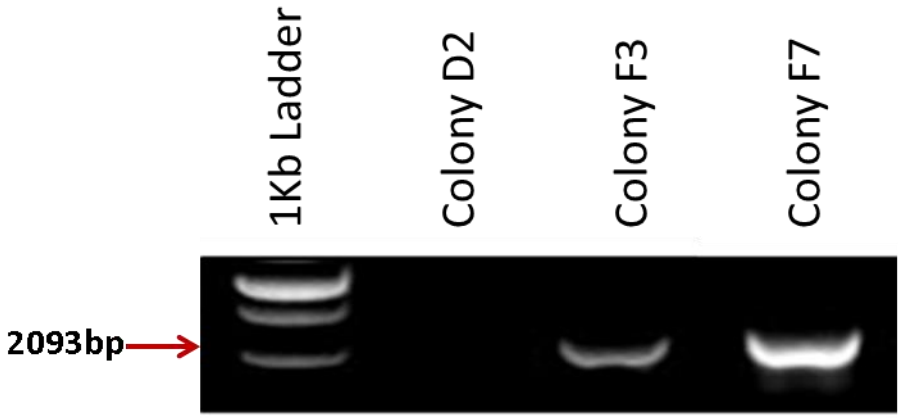


Figure 5.13 Confirmation of positive PCR screen for single cell diluted Mir-137 KO clone.
Agarose gel showing PCR fragments of 2093bp confirming positive targeting events in colony F3 and colony F7 following single cell diluition of the Mir-137 KO.

5.4 Discussion

In this chapter I have presented an efficient gene editing strategy, combining the CRISPR-Cas system with traditional AAV targeting to create a hemizygous KO of the microRNA Mir137 in SH-SY5Y cells.

As described in Section 5.1 the CRISPR-Cas systems have gained much attention as a transformative gene editing technology, owing to their low cost, high efficiency, simplicity and broad range of application (Mali et al., 2013b, Sander and Joung, 2014). However there is concern regarding off target effects of CRISPR, (reviewed in Cho et al., 2014), although these are currently being addressed (Shen et al., 2014). rAAV gene targeting is a well validated approach to gene editing (Russell and Hirata, 1998) and is known to be a high precision technique, stimulating homologous recombination directly at 100x higher efficiency than plasmids (Russell and Hirata, 1998). rAAV targeting is however a low throughput technique determined by the vast time and resources required to screening and identify correctly targeted clones.

By combining the two technologies in one application it was possible to significantly increase targeting efficiency while maintaining the high fidelity of AAV (to be confirmed). By co-transfecting with CRISPR-Cas plasmids designed to target the Mir-137 locus (Section5.3.3, Figure 5.8a and Figure 5.14b) double strand breaks were introduced (Figure 5.14c) which can be repaired by one of two natural cellular processes the error prone NHEJ or the high fidelity HR (Figure 5.14d) . Following the co-transfection of SH-SY5Y cells with CRISPR-Cas plasmids (Mir137-gRNA and Cas9WT) cell were infected with rAAV-Mir-137KO which provided the homologous donor DNA (Figure 5.14e), to reengineer the locus to delete Mir-137 via homologous directed repair (HDR).

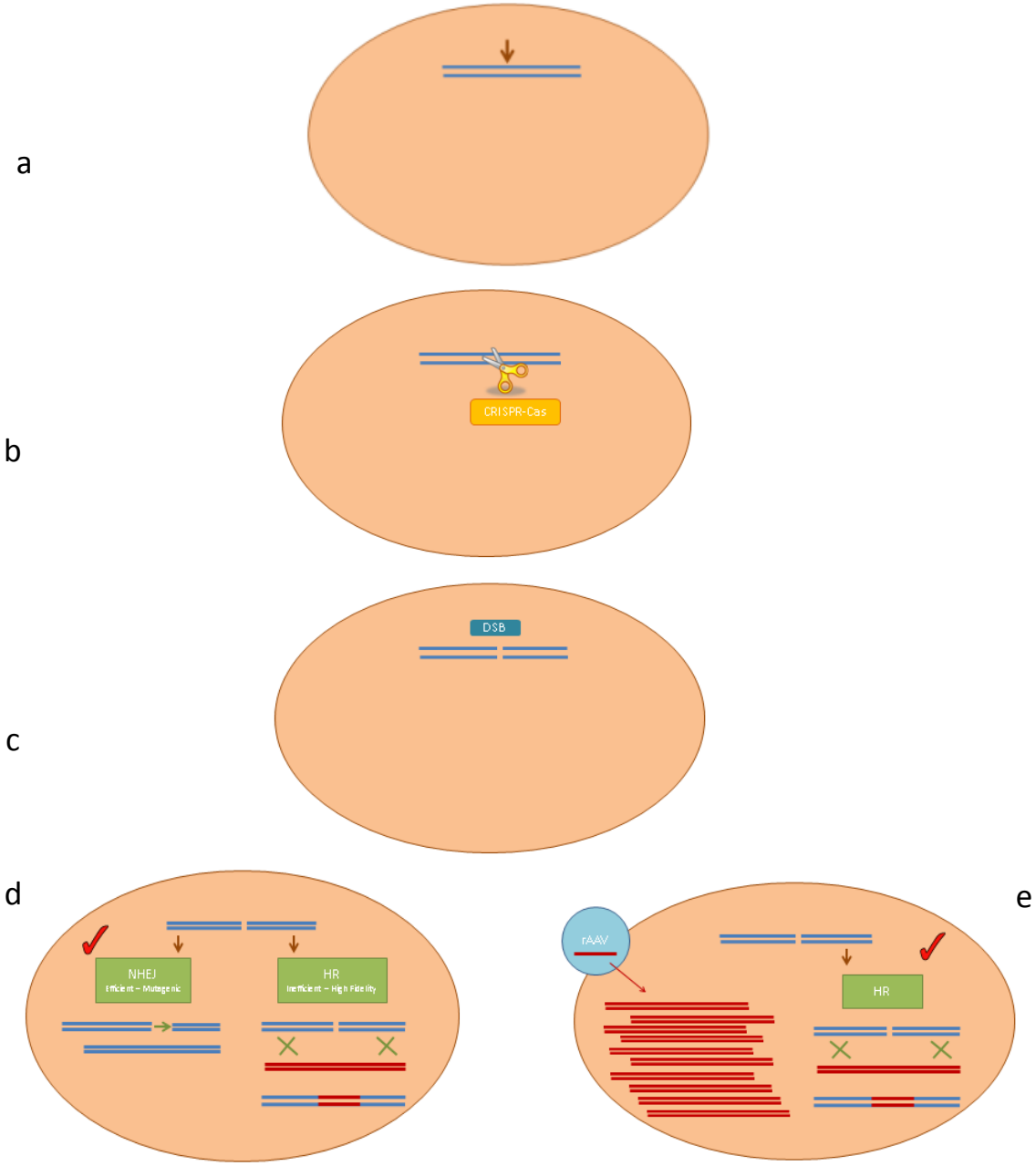


Figure 5.14 A strategy to KO Mir-137 using CRISPR-Cas & AAV.

A schematic diagram depicting the series of events occurring inside the nucleus during CRISPR-Cas and AAV targeting. a.Mir-137 locus b. CRISPR-Cas targeting c. introduction of double strand breaks d. two natural cellular processes available to repair the DSB's the error prone NHEJ or the high fidelity HR. e. rAAV-Mir-137KO acts as homologous donor DNA shifting the preference of repair to homologous directed repair (HDR)

The targeting efficiencies of all approaches used throughout the study are outlined in Table 5.2.

	Colony formation	+ve recombination prior to SCD	+ve recombination post SCD	+ve Insertion into gDNA
rAAV-G4/LoxP-CT-1st	**0%**	**N/A**	**N/A**	**N/A**
rAAV-G4/LoxP-CT-2nd	**21%**	**0.2%**	**0.03%**	**0.03%**
rAAV-Mir-137	**32%**	**0%**	**N/A**	**N/A**
CRISPR & rAAV-Mir-137	**72%**	**67%**	**100%**	**66%**

Table 5.2 Comparison of targeting efficiencies between various approaches used.

In addition to the overall increased targeting frequency, it should be noted that the combined CRISPR & AAV approach significantly reduces the time and cost of targeting. I estimate an initial saving of £1800 in tissue culture and PCR screening resources in addition to 3 months salaried time. This is primarily due to the fact that as a result of the increased targeting efficiency, the 1st 96 well plate to reach 50% confluency was screened and contained enough positive recombination events to expand and SCD. Therefore preventing the need to wait for and screen colonies forming within the remaining 2496 wells. Based on this information, further savings could be made as the amount of rAAV virus produced and the range of MOI and cell

densities could be further reduced, to creating a further streamlined yet robust gene editing system.

A further consideration which has been highlighted through this combined targeting strategy is the potential to further increase targeting frequency through a second round of CRISPR-Cas co-transfcetion to make use of the episomal AAV. Following infection with rAAV a proportion of the rAAV will remain as a stable episome (Duan et al., 1998), episomal AAV transgene expression has been observed for up to one year (Haidet et al., 2008, Le Hir et al., 2013). During this time cells could undergo a second round of CRISPR-Cas cotransfection which would re-target the loci, of either those cells which had failed to be targeted in the 1st round or the untargeted allele, as only these untargeted cells / alleles would possess the complement sequence for recognition by the gRNA. This would introduce ds breaks and the episomal rAAV would be the homologous donor sequence thus re-instigating HDR to reengineer the region as desired, this also has implications in homozygous targeting as it could increase the chance of a homozygous recombination event. At this time a study that has not been done.

Appendix

Primers	Applications	Primers sequence	Region amplified	Expected product size	PCR conditions
5' LPR VNTR_Fwd 5' LPR VNTR_rev	ChIP Screening	5' ATGCCAGCACCTAACCCCTAATGT 3' 5' GGACCGCAAGGTGGGCGGGA 3'	SLC6A4 5' LPR VNTR chr17:28564079- 28564497	419bp + VNTR length	Touchdown PCR: 1 cycle: 94°C 3 (min) 3 cycles: 94°C 15 (sec) 64°C 30 (sec) 70°C 2 (min) 3 cycles: 94°C 15 (sec) 61°C 30 (sec) 70°C 2 (min) 3 cycles: 94°C 15 (sec) 58°C 30 (sec) 70°C 2 (min) 3 cycles: 94°C 15 (sec) 58°C 30 (sec) 70°C 2 (min) 35 cycles: 94°C 15 (sec) 57°C 30 (sec) 70°C 2 (min) 1 cycle: 70C 5 (min)
NHE F NHE R	ChIP Screening & Cloning	5' **atggcgcgccggtacc**tcacaggctgcgagtaga 3' 5' **aacggcgcgcctcga**gtggcctctcaagagga 3'	CT element c-MYC	550bp	1 cycle: 96°C (3 min) 35 cycles: 95°C (1 min), 54°C (1min), 72°C (1min)
MAOA proximal F MAOA Proximal R	ChIP screening mRNA analysis	5' ACAGCCTGACCGTGGAGAAG 3' 5' TCCGAATGGAGCGTCCGTTC 3'	MAOA µVNTR chrX:43514243+43514566	324 bp	1 cycle: 96°C (3 min) 36 cycles: 94°C (1 min), 55°C (1 min), 72°C (30 sec)

Primers	Applications	Primers sequence	Region amplified	Expected product size	PCR conditions
MAOA Distal F MAOA Distal R	ChIP screening mRNA analysis	5’ GGGTTAAGCGCCTCAGCTTG 3’ 5’ CTGCTTCCTTAAGTCCACTCTTG 3’	MAOA Distal VNTR chrX:43513612+43513976	365 bp	Touchdown PR as for 5’LPR above
CT-AAV-WT-Fwd CT-AAV-WT-Fwd	AAV screening	5’ gcgaccttgcttgccctctct 3’ 5’ tctctgggtgcttggatttgtggat 3’	CT element LHA+gDNA	1141 bp	1 cycle: 96°C (3 min) 36 cycles: 94°C (1 min), 59°C (30 sec), 72°C (1min)
CT-AAV-+ve-Fwd CT-AAV-+ve-Fwd	AAV screening	5’actgggcactgtcgcaggtc 3’ 5’ catggccggcgaactgggct 3’	CT element gDNA + Neo	1810 bp	1 cycle: 96°C (3 min) 35 cycles: 94°C (30 sec), 59°C (30 sec), 72°C (90 sec)
Mir-137-CRISPR_Fwd Mir-137-CRISPR_Rev	CRISPR-gRNA	5’TATTCTTGGGTGGATAATAGT 3’ 5’ TATTATCCACCCAAGAATACG 3’	Mir-137	102 bp	1 cycle: 95°C (5 min)
Mir-137 Fwd Mir-137 Rev	CRISPR-T7E1	5’ CAGAGGAAAGCACTGGGAGA 3’ 5’ CTTGGCAACCACGGGCGTTTA 3’	chr1:98511564-98512101 Mir-137	538bp	1 cycle: 96°C (3 min) 35 cycles: 94°C (30 sec), 62°C (30 sec), 72°C (30 sec)
Mir-137-AAV_pos1F Mir-137-AAV_pos1R	CRISPR & AAV screen	5’ TTGTCACTCCCACTTGTGCC 3’ 5’ GTGGATGTGGAATGTGTGCG 3’	Mir-137 LHA + Neo	513 bp	1 cycle: 96°C (3 min) 35 cycles: 94°C (30 sec), 60°C (30 sec), 72°C (30 sec)
Mir-137-AAV pos 1 F2 Mir-137AAV pos 1 R	CRISPR & AAV screen	5’ TGTGCCTCCCTTGACCATC 3’ 5’ GTGGATGTGGAATGTGTGCG 3’	Mir-137 gDNA + Neo	2093 bp	Touchdown PR as for 5’LPR above
Mir-137-AAV_negF Mir-137-AAV_negR	CRISPR & AAV screen	5’ TTGTCACTCCCACTTGTGCC 3’ 5’GAACCAGTGCGAAAACACCC 3’	Mir-137 LHA +/- Neo	545/ 1911 bp	1 cycle: 96°C (3 min) 35 cycles: 94°C (30 sec), 60°C (30 sec), 72°C (90 sec)

Table 6.1: Primer sequences and PCR conditions used during this study.

Vector	Discription
pAAV-G4-CT	pAAV-TK-accepter with 1.5kb Homology c-MYC 700bp+/-CT element (G4-KO mutated CT element- Figure 4.1b)
pAAV-LoxP-CT	pAAV-TK-accepter with 1.5kb Homology c-MYC 700bp+/-CT element (Figure 4.1b)
pAAV-Sp1-CT	pAAV-TK-accepter with 1.5kb Homology c-MYC 700bp+/-CT element (Sp1-KO mutated CT element- Figure 4.1b)
pAAV-G4-VEGFA	pAAV-TK-accepter with 1.5kb Homology VEGFA700bp+/-CT element (G4-KO mutated CT element- Figure 4.1d)
pAAV-LoxP-VEGFA	pAAV-TK-accepter with 1.5kb Homology VEGFA700bp+/-CT element (Figure 4.1d)
pAAV-Sp1-VEGFA	pAAV-TK-accepter with 1.5kb Homology VEGFA700bp+/-CT element (Sp1-KO mutated CT element- Figure 4.1d)

Table 6.2 Vectors cloned during this study.

Virus	Discription
rAAV-G4-CT	rAAV-TK-accepter with 1.5kb Homology c-MYC 700bp+/-CT element (G4-KO mutated CT element- Figure 4.1b)
rAAV-LoxP-CT	rAAV-TK-accepter with 1.5kb Homology c-MYC 700bp+/-CT element (Figure 4.1b)
rAAV-Mir-137 KO	rAAV-TK-accepter with 1.5kb Homology c-MYC 700bp+/-CT element (Sp1-KO mutated CT element- Figure 4.1b)

Table 6.3 Virus generated during this study

Antibody	Clonality	Host species	Immunogen	Source	Catalogue Number
anti-CTCF	Monoclonal	Mouse	His-tagged recombinant protein corresponding to human CTCF	Millipore	17-10044
anti-Histone H3	Polyclonal	Rabbit	Synthetic peptide conjugated to KLH derived from within residues 100 to the C-terminus of Human Histone H3	abcam	ab1791
anti-Nucleolin	Monoclonal	Mouse	Human nucleolin protein from Raji cell extract.	abcam	ab13541
anti-Histone H3 (tri methyl K9)	Polyclonal	Rabbit	Synthetic peptide conjugated to KLH derived from within residues 1 - 100 of Human Histone H3, tri methylated at K9.	abcam	ab8898
anti-hnRNP K	Polyclonal	Rabbit	Synthetic peptide corresponding to a sequence from the C-terminus of isoform a of human hnRNP K. (NP_002131.2)	abcam	ab70492
Anti-CNBP	Polyclonal	Goat	Synthetic peptide: GESGHLARECTIE , corresponding to C terminal amino acids 162-174 of Human CNBP.	abcam	ab48027
anti-RNA pol II CTD phospho Ser5	Polyclonal	Rabbit	This RNA pol II CTD phospho Ser5 antibody was raised against synthetic peptide containing the RNA Pol II heptad repeat consensus sequence phosphorylated at serine 5.	Active-Motif	39233

Table 6.4 Antibodies used in Chromatin Immunoprecipitation (ChIP).
Clonality, host species, target peptide or protein region, source and catalogue number used in ChIP reactions are shown.

References

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.

AMRANE, S., KERKOUR, A., BEDRAT, A., VIALET, B., ANDREOLA, M. L. & MERGNY, J. L. 2014. Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development. *J Am Chem Soc,* 136**,** 5249-52.

AOYAGI, S., NARLIKAR, G., ZHENG, C., SIF, S., KINGSTON, R. E. & HAYES, J. J. 2002. Nucleosome remodeling by the human SWI/SNF complex requires transient global disruption of histone-DNA interactions. *Mol Cell Biol,* 22**,** 3653-62.

ARANGO, V., UNDERWOOD, M. D. & MANN, J. J. 2002. Serotonin brain circuits involved in major depression and suicide. *Prog Brain Res,* 136**,** 443-53.

ARMAS, P., AGUERO, T. H., BORGOGNONE, M., AYBAR, M. J. & CALCATERRA, N. B. 2008a. Dissecting CNBP, a zinc-finger protein required for neural crest development, in its structural and functional domains. *J Mol Biol,* 382**,** 1043-56.

ARMAS, P., NASIF, S. & CALCATERRA, N. B. 2008b. Cellular nucleic acid binding protein binds G-rich single-stranded nucleic acids and may function as a nucleic acid chaperone. *J Cell Biochem,* 103**,** 1013-36.

BARANELLO, L., LEVENS, D., GUPTA, A. & KOUZINE, F. 2012. The importance of being supercoiled: how DNA mechanics regulate dynamic processes. *Biochim Biophys Acta,* 1819**,** 632-8.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 315**,** 1709-12.

BECKER, P. B. 2002. Nucleosome sliding: facts and fiction. *EMBO J,* 21**,** 4749-53.

BERBERICH, S. J. & POSTEL, E. H. 1995. PuF/NM23-H2/NDPK-B transactivates a human c-myc promoter-CAT gene via a functional nuclease hypersensitive element. *Oncogene,* 10**,** 2343-7.

BHAYA, D., DAVISON, M. & BARRANGOU, R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet,* 45**,** 273-97.

BIFFI, G., TANNAHILL, D., MCCAFFERTY, J. & BALASUBRAMANIAN, S. 2013. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem,* 5**,** 182-6.

BIRD, A. P. 1986. CpG-rich islands and the function of DNA methylation. *Nature,* 321**,** 209-13.

BLAKELY, R. D., BERSON, H. E., FREMEAU, R. T., JR., CARON, M. G., PEEK, M. M., PRINCE, H. K. & BRADLEY, C. C. 1991. Cloning and expression of a functional serotonin transporter from rat brain. *Nature,* 354**,** 66-70.

BOEGER, H., BUSHNELL, D. A., DAVIS, R., GRIESENBECK, J., LORCH, Y., STRATTAN, J. S., WESTOVER, K. D. & KORNBERG, R. D. 2005. Structural basis of eukaryotic gene transcription. *FEBS Lett,* 579**,** 899-903.

BOEGER, H., GRIESENBECK, J. & KORNBERG, R. D. 2008. Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell,* 133**,** 716-26.

BOMSZTYK, K., VAN SEUNINGEN, I., SUZUKI, H., DENISENKO, O. & OSTROWSKI, J. 1997. Diverse molecular interactions of the hnRNP K protein. *FEBS Lett,* 403**,** 113-5.

BOYLE, A. P., DAVIS, S., SHULHA, H. P., MELTZER, P., MARGULIES, E. H., WENG, Z., FUREY, T. S. & CRAWFORD, G. E. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell,* 132**,** 311-22.

BRADDOCK, D. T., LOUIS, J. M., BABER, J. L., LEVENS, D. & CLORE, G. M. 2002. Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature,* 415**,** 1051-6.

BRANDEIS, M., FRANK, D., KESHET, I., SIEGFRIED, Z., MENDELSOHN, M., NEMES, A., TEMPER, V., RAZIN, A. & CEDAR, H. 1994. Sp1 elements protect a CpG island from de novo methylation. *Nature,* 371**,** 435-8.

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.

BRENNEMAN, M., GIMBLE, F. S. & WILSON, J. H. 1996. Stimulation of intrachromosomal homologous recombination in human cells by electroporation with site-specific endonucleases. *Proc Natl Acad Sci U S A,* 93**,** 3608-12.

BRIGGS, M. R., KADONAGA, J. T., BELL, S. P. & TJIAN, R. 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science,* 234**,** 47-52.

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.

BROOKS, T. A. & HURLEY, L. H. 2010. Targeting MYC Expression through G-Quadruplexes. *Genes Cancer,* 1**,** 641-649.

BROUNS, S. J., JORE, M. M., LUNDGREN, M., WESTRA, E. R., SLIJKHUIS, R. J., SNIJDERS, A. P., DICKMAN, M. J., MAKAROVA, K. S., KOONIN, E. V. & VAN DER OOST, J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science,* 321**,** 960-4.

CANAFF, L., ZHOU, X. & HENDY, G. N. 2008. The proinflammatory cytokine, interleukin-6, up-regulates calcium-sensing receptor gene transcription via Stat1/3 and Sp1/3. *J Biol Chem,* 283**,** 13586-600.

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.

CASPI, A., SUGDEN, K., MOFFITT, T. E., TAYLOR, A., CRAIG, I. W., HARRINGTON, H., MCCLAY, J., MILL, J., MARTIN, J., BRAITHWAITE, A. & POULTON, R. 2003. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science,* 301**,** 386-9.

CHEN, W., LIANG, Y., DENG, W., SHIMIZU, K., ASHIQUE, A. M., LI, E. & LI, Y. P. 2003. The zinc-finger protein CNBP is required for forebrain formation in the mouse. *Development,* 130**,** 1367-79.

CHENG, S. W., DAVIES, K. P., YUNG, E., BELTRAN, R. J., YU, J. & KALPANA, G. V. 1999. c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. *Nat Genet,* 22**,** 102-5.

CHO, S. W., KIM, S., KIM, J. M. & KIM, J. S. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol,* 31**,** 230-2.

CHO, S. W., KIM, S., KIM, Y., KWEON, J., KIM, H. S., BAE, S. & KIM, J. S. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res,* 24**,** 132-41.

CHUNG, H. J. & LEVENS, D. 2005. c-myc expression: keep the noise down! *Mol Cells,* 20**,** 157-66.

CLUZEL, P., LEBRUN, A., HELLER, C., LAVERY, R., VIOVY, J. L., CHATENAY, D. & CARON, F. 1996. DNA: an extensible molecule. *Science,* 271**,** 792-4.

COOK, P. R. 1995. A chromomeric model for nuclear and chromosome structure. *J Cell Sci,* 108 (Pt 9)**,** 2927-35.

CORE, L. J. & LIS, J. T. 2008. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science,* 319**,** 1791-2.

COWLING, V. H. & COLE, M. D. 2006. Mechanism of transcriptional activation by the Myc oncoproteins. *Semin Cancer Biol,* 16**,** 242-52.

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.

DARZACQ, X., SHAV-TAL, Y., DE TURRIS, V., BRODY, Y., SHENOY, S. M., PHAIR, R. D. & SINGER, R. H. 2007. *In vivo* dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol,* 14**,** 796-806.

DAVIS, T. L., FIRULLI, A. B. & KINNIBURGH, A. J. 1989. Ribonucleoprotein and protein factors bind to an H-DNA-forming c-myc DNA element: possible regulators of the c-myc gene. *Proc Natl Acad Sci U S A,* 86**,** 9682-6.

DENG, C. & CAPECCHI, M. R. 1992. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol Cell Biol,* 12**,** 3365-71.

DENIAUD, E., BAGUET, J., CHALARD, R., BLANQUIER, B., BRINZA, L., MEUNIER, J., MICHALLET, M. C., LAUGRAUD, A., AH-SOON, C., WIERINCKX, A., CASTELLAZZI, M., LACHUER, J., GAUTIER, C., MARVEL, J. & LEVERRIER, Y. 2009. Overexpression of transcription factor Sp1 leads to gene expression perturbations and cell cycle inhibition. *PLoS One,* 4**,** e7035.

DESJARDINS, E. & HAY, N. 1993. Repeated CT elements bound by zinc finger proteins control the absolute and relative activities of the two principal human c-myc promoters. *Mol Cell Biol,* 13**,** 5710-24.

DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. & REN, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature,* 485**,** 376-80.

DOCQUIER, F., FARRAR, D., D'ARCY, V., CHERNUKHIN, I., ROBINSON, A. F., LOUKINOV, D., VATOLIN, S., PACK, S., MACKAY, A., HARRIS, R. A., DORRICOTT, H., O'HARE, M. J., LOBANENKOV, V. & KLENOVA, E. 2005. Heightened expression of CTCF in breast cancer cells is associated with resistance to apoptosis. *Cancer Res,* 65**,** 5112-22.

DU, Z., ZHAO, Y. & LI, N. 2008. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res,* 18**,** 233-41.

DUAN, D., SHARMA, P., YANG, J., YUE, Y., DUDUS, L., ZHANG, Y., FISHER, K. J. & ENGELHARDT, J. F. 1998. Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. *J Virol,* 72**,** 8568-77.

DUNCAN, R., BAZAR, L., MICHELOTTI, G., TOMONAGA, T., KRUTZSCH, H., AVIGAN, M. & LEVENS, D. 1994. A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. *Genes Dev,* 8**,** 465-80.

EDDY, J. & MAIZELS, N. 2006. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res,* 34**,** 3887-96.

EILERS, M. & EISENMAN, R. N. 2008. Myc's broad reach. *Genes Dev,* 22**,** 2755-66.

EISENMAN, R. N. 2001. Deconstructing myc. *Genes Dev,* 15**,** 2023-30.

ELF, J., LI, G. W. & XIE, X. S. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science,* 316**,** 1191-4.

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.

FARLA, P., HERSMUS, R., GEVERTS, B., MARI, P. O., NIGG, A. L., DUBBINK, H. J., TRAPMAN, J. & HOUTSMULLER, A. B. 2004. The androgen receptor ligand-binding domain stabilizes DNA binding in living cells. *J Struct Biol,* 147**,** 50-61.

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

FERNANDEZ, P. C., FRANK, S. R., WANG, L., SCHROEDER, M., LIU, S., GREENE, J., COCITO, A. & AMATI, B. 2003. Genomic targets of the human c-Myc protein. *Genes Dev,* 17**,** 1115-29.

FILIPPOVA, G. N., FAGERLIE, S., KLENOVA, E. M., MYERS, C., DEHNER, Y., GOODWIN, G., NEIMAN, P. E., COLLINS, S. J. & LOBANENKOV, V. V. 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol,* 16**,** 2802-13.

FILIPPOVA, G. N., LINDBLOM, A., MEINCKE, L. J., KLENOVA, E. M., NEIMAN, P. E., COLLINS, S. J., DOGGETT, N. A. & LOBANENKOV, V. V. 1998. A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. *Genes Chromosomes Cancer,* 22**,** 26-36.

FILIPPOVA, G. N., QI, C. F., ULMER, J. E., MOORE, J. M., WARD, M. D., HU, Y. J., LOUKINOV, D. I., PUGACHEVA, E. M., KLENOVA, E. M., GRUNDY, P. E., FEINBERG, A. P., CLETON-JANSEN, A. M., MOERLAND, E. W., CORNELISSE, C. J., SUZUKI, H., KOMIYA, A., LINDBLOM, A., DORION-BONNET, F., NEIMAN, P. E., MORSE, H. C., 3RD, COLLINS, S. J. & LOBANENKOV, V. V. 2002. Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter tts DNA-binding specificity. *Cancer Res,* 62**,** 48-52.

FISKERSTRAND, C. E., LOVEJOY, E., GERRARD, L. & QUINN, J. P. 1999. An intronic domain within the rat preprotachykinin-A gene containing a CCCT repetitive motif acts as an enhancer in differentiating embryonic stem cells. *Neurosci Lett,* 263**,** 141-4.

FLETCHER, T. M., XIAO, N., MAUTINO, G., BAUMANN, C. T., WOLFORD, R., WARREN, B. S. & HAGER, G. L. 2002. ATP-dependent mobilization of the glucocorticoid receptor during chromatin remodeling. *Mol Cell Biol,* 22**,** 3255-63.

FLINK, I. L. & MORKIN, E. 1995. Alternatively processed isoforms of cellular nucleic acid-binding protein interact with a suppressor region of the human beta-myosin heavy chain gene. *J Biol Chem,* 270**,** 6959-65.

FREEMAN, L. A. & GARRARD, W. T. 1992. DNA supercoiling in chromatin structure and gene expression. *Crit Rev Eukaryot Gene Expr,* 2**,** 165-209.

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

GELTINGER, C., HORTNAGEL, K. & POLACK, A. 1996. TATA box and Sp1 sites mediate the activation of c-myc promoter P1 by immunoglobulin kappa enhancers. *Gene Expr,* 6**,** 113-27.

GIBSON, D. G., YOUNG, L., CHUANG, R. Y., VENTER, J. C., HUTCHISON, C. A., 3RD & SMITH, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods,* 6**,** 343-5.

GILBERT, D. E. & FEIGON, J. 1999. Multistranded DNA structures. *Curr Opin Struct Biol,* 9**,** 305-14.

GINISTY, H., SICARD, H., ROGER, B. & BOUVET, P. 1999. Structure and functions of nucleolin. *J Cell Sci,* 112 (Pt 6)**,** 761-72.

GODDE, J. S. & WIDOM, J. 1992. Chromatin structure of Schizosaccharomyces pombe. A nucleosome repeat length that is shorter than the chromatosomal DNA length. *J Mol Biol,* 226**,** 1009-25.

GONZALEZ, V. & HURLEY, L. H. 2010a. The c-MYC NHE III(1): function and regulation. *Annu Rev Pharmacol Toxicol,* 50**,** 111-29.

GONZALEZ, V. & HURLEY, L. H. 2010b. The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry,* 49**,** 9706-14.

GOODRICH, J. A., CUTLER, G. & TJIAN, R. 1996. Contacts in context: promoter specificity and macromolecular interactions in transcription. *Cell,* 84**,** 825-30.

GORMAN, J. & GREENE, E. C. 2008. Visualizing one-dimensional diffusion of proteins along DNA. *Nat Struct Mol Biol,* 15**,** 768-74.

GORSKI, S. A., DUNDR, M. & MISTELI, T. 2006. The road much traveled: trafficking in the cell nucleus. *Curr Opin Cell Biol,* 18**,** 284-90.

GOWERS, D. M., WILSON, G. G. & HALFORD, S. E. 2005. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc Natl Acad Sci U S A,* 102**,** 15883-8.

GRAND, C. L., HAN, H., MUNOZ, R. M., WEITMAN, S., VON HOFF, D. D., HURLEY, L. H. & BEARSS, D. J. 2002. The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth *in vivo*. *Mol Cancer Ther,* 1**,** 565-73.

GRINSTEIN, E., DU, Y., SANTOURLIDIS, S., CHRIST, J., UHRBERG, M. & WERNET, P. 2007. Nucleolin regulates gene expression in CD34-positive hematopoietic cells. *J Biol Chem,* 282**,** 12439-49.

GRINSTEIN, E., WERNET, P., SNIJDERS, P. J., ROSL, F., WEINERT, I., JIA, W., KRAFT, R., SCHEWE, C., SCHWABE, M., HAUPTMANN, S., DIETEL, M., MEIJER, C. J. & ROYER, H. D. 2002. Nucleolin as activator of human papillomavirus type 18 oncogene transcription in cervical cancer. *J Exp Med,* 196**,** 1067-78.

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*.

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.

HAGER, G. L., MCNALLY, J. G. & MISTELI, T. 2009. Transcription dynamics. *Mol Cell,* 35**,** 741-53.

HAIDET, A. M., RIZO, L., HANDY, C., UMAPATHI, P., EAGLE, A., SHILLING, C., BOUE, D., MARTIN, P. T., SAHENK, Z., MENDELL, J. R. & KASPAR, B. K. 2008. Long-term enhancement of skeletal muscle mass and strength by single gene administration of myostatin inhibitors. *Proc Natl Acad Sci U S A,* 105**,** 4318-22.

HALDER, R., HALDER, K., SHARMA, P., GARG, G., SENGUPTA, S. & CHOWDHURY, S. 2010. Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide. *Mol Biosyst,* 6**,** 2439-47.

HALE, C. R., MAJUMDAR, S., ELMORE, J., PFISTER, N., COMPTON, M., OLSON, S., RESCH, A. M., GLOVER, C. V., 3RD, GRAVELEY, B. R., TERNS, R. M. & TERNS, M. P. 2012. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol Cell,* 45**,** 292-302.

HALFORD, S. E. & MARKO, J. F. 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res,* 32**,** 3040-52.

HANAHAN, D. & WEINBERG, R. A. 2000. The hallmarks of cancer. *Cell,* 100**,** 57-70.

HAY, N., BISHOP, J. M. & LEVENS, D. 1987. Regulatory elements that modulate expression of human c-myc. *Genes Dev,* 1**,** 659-71.

HE, Y., FANG, J., TAATJES, D. J. & NOGALES, E. 2013. Structural visualization of key steps in human transcription initiation. *Nature,* 495**,** 481-6.

HESSELBERTH, J. R., CHEN, X., ZHANG, Z., SABO, P. J., SANDSTROM, R., REYNOLDS, A. P., THURMAN, R. E., NEPH, S., KUEHN, M. S., NOBLE, W. S., FIELDS, S. & STAMATOYANNOPOULOS, J. A. 2009. Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nat Methods,* 6**,** 283-9.

HIRATA, R. K. & RUSSELL, D. W. 2000. Design and packaging of adeno-associated virus gene targeting vectors. *J Virol,* 74**,** 4612-20.

HOLLER, M., WESTIN, G., JIRICNY, J. & SCHAFFNER, W. 1988. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev,* 2**,** 1127-35.

HORVATH, P. & BARRANGOU, R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science,* 327**,** 167-70.

HTUN, H. & DAHLBERG, J. E. 1989. Topology and formation of triple-stranded H-DNA. *Science,* 243**,** 1571-6.

HUPPERT, J. L. & BALASUBRAMANIAN, S. 2005. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res,* 33**,** 2908-16.

HUPPERT, J. L. & BALASUBRAMANIAN, S. 2007. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res,* 35**,** 406-13.

IBORRA, F. J., JACKSON, D. A. & COOK, P. R. 2001. Coupled transcription and translation within nuclei of mammalian cells. *Science,* 293**,** 1139-42.

IZBAN, M. G. & LUSE, D. S. 1992. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J Biol Chem,* 267**,** 13647-55.

JIANG, W., BIKARD, D., COX, D., ZHANG, F. & MARRAFFINI, L. A. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol,* 31**,** 233-9.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337**,** 816-21.

JOHN, S., SABO, P. J., JOHNSON, T. A., SUNG, M. H., BIDDIE, S. C., LIGHTMAN, S. L., VOSS, T. C., DAVIS, S. R., MELTZER, P. S., STAMATOYANNOPOULOS, J. A. & HAGER, G. L. 2008. Interaction of the glucocorticoid receptor with the chromatin landscape. *Mol Cell,* 29**,** 611-24.

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*.

KADONAGA, J. T., CARNER, K. R., MASIARZ, F. R. & TJIAN, R. 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell,* 51**,** 1079-90.

KADONAGA, J. T. & TJIAN, R. 1986. Affinity purification of sequence-specific DNA binding proteins. *Proc Natl Acad Sci U S A,* 83**,** 5889-93.

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.

KIM, J. S., BONIFANT, C., BUNZ, F., LANE, W. S. & WALDMAN, T. 2008. Epitope tagging of endogenous genes in diverse human cell lines. *Nucleic Acids Res,* 36**,** e127.

KIM, S. M., BOWERS, P. M., PAL, D., STRONG, M., TERWILLIGER, T. C., KAUFMANN, M. & EISENBERG, D. 2007. Functional linkages can reveal protein complexes for structure determination. *Structure,* 15**,** 1079-89.

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.

KLENOVA, E. M., NICOLAS, R. H., PATERSON, H. F., CARNE, A. F., HEATH, C. M., GOODWIN, G. H., NEIMAN, P. E. & LOBANENKOV, V. V. 1993. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol,* 13**,** 7612-24.

KOHWI, Y. & KOHWI-SHIGEMATSU, T. 1991. Altered gene expression correlates with DNA structure. *Genes Dev,* 5**,** 2547-54.

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

KORNBERG, R. D. & THOMAS, J. O. 1974. Chromatin structure; oligomers of the histones. *Science,* 184**,** 865-8.

KOUZINE, F., GUPTA, A., BARANELLO, L., WOJTOWICZ, D., BEN-AISSA, K., LIU, J., PRZYTYCKA, T. M. & LEVENS, D. 2013a. Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat Struct Mol Biol,* 20**,** 396-403.

KOUZINE, F. & LEVENS, D. 2007. Supercoil-driven DNA structures regulate genetic transactions. *Front Biosci,* 12**,** 4409-23.

KOUZINE, F., SANFORD, S., ELISHA-FEIL, Z. & LEVENS, D. 2008. The functional response of upstream DNA to dynamic supercoiling *in vivo*. *Nat Struct Mol Biol,* 15**,** 146-54.

KOUZINE, F., WOJTOWICZ, D., YAMANE, A., RESCH, W., KIEFFER-KWON, K. R., BANDLE, R., NELSON, S., NAKAHASHI, H., AWASTHI, P., FEIGENBAUM, L., MENONI, H., HOEIJMAKERS, J., VERMEULEN, W., GE, H., PRZYTYCKA, T. M., LEVENS, D. & CASELLAS, R. 2013b. Global regulation of promoter melting in naive lymphocytes. *Cell,* 153**,** 988-99.

KREEK, M. J., NIELSEN, D. A., BUTELMAN, E. R. & LAFORGE, K. S. 2005. Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction. *Nat Neurosci,* 8**,** 1450-7.

LAVELLE, C. 2009. Forces and torques in the nucleus: chromatin under mechanical constraints. *Biochem Cell Biol,* 87**,** 307-22.

LE HIR, M., GOYENVALLE, A., PECCATE, C., PRECIGOUT, G., DAVIES, K. E., VOIT, T., GARCIA, L. & LORAIN, S. 2013. AAV genome loss from dystrophic mouse muscles during AAV-U7 snRNA-mediated exon-skipping therapy. *Mol Ther,* 21**,** 1551-8.

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

LESCH, K. P., BALLING, U., GROSS, J., STRAUSS, K., WOLOZIN, B. L., MURPHY, D. L. & RIEDERER, P. 1994. Organization of the human serotonin transporter gene. *J Neural Transm Gen Sect,* 95**,** 157-62.

LESCH, K. P. & MERSCHDORF, U. 2000. Impulsivity, aggression, and serotonin: a molecular psychobiological perspective. *Behav Sci Law,* 18**,** 581-604.

LESCH, K. P., ZENG, Y., REIF, A. & GUTKNECHT, L. 2003. Anxiety-related traits in mice with modified genes of the serotonergic pathway. *Eur J Pharmacol,* 480**,** 185-204.

LEVENS, D., DUNCAN, R. C., TOMONAGA, T., MICHELOTTI, G. A., COLLINS, I., DAVIS-SMYTH, T., ZHENG, T. & MICHELOTTI, E. F. 1997. DNA conformation, topology, and the regulation of c-myc expression. *Curr Top Microbiol Immunol,* 224**,** 33-46.

LI, Z., VAN CALCAR, S., QU, C., CAVENEE, W. K., ZHANG, M. Q. & REN, B. 2003. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc Natl Acad Sci U S A,* 100**,** 8164-9.

LIU, J. & LEVENS, D. 2006. Making myc. *Curr Top Microbiol Immunol,* 302**,** 1-32.

LIU, L. F. & WANG, J. C. 1987. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A,* 84**,** 7024-7.

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.

LOTRICH, F. E. & POLLOCK, B. G. 2004. Meta-analysis of serotonin transporter polymorphisms and affective disorders. *Psychiatr Genet,* 14**,** 121-9.

LOVEJOY, E. A., SCOTT, A. C., FISKERSTRAND, C. E., BUBB, V. J. & QUINN, J. P. 2003. The serotonin transporter intronic VNTR enhancer correlated with a predisposition to affective disorders has distinct regulatory elements within the domain based on the primary DNA sequence of the repeat unit. *Eur J Neurosci,* 17**,** 417-20.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature,* 389**,** 251-60.

MACKENZIE, A., PAYNE, C., BOYLE, S., CLARKE, A. R. & QUINN, J. P. 2000. The human preprotachykinin-A gene promoter has been highly conserved and can drive human-like marker gene expression in the adult mouse CNS. *Mol Cell Neurosci,* 16**,** 620-30.

MACKENZIE, A. & QUINN, J. 1999. A serotonin transporter gene intron 2 polymorphic region, correlated with affective disorders, has allele-dependent differential enhancer-like properties in the mouse embryo. *Proc Natl Acad Sci U S A,* 96**,** 15251-5.

MALI, P., AACH, J., STRANGES, P. B., ESVELT, K. M., MOOSBURNER, M., KOSURI, S., YANG, L. & CHURCH, G. M. 2013a. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol,* 31**,** 833-8.

MALI, P., ESVELT, K. M. & CHURCH, G. M. 2013b. Cas9 as a versatile tool for engineering biology. *Nat Methods,* 10**,** 957-63.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013c. RNA-guided human genome engineering via Cas9. *Science,* 339**,** 823-6.

MARCU, K. B. 1987. Regulation of expression of the c-myc proto-oncogene. *Bioessays,* 6**,** 28-32.

MARCU, K. B., BOSSONE, S. A. & PATEL, A. J. 1992. myc function and regulation. *Annu Rev Biochem,* 61**,** 809-60.

MARGARITIS, T. & HOLSTEGE, F. C. 2008. Poised RNA polymerase II gives pause for thought. *Cell,* 133**,** 581-4.

MASTRANGELO, I. A., COUREY, A. J., WALL, J. S., JACKSON, S. P. & HOUGH, P. V. 1991. DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement. *Proc Natl Acad Sci U S A,* 88**,** 5670-4.

MATSUI, T., SEGALL, J., WEIL, P. A. & ROEDER, R. G. 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J Biol Chem,* 255**,** 11992-6.

MATUNIS, M. J., MATUNIS, E. L. & DREYFUSS, G. 1992. Isolation of hnRNP complexes from Drosophila melanogaster. *J Cell Biol,* 116**,** 245-55.

MCDONOUGH, P. M., HANFORD, D. S., SPRENKLE, A. B., MELLON, N. R. & GLEMBOTSKI, C. C. 1997. Collaborative roles for c-Jun N-terminal kinase, c-Jun, serum response factor, and Sp1 in calcium-regulated myocardial gene expression. *J Biol Chem,* 272**,** 24046-53.

MCGHEE, J. D. & VON HIPPEL, P. H. 1975. Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases. *Biochemistry,* 14**,** 1281-96.

MCLAUGHLIN, D. P., LITTLE, K. Y., LOPEZ, J. F. & WATSON, S. J. 1996. Expression of serotonin transporter mRNA in human brainstem raphe nuclei. *Neuropsychopharmacology,* 15**,** 523-9.

MCLUCKIE, K. I., DI ANTONIO, M., ZECCHINI, H., XIAN, J., CALDAS, C., KRIPPENDORFF, B. F., TANNAHILL, D., LOWE, C. & BALASUBRAMANIAN, S. 2013. G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. *J Am Chem Soc,* 135**,** 9640-3.

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***,** 1-16.

MELLOR, J. 2005. The dynamics of chromatin remodeling at promoters. *Mol Cell,* 19**,** 147-57.

MICHELOTTI, E. F., MICHELOTTI, G. A., ARONSOHN, A. I. & LEVENS, D. 1996a. 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.

MICHELOTTI, G. A., MICHELOTTI, E. F., PULLNER, A., DUNCAN, R. C., EICK, D. & LEVENS, D. 1996b. Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene *in vivo*. *Mol Cell Biol,* 16**,** 2656-69.

MILLER, D. G. 2011. AAV-mediated gene targeting. *Methods Mol Biol,* 807**,** 301-15.

MILLER, D. G., PETEK, L. M. & RUSSELL, D. W. 2003. Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. *Mol Cell Biol,* 23**,** 3550-7.

MISTELI, T. 2001. Protein dynamics: implications for nuclear architecture and gene expression. *Science,* 291**,** 843-7.

MONGELARD, F. & BOUVET, P. 2007. Nucleolin: a multiFACeTed protein. *Trends Cell Biol,* 17**,** 80-6.

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*.

MULLER, S., SANDERS, D. A., DI ANTONIO, M., MATSIS, S., RIOU, J. F., RODRIGUEZ, R. & BALASUBRAMANIAN, S. 2012. Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. *Org Biomol Chem,* 10**,** 6537-46.

MURAT, P., GORMALLY, M. V., SANDERS, D., DI ANTONIO, M. & BALASUBRAMANIAN, S. 2013. Light-mediated in cell downregulation of G-quadruplex-containing genes using a photo-caged ligand. *Chem Commun (Camb),* 49**,** 8453-5.

MUSE, G. W., GILCHRIST, D. A., NECHAEV, S., SHAH, R., PARKER, J. S., GRISSOM, S. F., ZEITLINGER, J. & ADELMAN, K. 2007. RNA polymerase is poised for activation across the genome. *Nat Genet,* 39**,** 1507-11.

NAGAICH, A. K. & HAGER, G. L. 2004. UV laser cross-linking: a real-time assay to study dynamic protein/DNA interactions during chromatin remodeling. *Sci STKE,* 2004**,** pl13.

NIE, Z., HU, G., WEI, G., CUI, K., YAMANE, A., RESCH, W., WANG, R., GREEN, D. R., TESSAROLLO, L., CASELLAS, R., ZHAO, K. & LEVENS, D. 2012. c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell,* 151**,** 68-79.

OHLSSON, R., RENKAWITZ, R. & LOBANENKOV, V. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet,* 17**,** 520-7.

OLOFSSON, B. A., KELLY, C. M., KIM, J., HORNSBY, S. M. & AZIZKHAN-CLIFFORD, J. 2007. Phosphorylation of Sp1 in response to DNA damage by ataxia telangiectasia-mutated kinase. *Mol Cancer Res,* 5**,** 1319-30.

ORPHANIDES, G., LAGRANGE, T. & REINBERG, D. 1996. The general transcription factors of RNA polymerase II. *Genes Dev,* 10**,** 2657-83.

ORPHANIDES, G. & REINBERG, D. 2000. RNA polymerase II elongation through chromatin. *Nature,* 407**,** 471-5.

OSBORNE, C. S., CHAKALOVA, L., BROWN, K. E., CARTER, D., HORTON, A., DEBRAND, E., GOYENECHEA, B., MITCHELL, J. A., LOPES, S., REIK, W. & FRASER, P. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet,* 36**,** 1065-71.

OSTARECK-LEDERER, A., OSTARECK, D. H. & HENTZE, M. W. 1998. Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2. *Trends Biochem Sci,* 23**,** 409-11.

OSTER, S. K., HO, C. S., SOUCIE, E. L. & PENN, L. Z. 2002. The myc oncogene: MarvelouslY Complex. *Adv Cancer Res,* 84**,** 81-154.

OWENS, M. J. & NEMEROFF, C. B. 1994. Role of serotonin in the pathophysiology of depression: focus on the serotonin transporter. *Clin Chem,* 40**,** 288-95.

OWENS, M. J. & NEMEROFF, C. B. 1998. The serotonin transporter and depression. *Depress Anxiety,* 8 Suppl 1**,** 5-12.

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.

PARISI, F., WIRAPATI, P. & NAEF, F. 2007. Identifying synergistic regulation involving c-Myc and sp1 in human tissues. *Nucleic Acids Res,* 35**,** 1098-107.

PEARSON, C. E. & SINDEN, R. R. 1996. Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry,* 35**,** 5041-53.

PETESCH, S. J. & LIS, J. T. 2008. Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. *Cell,* 134**,** 74-84.

PHAIR, R. D., GORSKI, S. A. & MISTELI, T. 2004. Measurement of dynamic protein binding to chromatin *in vivo*, using photobleaching microscopy. *Methods Enzymol,* 375**,** 393-414.

PHILLIPS, J. E. & CORCES, V. G. 2009. CTCF: master weaver of the genome. *Cell,* 137**,** 1194-211.

POOREY, K., VISWANATHAN, R., CARVER, M. N., KARPOVA, T. S., CIRIMOTICH, S. M., MCNALLY, J. G., BEKIRANOV, S. & AUBLE, D. T. 2013. Measuring chromatin interaction dynamics on the second time scale at single-copy genes. *Science,* 342**,** 369-72.

PORTEUS, M. H., CATHOMEN, T., WEITZMAN, M. D. & BALTIMORE, D. 2003. Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Mol Cell Biol,* 23**,** 3558-65.

PTASHNE, M. & GANN, A. 1997. Transcriptional activation by recruitment. *Nature,* 386**,** 569-77.

QUINN, J. P., HOLBROOK, N. & LEVENS, D. 1987. Binding of a cellular protein to the gibbon ape leukemia virus enhancer. *Mol Cell Biol,* 7**,** 2735-44.

QUINN, J. P., MCGREGOR, R. A., FISKERSTRAND, C. E., DAVEY, C., ALLAN, J. & DALZIEL, R. G. 1998. Identification of a novel multifunctional structural domain in the herpes simplex virus type 1 genome: implications for virus latency. *J Gen Virol,* 79 (Pt 10)**,** 2529-32.

QUINN, J. P., TAKIMOTO, M., IADAROLA, M., HOLBROOK, N. & LEVENS, D. 1989. Distinct factors bind the AP-1 consensus sites in gibbon ape leukemia virus and simian virus 40 enhancers. *J Virol,* 63**,** 1737-42.

RAHL, P. B., LIN, C. Y., SEILA, A. C., FLYNN, R. A., MCCUINE, S., BURGE, C. B., SHARP, P. A. & YOUNG, R. A. 2010. c-Myc regulates transcriptional pause release. *Cell,* 141**,** 432-45.

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.

RASKO, J. E., KLENOVA, E. M., LEON, J., FILIPPOVA, G. N., LOUKINOV, D. I., VATOLIN, S., ROBINSON, A. F., HU, Y. J., ULMER, J., WARD, M. D., PUGACHEVA, E. M., NEIMAN, P. E., MORSE, H. C., 3RD, COLLINS, S. J. & LOBANENKOV, V. V. 2001. Cell growth inhibition by the multifunctional multivalent zinc-finger factor CTCF. *Cancer Res,* 61**,** 6002-7.

RECOUVREUX, P., LAVELLE, C., BARBI, M., CONDE, E. S. N., LE CAM, E., VICTOR, J. M. & VIOVY, J. L. 2011. Linker histones incorporation maintains chromatin fiber plasticity. *Biophys J,* 100**,** 2726-35.

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. 2013. MAOA and mechanisms of panic disorder revisited: from bench to molecular psychotherapy. *Mol Psychiatry*.

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-µVNTR in panic disorder in four European populations. *Am J Med Genet B Neuropsychiatr Genet,* 159B**,** 786-93.

RESSLER, K. J. & NEMEROFF, C. B. 2000. Role of serotonergic and noradrenergic systems in the pathophysiology of depression and anxiety disorders. *Depress Anxiety,* 12 Suppl 1**,** 2-19.

RHEE, H. S. & PUGH, B. F. 2011. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell,* 147**,** 1408-19.

RICH, A. & ZHANG, S. 2003. Timeline: Z-DNA: the long road to biological function. *Nat Rev Genet,* 4**,** 566-72.

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.

ROCA, J. 2011a. The torsional state of DNA within the chromosome. *Chromosoma,* 120**,** 323-34.

ROCA, J. 2011b. Transcriptional inhibition by DNA torsional stress. *Transcription,* 2**,** 82-85.

ROEDER, R. G. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci,* 21**,** 327-35.

ROEDER, R. G. 2005. Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett,* 579**,** 909-15.

ROMAN, D. L., WALLINE, C. C., RODRIGUEZ, G. J. & BARKER, E. L. 2003. Interactions of antidepressants with the serotonin transporter: a contemporary molecular analysis. *Eur J Pharmacol,* 479**,** 53-63.

RUDNICK, G. & WALL, S. C. 1993. Non-neurotoxic amphetamine derivatives release serotonin through serotonin transporters. *Mol Pharmacol,* 43**,** 271-6.

RUSSELL, D. W. & HIRATA, R. K. 1998. Human gene targeting by viral vectors. *Nat Genet,* 18**,** 325-30.

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

SANDER, J. D. & JOUNG, J. K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol,* 32**,** 347-55.

SARGENT, R. G., BRENNEMAN, M. A. & WILSON, J. H. 1997. Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol Cell Biol,* 17**,** 267-77.

SCHMIEDEBERG, L., SKENE, P., DEATON, A. & BIRD, A. 2009. A temporal threshold for formaldehyde crosslinking and fixation. *PLoS One,* 4**,** e4636.

SEILA, A. C., CORE, L. J., LIS, J. T. & SHARP, P. A. 2009. Divergent transcription: a new feature of active promoters. *Cell Cycle,* 8**,** 2557-64.

SHEN, B., ZHANG, W., ZHANG, J., ZHOU, J., WANG, J., CHEN, L., WANG, L., HODGKINS, A., IYER, V., HUANG, X. & SKARNES, W. C. 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods,* 11**,** 399-402.

SHKLOVER, J., WEISMAN-SHOMER, P., YAFE, A. & FRY, M. 2010. Quadruplex structures of muscle gene promoter sequences enhance *in vivo* MyoD-dependent gene expression. *Nucleic Acids Res,* 38**,** 2369-77.

SHLYAKHTENKO, L. S., POTAMAN, V. N., SINDEN, R. R. & LYUBCHENKO, Y. L. 1998. Structure and dynamics of supercoil-stabilized DNA cruciforms. *J Mol Biol,* 280**,** 61-72.

SHUMAY, E. & FOWLER, J. S. 2010. Identification and characterization of putative methylation targets in the MAOA locus using bioinformatic approaches. *Epigenetics,* 5**,** 325-42.

SHUMAY, E., LOGAN, J., VOLKOW, N. D. & FOWLER, J. S. 2012. Evidence that the methylation state of the monoamine oxidase A (MAOA) gene predicts brain activity of MAO A enzyme in healthy men. *Epigenetics,* 7**,** 1151-60.

SIDDIQUI-JAIN, A., GRAND, C. L., BEARSS, D. J. & HURLEY, L. H. 2002. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A,* 99**,** 11593-8.

SIMONSSON, T. 2001. G-quadruplex DNA structures--variations on a theme. *Biol Chem,* 382**,** 621-8.

SIOMIN, Y. A., SIMONOV, V. V. & POVERENNY, A. M. 1973. The reaction of formaldehyde with deoxynucleotides and DNA in the presence of amino acids and lysine-rich histone. *Biochim Biophys Acta,* 331**,** 27-32.

SMALE, S. T. & KADONAGA, J. T. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem,* 72**,** 449-79.

SMIH, F., ROUET, P., ROMANIENKO, P. J. & JASIN, M. 1995. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res,* 23**,** 5012-9.

SOEBY, K., LARSEN, S. A., OLSEN, L., RASMUSSEN, H. B. & WERGE, T. 2005. Serotonin transporter: evolution and impact of polymorphic transcriptional regulation. *Am J Med Genet B Neuropsychiatr Genet,* 136B**,** 53-7.

SON, L. S., BACOLLA, A. & WELLS, R. D. 2006. Sticky DNA: *in vivo* formation in E. coli and *in vitro* association of long GAA*TTC tracts to generate two independent supercoiled domains. *J Mol Biol,* 360**,** 267-84.

SONG, J., UGAI, H., OGAWA, K., WANG, Y., SARAI, A., OBATA, Y., KANAZAWA, I., SUN, K., ITAKURA, K. & YOKOYAMA, K. K. 2001. Two consecutive zinc fingers in Sp1 and in MAZ are essential for interactions with cis-elements. *J Biol Chem,* 276**,** 30429-34.

SPENCER, C. A. & GROUDINE, M. 1991. Control of c-myc regulation in normal and neoplastic cells. *Adv Cancer Res,* 56**,** 1-48.

SPRAGUE, B. L., PEGO, R. L., STAVREVA, D. A. & MCNALLY, J. G. 2004. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophys J,* 86**,** 3473-95.

STORCK, S., SHUKLA, M., DIMITROV, S. & BOUVET, P. 2007. Functions of the histone chaperone nucleolin in diseases. *Subcell Biochem,* 41**,** 125-44.

SU, W., JACKSON, S., TJIAN, R. & ECHOLS, H. 1991. DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. *Genes Dev,* 5**,** 820-6.

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.

SUZUKI, T., KIMURA, A., NAGAI, R. & HORIKOSHI, M. 2000. Regulation of interaction of the acetyltransferase region of p300 and the DNA-binding domain of Sp1 on and through DNA binding. *Genes Cells,* 5**,** 29-41.

TERNS, M. P. & TERNS, R. M. 2011. CRISPR-based adaptive immune systems. *Curr Opin Microbiol,* 14**,** 321-7.

TEYTELMAN, L., THURTLE, D. M., RINE, J. & VAN OUDENAARDEN, A. 2013. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A,* 110**,** 18602-7.

THOMAS, K. R. & CAPECCHI, M. R. 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell,* 51**,** 503-12.

TOMONAGA, T. & LEVENS, D. 1995. Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *J Biol Chem,* 270**,** 4875-81.

TOMONAGA, T. & LEVENS, D. 1996. Activating transcription from single stranded DNA. *Proc Natl Acad Sci U S A,* 93**,** 5830-5.

TORRANO, V., CHERNUKHIN, I., DOCQUIER, F., D'ARCY, V., LEON, J., KLENOVA, E. & DELGADO, M. D. 2005. CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. *J Biol Chem,* 280**,** 28152-61.

TRAVERS, A. & MUSKHELISHVILI, G. 2007. A common topology for bacterial and eukaryotic transcription initiation? *EMBO Rep,* 8**,** 147-51.

UHER, R. & MCGUFFIN, P. 2008. The moderation by the serotonin transporter gene of environmental adversity in the aetiology of mental illness: review and methodological analysis. *Mol Psychiatry,* 13**,** 131-46.

VASILEVA, A. & JESSBERGER, R. 2005. Precise hit: adeno-associated virus in gene targeting. *Nat Rev Microbiol,* 3**,** 837-47.

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, A., HALDER, K., HALDER, R., YADAV, V. K., RAWAL, P., THAKUR, R. K., MOHD, F., SHARMA, A. & CHOWDHURY, S. 2008. Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J Med Chem,* 51**,** 5641-9.

VOLOGODSKII, A. V., LEVENE, S. D., KLENIN, K. V., FRANK-KAMENETSKII, M. & COZZARELLI, N. R. 1992. Conformational and thermodynamic properties of supercoiled DNA. *J Mol Biol,* 227**,** 1224-43.

WANG, Q., XU, X., LI, J., LIU, J., GU, H., ZHANG, R., CHEN, J., KUANG, Y., FEI, J., JIANG, C., WANG, P., PEI, D., DING, S. & XIE, X. 2011. Lithium, an anti-psychotic drug, greatly enhances the generation of induced pluripotent stem cells. *Cell Res,* 21**,** 1424-35.

WARBURTON, A., BREEN, G., RUJESCU, D., BUBB, V. J. & QUINN, J. P. 2014. Characterization of a REST-Regulated Internal Promoter in the Schizophrenia Genome-Wide Associated Gene MIR137. *Schizophr Bull*.

WASYLISHEN, A. R. & PENN, L. Z. 2010. Myc: the beauty and the beast. *Genes Cancer,* 1**,** 532-41.

WEINSTEIN, I. B. 2002. Cancer. Addiction to oncogenes--the Achilles heal of cancer. *Science,* 297**,** 63-4.

WEINSTEIN, I. B., BEGEMANN, M., ZHOU, P., HAN, E. K., SGAMBATO, A., DOKI, Y., ARBER, N., CIAPARRONE, M. & YAMAMOTO, H. 1997. Disorders in cell circuitry associated with multistage carcinogenesis: exploitable targets for cancer prevention and therapy. *Clin Cancer Res,* 3**,** 2696-702.

WIEDENHEFT, B., STERNBERG, S. H. & DOUDNA, J. A. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature,* 482**,** 331-8.

WIERSTRA, I. & ALVES, J. 2006. FOXM1c transactivates the human c-myc promoter directly via the two TATA boxes P1 and P2. *FEBS J,* 273**,** 4645-67.

WONG, B., CHEN, S., KWON, J. A. & RICH, A. 2007. Characterization of Z-DNA as a nucleosome-boundary element in yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A,* 104**,** 2229-34.

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.

WOODCOCK, C. L., SKOULTCHI, A. I. & FAN, Y. 2006. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res,* 14**,** 17-25.

WU, C., WONG, Y. C. & ELGIN, S. C. 1979. The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell,* 16**,** 807-14.

YAFE, A., SHKLOVER, J., WEISMAN-SHOMER, P., BENGAL, E. & FRY, M. 2008. Differential binding of quadruplex structures of muscle-specific genes regulatory sequences by MyoD, MRF4 and myogenin. *Nucleic Acids Res,* 36**,** 3916-25.

YAKOVCHUK, P., PROTOZANOVA, E. & FRANK-KAMENETSKII, M. D. 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res,* 34**,** 564-74.

ZEITLINGER, J., STARK, A., KELLIS, M., HONG, J. W., NECHAEV, S., ADELMAN, K., LEVINE, M. & YOUNG, R. A. 2007. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet,* 39**,** 1512-6.

ZELLER, K. I., ZHAO, X., LEE, C. W., CHIU, K. P., YAO, F., YUSTEIN, J. T., OOI, H. S., ORLOV, Y. L., SHAHAB, A., YONG, H. C., FU, Y., WENG, Z., KUZNETSOV, V. A., SUNG, W. K., RUAN, Y., DANG, C. V. & WEI, C. L. 2006. Global mapping of c-Myc binding sites and target gene networks in human B cells. *Proc Natl Acad Sci U S A,* 103**,** 17834-9.

ZHAO, S., VENKATASUBBARAO, K., LI, S. & FREEMAN, J. W. 2003. Requirement of a specific Sp1 site for histone deacetylase-mediated repression of transforming growth factor beta Type II receptor expression in human pancreatic cancer cells. *Cancer Res,* 63**,** 2624-30.

ZHOU, F. C., SARI, Y. & ZHANG, J. K. 2000. Expression of serotonin transporter protein in developing rat brain. *Brain Res Dev Brain Res,* 119**,** 33-45.

ZLATANOVA, J. & VICTOR, J. M. 2009. How are nucleosomes disrupted during transcription elongation? *HFSP J,* 3**,** 373-8.