THE UNIVERSITY *of* EDINBURGH

Genotype-phenotype correlations in β-catenin

Anagha Krishna

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

February 2019

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

Name Anagha Krishna. Date 28/02/19

Acknowledgements

It is with immense pleasure I would like to thank everyone who made my PhD journey the most memorable and cherishable experience.

I would like to thank Peter for giving me the opportunity to be a part of his lab and for all the guidance and support.

My greatest thanks go to Derya, for her constant guidance, her genius ideas and for being with me and helping me in every which ways possible

I am deeply grateful to Shahida for her encouragement, and for all the long conversations which provided the much needed break from my often crazy work days. I would like to thank Raffa and Chris, for being the most joyous people to work with. A Big thank you to Paul, who tirelessly carried tons of my samples to the sequencing facility day in and day out and also for reading through my thesis and correcting my english.

I would like to thank the scientific support staff of Roslin Institute, especially Graeme and Bob for all the help with FACS sort.

I would like to thank Dr. Helen Brown for providing help with the statistical analysis.

I would also like to thank Andrew Wood and Martijn Kelder for their help in analysis of the saturation data. Special thanks to Ailith Ewing for her help with analyzing the mutational background. Also, a big thanks to Deepti for her help with the graphs.

Most importantly, I would like to thank my family, for their constant support and encouragement. Thank you amma, ayya, Abhignya and Alwar, for always encouraging me to pursue my dreams.

Finally, I would like to thank the College of Medicine and Veterinary Medicine for funding my studies.

Abstract

Canonical Wnt signaling is one of the most extensively studied signaling cascades, with a multifunctional role in development and disease. Activating mutations in β-catenin, the key regulator of this pathway, have been observed in many types of cancer. The general hypothesis is that, all these activating mutations in β-catenin affect the protein turnover and localization, and thus aberrantly activate the Wnt target genes that are capable of modulating multiple aspects of tumourigenesis. Although various residues in β-catenin were found to be mutated in different cancer types, a detailed analysis of the individual mutations has been lacking so far.

In this project, firstly I sought to explore the frequency and distribution of mutations in the β-catenin gene across different cancer types using the COSMIC (Catalogue of Somatic Mutations in Cancer) database. The analysis of the β-catenin mutational spectrum revealed a preferential selection of different residues and amino acid substitutions among the different cancer types. This specific selectivity of mutations indicated a difference in phenotypic effect from different β-catenin mutations. Furthermore, through "saturation mutagenesis" of the β-catenin hotspot region (L31-G50) and by generating independent clonal cell lines harbouring various substitution mutations at the selected top six β-catenin residues through "multiplex targeting", I tried to understand whether or not there exists a genotype-phenotype correlation among the β-catenin mutations. Using these complementary experimental approaches, I convincingly demonstrated the presence of an allele specific β-catenin activity level conferred by these mutational variants, which cannot be explained by the current model of β-catenin regulation.

The results of this study imply a fundamental difference between these mutations, the existence of a genotype-phenotype correlation based on β-catenin activity, and challenges the prevailing dogma of β-catenin regulation, thus emphasizing the need to further examine the underlying mechanistic process involved in the observed differential phenotypic response.

Lay Summary

Genes are functional units of DNA, many of which code for proteins responsible for controlling the various functions of a cell. A number of these processes are controlled by a coordinated network of protein-protein interaction referred to as signaling pathways. Often, genes encoding the proteins of these signaling pathway are susceptible to mutations (changes in the DNA sequence) which when remained unrepaired may result in various diseases including cancer. Analysis of the DNA sequence of tumour samples from different types of cancers have revealed mutations in various genes including the gene encoding the β-catenin protein which is an important component of the Wnt signaling pathway.

The current accepted model of β-catenin activity is based on an all or none response governed solely by the stability of β-catenin protein. The β-catenin mutations observed in many types of cancer are all considered to affect the stability of the protein and have the same effect. However my analysis of β-catenin mutations revealed that different tumour types have preferential selection of mutations. This preferential selection indicates the presence of a specific mutation-effect correlation.

In this project, I studied this mutation-effect correlation among the β-catenin mutations. Our study was made possible by the advanced gene editing technique known as the CRISPR/Cas9 system. Using the CRISPR/Cas9 system I was able to change the normal β-catenin gene to create the different mutations artificially in the genome of mouse embryonic stem cells. By creating the different mutations in the β-catenin gene using the CRISPR/Cas9 system and analyzing the activity corresponding to each variant, I was able to show the existence of differences in β-catenin activity response among the various mutants. In conclusion, this study lays a strong ground work and highlights the importance of the variation in functional outcome resulting from the specific mutations in the different cancer types in not only understanding their role in process of cancer but also for further therapeutic applications.

Contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of tables

Chapter 4

Chapter 7

List of Abbreviations

AAV	adeno- associated virus
ANOVA	Analysis of variance
BER	base excision repair
bGh	bovine growth hormone
bp	base pair
BSA	bovine serum albumin
Cas9	CRISPR associated protein 9
cDNA	complementary DNA
CGP	Cancer Genome Project
CNS	central nervous system
COSMIC	Catalogue of somatic mutations in cancer
CRCs	colorectal cancers
CRISPR	Clustered Regularly Interspersed Palindromic Repeat
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
dCas9	dead Cas9
DDR	DNA damage response
DMSO	dimethyl sulfoxide
Ds	double stranded
DSBs	double-strand breaks
ER	endoplasmic reticulum
F primer	forward primer
FACS	fluorescence activated cell sorting
FAP	familial adenomatous polyposis

FBS	fetal bovine serum
FIAU	Fialuridine or 1-(2-deoxy-2-fluoro-1-D-arabinofuranosyl)-5-iodouracil
FITC	Fluorescein isothiocynate
FLP	flippase
FSC	forward scatter
GDC	Genomic Data Commons
HAT	histone acetyltransferase
HCC	hepatocellular carcinomas
HDR	homology directed repair
Het	heterozygous
Hom	homozygous
HR	homologous recombination
HSV1	herpes simplex virus type-1
ICGC	International Cancer Genome Consortium
Indel	insertions deletions
LOH	loss of heterozygosity
MCR	mutation cluster region
mESCs	mouse embryonic stem cells
MMTV	mouse mammary tumour virus
mRNA	messenger RNA
NCI	National Cancer Institute
NHEJ	non-homologous end joining
NIH	National Institutes of Health
NLS	nuclear localization signal
nt	nucleotide

PAM	protospacer adjacent motif
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PNS	positive negative strategy
QE	QuickExtract
R primer	reverse primer
RVD	repeat variable di-residue
SDS	sodium dodecyl sulfate
SDSA	synthesis dependent strand annealing
sgRNA	single guide RNA
SNV	single nucleotide variant
SSC	side scatter
ssDNA	single stranded DNA
ssODN	short single-stranded oligodinucleotide
TALENs	transcription activator like effector nucleases
TCGA	The Cancer Genome Atlas
TFIID	transcription factor II D
Tk	thymidine kinase
tracr RNA	trans activating RNA
TVs	targeting vectors
UV	ultraviolet
WT	wild type
ZFNs	zinc- finger nucleases
ZFPs	zinc finger proteins

Chapter 1 Introduction

1.1 Background

Deregulation of cellular signaling is central to the multistep process of tumourigenesis, largely resulting from malfunctioning of the key mediators of the circuitry. Studies over the years, have revealed the underlying genetic and epigenetic changes in various components of the signaling cascades, which lead to the deviation from normal physiology, and have also highlighted their crucial role in the acquisition of characteristic cellular advantages to cancer cells. The most significant among them being, the gain of function and loss of function mutations in 'proto oncogenes' and 'tumour suppressor genes', respectively, that code for the major mediator proteins of cellular processes. The genetic alterations in these genes are especially crucial for hijacking the multiple mechanisms key for neoplastic transformation, including the process of cell proliferation and cell death. Characterization of the tumorigenic potential of the aberrant counterparts of cellular oncogene and tumour suppressor genes have not only provided key insights into the disease pathogenesis, but in addition helped to further understand the role of these mediators in the regulation of the molecular mechanism of cellular signaling pathways. The advancements in the genome engineering techniques such as the Clustered Regularly Interspersed Palindromic Repeat/CRISPR associated protein 9 (CRISPR/Cas9) system have provided the perfect platform for such investigations. In this view, this project is an attempt to understand the genotype-phenotype correlations in the oncogenic protein β-catenin, and further envisage the yet unresolved mechanistic details governing the β-catenin mediated signal transduction cascade.

1.2 The Wnt signaling cascade

The diverse role of Wnt signaling in embryonic development and maintenance of adult homeostasis is coordinated by a network of multiple receptors, ligands, phosphokinases, phosphatases, secondary metabolites, G proteins and various other proteins orchestrating the Wnt mediated signaling cascades. The role of the majority of these mediators in Wnt signaling, and their importance in development, have been understood through the functional studies of their respective homologues in model systems, including *Drosophilla*, *Xenopus laevis*, *C elegans* and *Mus musculus*.

In the year 1982, using the method of proviral tagging, Nusse and collagues identified the proviral integration site of mouse mammary tumour virus (MMTV) to be within the *Int1* (int for integration) locus in multiple tumour samples, making *Int1* a putative proto-oncogene activated by viral integration (Nusse and Varmus, 1982). The oncogenic activity of *Int1* was later confirmed by the tumour initiating effect of *Int1* transgene under the transcriptional regulation of MMTV (Varmus, 1988). The conservative nature of *Int1* gene allowed identification of *Int1* homologue in *Drosophilla,* that was later mapped to the *Wingless* (*Wg*) gene responsible for segment polarity (Rijsewijk *et al.*, 1987). To avoid confusions in nomenclature, the *Int1* and *Wg* members were named *Wnt*-for Wingless related integration site (Nusse *et al.*, 1991). *Wnt* is a large gene family consisting of variable number of *Wnt* related genes in different species. In mammals, 19 *Wnt* related genes have been described, expressing the secretory Wnt proteins with distinct developmental roles (Garriock *et al.*, 2007).

The biogenesis of Wnt (Fig 1-1) involves a complex process, including the crucial palmitoylation by the membrane bound O-acyltransferase protein porcupine in the endoplasmic reticulum (ER) (Kadowaki *et al.*, 1996). This post translational modification of Wnt proteins is necessary for both secretion, and its interaction with Frizzled receptor (Bazan and de Sauvage, 2009).

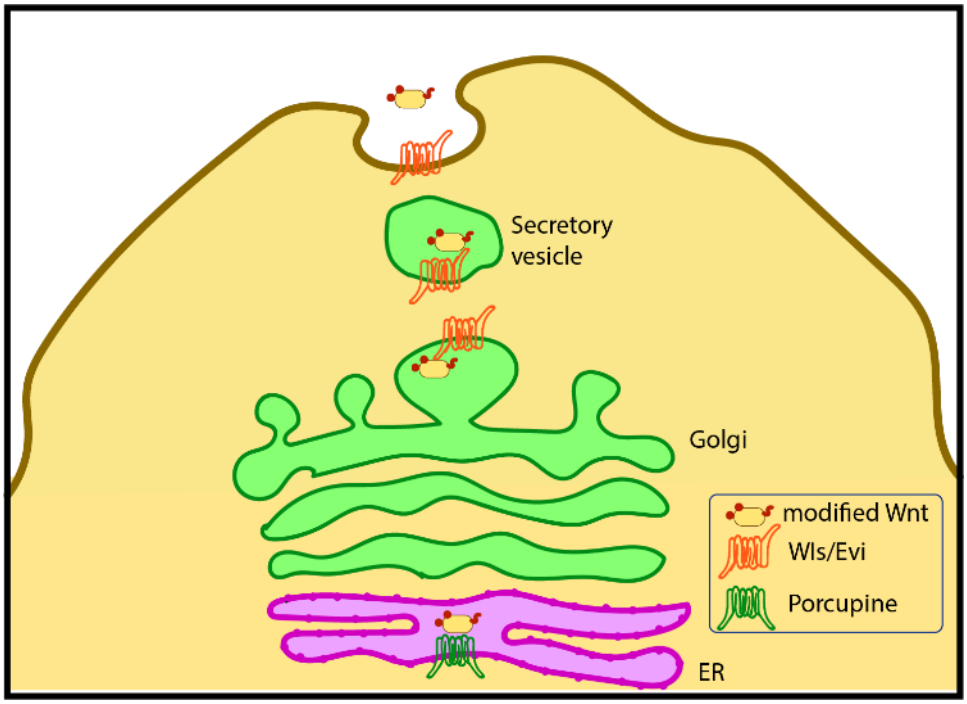


Figure 1-1: Biogenesis of Wnt. The translated Wnt proteins are modified in the ER by the membrane bound porcupine and then transported to the golgi apparatus, where the WIs//Evi act as chaperones and assist the transport of the modified Wnt proteins across the golgi to the extracellular space.

Once transported to the golgi apparatus, the lipid modified Wnts are recognized by Wntless/Evenness Interupted (WIs/Evi). Acting as chaperones, WIs assist trafficking of Wnt proteins across the trans golgi network to the plasma membrane (Port and Basler, 2010). Binding of various Wnt ligands to its cognate receptor Frizzled, induces the activation of both canonical and non-canonical signaling cascades. To date, a diverse range of Wnt mediated, β-catenin independent non canonical pathways have been described of which the planar cell polarity pathway and the Wnt/Ca^{2+} have been studied extensively.

The structured organization of the cellular architecture, requires directional alignment or polarization for localization of specific components, or performing specified functions, and the Wnt mediated planar cell polarity pathway plays a key role in coordinating these polarization events. The Wnt ligand, acting through Frizzled receptor, activates various downstream components, including Dishevelled (Dsh), which in turn activates the Rho

family of small GTPases including Rho and Rac, each of which then activates the ROCK and the JNK/P-38 type MAPK, respectively. The coordinated effort of these mediators is required for directional alignment of various structures, including hair follicles, sensory bristles, actin skeleton organization, and also in migration of dorsal mesodermal cells during gastrulation. The Wnt/Ca^{2+} pathway, acting through the phosphotidyl inositol signaling coupled to G-protein, triggers the release of Ca^{2+} from the ER, and these second messengers activate Ca^{2+}/ CaMKII and PKC, which in turn are responsible for activating the transcription factors enhancing expression of various genes. The Wnt/Ca^{2+} pathway is known to play an important role at various stages of gastrulation and organogenesis (Komiya and Habas, 2008). These non-canonical pathways do not act independently and there exists consistent cross talk between the β-catenin mediated canonical and non-canonical pathways, specifically the reciprocal interaction between the two cascades, involving various different mechanisms have been a subject of investigation in multiple contexts during both development and disease (Veeman, Axelrod and Moon, 2003; Toyama *et al.*, 2010).

1.2.1 Wnt/β-catenin mediated signalling

The canonical β-catenin mediated Wnt signal transduction pathway is one of the most extensively studied signaling cascades, with a multifunctional role in development and disease (Clevers, 2006). Absent in single cell organisms, the Wnt/β-catenin pathway evolved as early as in the sponges and is conserved across metazoa from the lower invertebrates to higher mammalian species (Croce, 2008). A large number of mediators of this pathway have been attributed with either tumour suppressor or oncogenic properties, owing to which, the pathway is subjected to high regulation at various levels, thus highlighting the importance of Wnt signaling in maintenance of normal tissue homeostasis (Polakis, 2000, 2012).

Central to the canonical Wnt signaling cascade is the β-catenin protein. Being a multifunctional protein, β-catenin is sequestered in different cellular compartments on account of the differential roles performed by this protein. In mammals, the proto oncogene *CTNNB1* encodes for β-catenin protein, and a major portion of the newly translated product of this gene associates with the cell membrane, where it acts as an adaptor protein, essential for cadherin mediated cell adhesion. The turnover of the

remaining pool of cellular and nuclear β-catenin is regulated by canonical Wnt signaling (Fig 1-2) (Willert and Nusse, 1998; Brembeck and Rosario, 2006).

1.2.2 β-catenin in adherens junction

Immunoprecipitation assays of Ca^{2+} dependent cell adhesion molecule uvomorulin (E-cadherin) performed by two groups independently in the 1980s, resulted in co-precipitation of three other proteins along with uvomorulin (Vestweber and Kemler, 1984; Peyrieras, Louvard and Jacob, 1985). Later, in 1989, Rolf Kemler and colleagues found these three proteins to be associated with the cytoplasmic domain of uvomorulin, possibly linking it to the cytoskeleton, and named them α β and Y catenins (Catena in Latin meaning chain).The α β and γ catenins were found to be structurally related in various species, including the mouse, humans and avians, implicating a functional conservation across species (Ozawa, Baribault and Kemler, 1989).

Furthermore, the elucidation of the molecular interaction of the cadherin-catenin complex, especially the more extensively studied E-cadherin, β and α-catenin interactions revealed that immediately following its synthesis, E cadherin associates with β-catenin through its cytoplasmic domain, and this interaction with β-catenin is essential for efficient exit of E-cadherin from ER to the baso-lateral membrane (Chen, Stewart and Nelson, 1999). Following membrane trafficking, the E-cadherin- β-catenin complex interacts with α-catenin (Hinck *et al.*, 1994). It is well established that α-catenin has binding sites for actin (Rimm *et al.*, 1995). However, contrary evidence exists whether or not α-catenin physically links the E-cadherin β-catenin complex to actin filaments. Studies have suggested the presence of a dynamic interaction of α-catenin with actin filament that is mutually exclusive to binding β-catenin, owing to the specificities conferred by the different molecular conformations of this allosteric protein (Drees *et al.*, 2005; Yamada *et al.*, 2005).

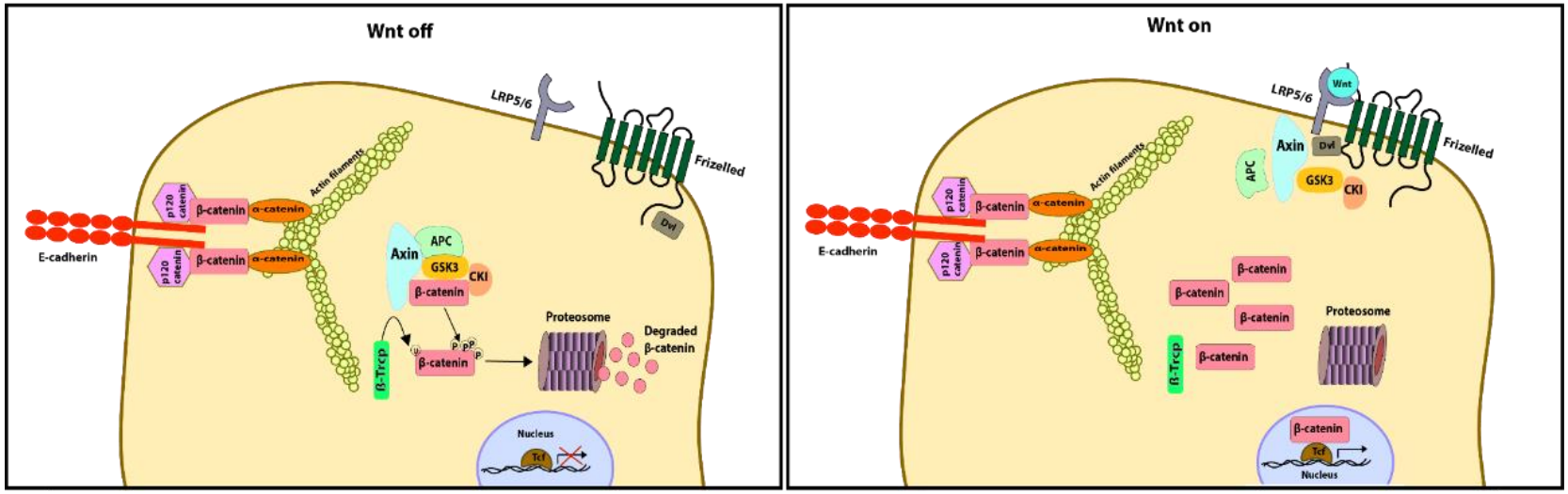


Figure 1-2: Canonical Wnt signalling cascade. In the membrane, β-catenin, acts as an adaptor protein essential for cadherin mediated cell adhesion. The turnover of the cytosolic and nuclear pool of β-catenin is mostly regulated by canonical Wnt signaling. The current canonical Wnt pathway is based on the on and off model, where in the absence of Wnt ligand, β-catenin is sequestered in the destruction complex and is sequentially phosphorylated at the serine and threonine residues which in-turn tags it for ubiquitin mediated proteosomal degradation. In the absence of β-catenin, the TCF/Lef transcription factors are bound by repressors such as Groucho thus preventing the expression of Wnt target genes. Binding of the Wnt ligand to the heterodimeric receptor leads to the stabilization of β-catenin which then translocates to the nucleus and interacts with TCF/Lef transcriptional factors activating the expression of Wnt target genes.

Although the intricate molecular details remain inconclusive, it is widely accepted that the core cadherin-catenin complex along with additional catenins, including p120-catenin, plakoglobin/γ-catenin, and various other junctional proteins, contribute to the formation of adherens junction, conferring cell-cell interaction through reorganization of the actin cytoskeleton mediated by α-catenin. The stability of E-cadherin based adherens junction is regulated by phosphorylation/dephosphorylation at various residues of the interacting proteins, including E-cadherin, β-catenin and p120-catenin (Bertocchi, Rao and Zaidel-bar, 2012). The phosphorylation especially at the tyrosine residues 142, 489 and 654 in the armadillo repeat of β-catenin inhibit the protein-protein interactions with α-catenin and E-cadherin, respectively (Roura *et al.*, 1999; Piedra *et al.*, 2003; Winter, Shasby and Shasby, 2008).

1.2.3 β-catenin protein structure

The characterization of protein structure has helped to deduce the pleiotropic association of β-catenin with various components involved in both cell adhesion and signal transduction. Biochemical assays studying the crystal structure of β-catenin complexed with binding partners including T-cell factor (TCF), inhibitor of β-catenin and TCF (ICAT), Beta-trasducin repeat-containing protein-S-phase kinase associated protein 1 (β-TrCP-Skp1), adenomatous polyposis coli (APC), Axin and E-cadherin, have helped to pinpoint exact domains and inter molecular bonding specificities required for protein-protein interactions.

In humans, the *CTNNB1* gene (41.02Kb – Ensemble Human GRCh38) consisting of 14 coding exons is translated into a 781 amino acid product, yielding a primary structure composed of three main domains – the NH_2 terminal domain, the central armadillo repeat domain and the COOH terminal domain (Fig 1-3). The highly conserved central Armadillo repeat domain consists of 12 arm repeats, each of approximately 42aa that forms three alpha helices linked by short loops (Riggleman, 1989; Huber, Nelson and Weis, 1997). The helices of the 12 continuous repeat units form a right handed superhelix, and in course a positively charged shallow groove is generated with potential binding surface for a number of β-catenin interacting proteins. The positively charged surface of the arm repeat domain, specifically the groove of the superhelix, is the predominant binding site for a majority of β-catenin partners involved in cell adhesion and Wnt signaling. The

binding sites for a large number of these proteins, including E-Cadherin, APC, TCF and the Wnt signaling inhibitor ICAT overlap with each other, and hence the interaction of these proteins with β-catenin is known to be mutually exclusive (Xu and Kimelman, 2007). In addition, the arm repeat domain forms the binding site for various co-activators and inhibitors of canonical Wnt signaling, making it a major focal point for interaction with β-catenin.

Although the three dimensional crystal structure of arm repeat domain has been studied in detail, not much is known about the N and C terminal domains. However, the N and C terminal domains are known to be negatively charged. A part of the C-terminal domain forms an alpha helix known as helix C, which in turn caps the hydrophobic residues of the 12^{th} arm repeat (Xing et al. 2008). The crystal structures obtained from β-catenin peptide complexes with β-TrCP have indicated the presence of a short helical structure even at the N-terminal domain (Megy *et al.*, 2005). The N-terminal domain DSG*φ*XS motif phosphorylated at the two serine residues 33 and 37 is the major recognition and site of contact for β-TrCP, which ubiquitinates β-catenin, flagging it for proteosomal degradation (Wu, Xu, B. a. Schulman, *et al.*, 2003). In addition, the extended helix present prior to the first arm repeat at the distal end of the N terminal domain (residue 118-146) provide binding sites for α-catenin (Pokutta and Weis, 2000).

The exact binding specificities of the C-terminal domain is yet to be resolved, however, β-catenin C-terminal domain and Gal4 fusion based artificial reporter systems have demonstrated this region to be a transactivation domain. Furthermore, the requirement of C-terminal domain in lymphoid enhancer-binding factor 1 (Lef1) mediated signaling by *in vivo* study in *Xenopus laevis*, complemented these findings, establishing the role of C-terminal domain as a scaffold domain for binding of transcriptional factors (Wetering *et al.*, 1997; Vleminckx, Kemler and Hecht, 1999).

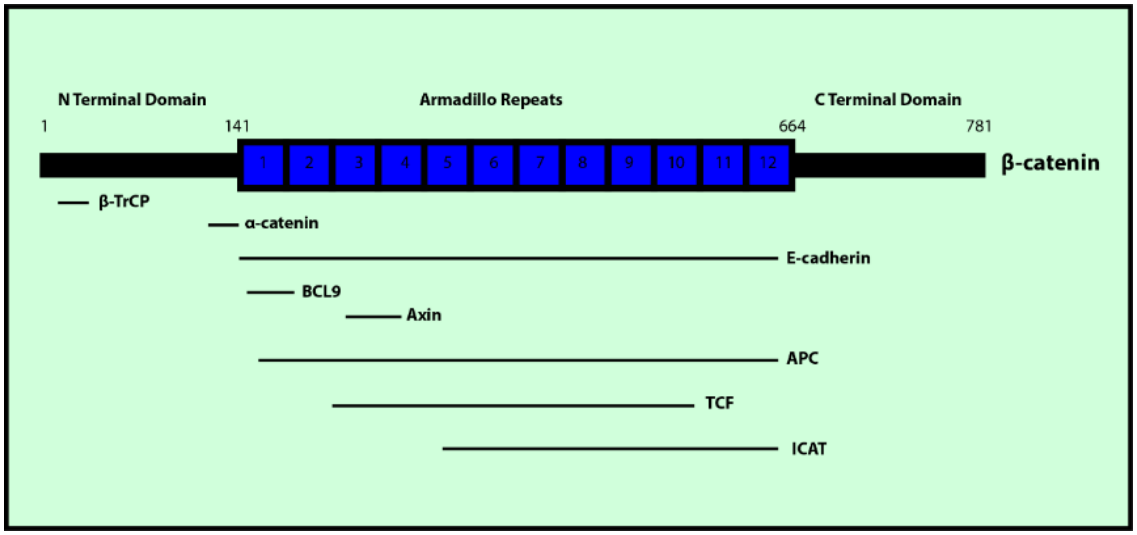


Figure 1-3: β-catenin protein domain structure. The N-terminal domain consists of the regulatory residues important for β-catenin stabilization and also contains binding site for α-catenin. The twelve armadillo repeats are the major hub for binding of majority of the interacting proteins including E-cadherin, BCL9, Axin, APC and ICAT. The C-terminal domain is known to be the transactivation domain of β-catenin.

1.2.4 On/Off model of Wnt/ β-catenin mediated pathway

According to the current on and off model of Wnt signaling, the kinase dependent ubiquitin mediated proteosomal degradation of β-catenin is the key regulatory event governing the stability of β-catenin, which in turn helps in maintaining low cytoplasmic level of this major downstream effector (Aberle *et al.*, 1997). In the absence of Wnt signaling, the so called destruction complex, consisting of the scaffold protein Axin, the APC tumour suppressor protein and the two kinases glycogen synthase kinase3 (GSK3) and Casein kinase 1 (CK1), sequesters and phosphorylates β-catenin (Ikeda *et al.*, 1998; Kishida *et al.*, 1998; Sakanaka, Weiss and Williams, 1998; Liu *et al.*, 2002). The initial phosphorylation of β-catenin by CK1 at residue S45 is the priming event for the subsequent GSK3β mediated sequential phosphorylation of residues T41, S37 and S33 (Fig 1-4) (Liu *et al.*, 2002). The phosphorylation of the two serine residues in the consensus degron motif DSG*φ*XS is especially considered to be crucial for the recognition and interaction by the F box protein β-TrCP, a component of the E3 ubiquitin ligase complex (Hart *et al.*, 1999). This recognition of the doubly phosphorylated motif by the WD40 domain of β-TrCP is followed by ubiquitination and proteosomal degradation

of β-catenin, and in the absence of nuclear β-catenin, the TCF/Lef interaction with repressors such as Groucho/TLE prevents the expression of Wnt target genes, keeping the pathway switched off (Hart *et al.*, 1999; Cadigan and Waterman, 2012).

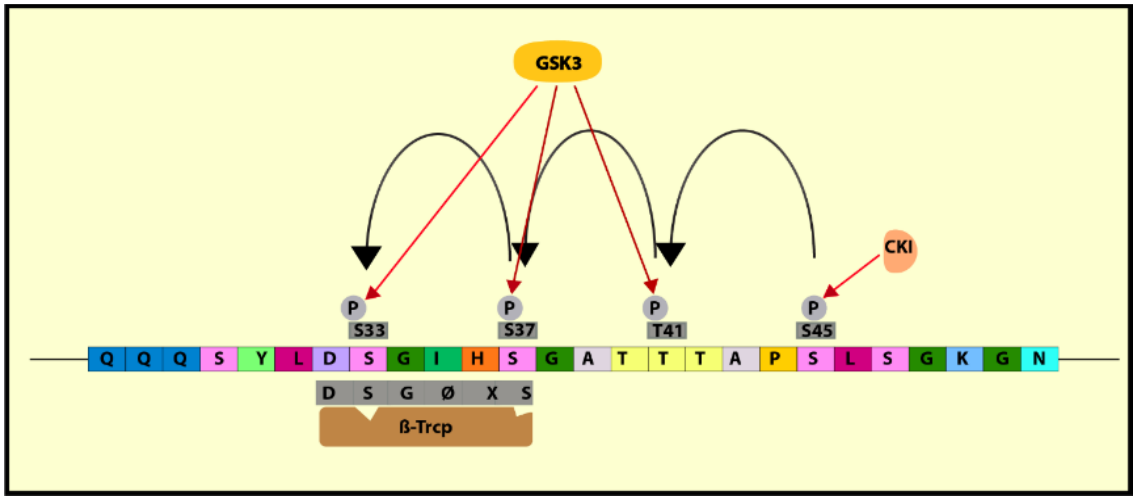


Figure 1-4: Regulation of β-catenin stability by sequential Phosphorylation and β-TrCP mediated proteosomal degradation. The initial phosphorylation of S45 by CKI acts as a priming event for sequential Phosphorylation of T41, S37 and S33 by GSK3β, forming the recognition site for β-TrCP, which then mediates ubiquitin tagging and thus flagging β-catenin for proteosomal degradation.

The canonical Wnt pathway is activated by binding of the secretory Wnt proteins to the serpentine (G protein coupled) receptors Frizzled and the single pass LRP5/6 co-receptors, which belong to the low density lipoprotein receptor family (Macdonald and He, 2012). Ligand binding to the heterodimeric receptor induces conformational change and phosphorylation of LRPs (Liu *et al.*, 2003; Zeng *et al.*, 2017). The recruitment of the effector protein Dsh and the Axin bound destruction complex by the activated receptor is known to play a crucial role in mediating β-catenin stability, however, the exact molecular mechanisms governing the stability of β-catenin upon activation of the pathway remains elusive, and is subject to controversy (Zeng *et al.*, 2017). Various different models have been proposed on the mechanism of stabilization of β-catenin, a common theme being the inhibition of β-catenin degradation, resulting in its cytosolic accumulation (Gerlach *et al.*, 2014). The cytoplasmic β-catenin is then translocated to the nucleus where it competes with the Groucho/TLE repressors for TCF/Lef binding site, and on interaction

with TCF/Lef, it acts as a transcriptional co-activator, leading to the expression of Wnt responsive genes and turning the pathway on (Daniels and Weis, 2005). However, the mechanism of shuttling of β-catenin between cytoplasmic and nuclear compartments again remains unclear. Reports suggests that, in the absence of a nuclear localization signal (NLS), β-catenin directly binds to the nuclear pore complex without the requirement for classical transport proteins, such as importin and Ran (Fagotto, Glück and Gumbiner, 1998; Yokoya *et al.*, 1999).

The bipartite transcriptional function of TCF/Lef is mediated by binding to various activators and co-repressors, acting likely through the chromatin remodeling ability of these direct/indirect binding partners. The covalent modifications, especially tinkering the acetylation and methylation state of the chromatin, plays a significant role in modulating the target gene transcription. Binding of the histone acetyltransferase (HAT) encoding proteins CBP/P300 to the C-terminal transactivation domain of β-catenin, results in increased acetylation of H3 and H4 subunits of the histone moiety, and enhances the transcriptional activation of Wnt target genes (Parker *et al.*, 2008). In addition to covalent modification, the Brg1 of the SWI/SNF complex, acting through their intrinsic ATPase domain are capable of remodeling the histone (Barker *et al.*, 2001). In addition, H3K4me3 catalysed by MLL/Set1 family of methyltransferases bound to β-catenin induces an open chromatin mark (Sierra *et al.*, 2006) (Wend *et al.*, 2013). The N terminal transactivation domain of β-catenin is also known to recruit adaptor protein Bcl9. The Bcl9 protein mediates the interaction between Pygopus and β-catenin, and this interaction is known to promote transcriptional activation. Studies suggest that Pygo 2 acts as a scaffold facilitating RNApol II mediated transcription of Wnt target genes, either indirectly by recruitment of the mediator complex, or by direct interaction with the transcription factor II D (TFIID) subunit of the basic transcriptional machinery (Carrera *et al.*, 2008; Wright and Tjian, 2009). Given the significance of Wnt signaling in a myriad of physiological processes, the target genes under direct transcriptional control of canonical Wnt signaling are differentially regulated by the coordinated action of the β-catenin interacting nuclear transcriptional machinery, thus orchestrating the activation of gene expression in a context dependent manner. The list of Wnt target gene is constantly increasing in number, and in addition to regulating the expression of proteins that govern the observed pleotropic effects of Wnt signaling, few of the mediators of the signaling cascade are

themselves targets of Wnt/β-catenin mediated signal transduction. The scaffold protein Axin2/Conductin is one such target gene under the control of Wnt signaling, and upregulation resulting from increased Wnt activation observed in tumours, is known to constitute a negative feedback loop, which in turn represses the Wnt cascade (Lustig *et al.*, 2002).

In addition to the feedback mechanisms, Wnt signaling is modulated by the regulatory action of various antagonists and agonists. The six major secreted antagonist including, secreted frizzled-related protein (Sfrp), Wnt inhibitory factor 1 (WIF1), Cerebrus, Wise/Sclerostin (SOST), insulin- like growth factor-binding protein 4 (IGFBP4) and Dickkopf 1 (DKK1), can be classified into two families based on their ability to either directly bind to Wnt, or to the Frizzled and LRP5/6 co receptors. In addition to secreted antagonists, multiple transmembrane bound proteins including Shisa, Waif1/5T4, APC down-regulated 1 (APCDD1) and Tiki1 are also known to negatively regulate Wnt signaling. Norrin and R Spondin are two of the well characterized agonist, activating Wnt signaling by binding to the Frizzled LRP complex, or the Leucine-rich repeat-containing G-protein coupled receptor 4/5 (LGR4/5) receptors, respectively (Kawano, 2003; Cruciat and Niehrs, 2013).

1.3 Wnt signaling in cancer

Since the initial discovery of oncogenic activation of *Int1* by insertion of MMTV, capable of initiating mammary tumours and adenocarcinomas in the mouse, several activating and loss of function mutations in various components of the Wnt signaling cascade have been reported in different cancer types.

The tumour suppressor gene *APC* was identified through its ability to induce colorectal adenomas (benign polyps) observed in variable numbers ranging from hundreds to thousands in familial adenomatous polyposis (FAP) patients. FAP is an inherited autosomal dominant disorder, and without surgical intervention the benign condition invariably progresses to malignancy. In addition to germline mutations in FAP, *APC* mutations have been frequently observed in sporadic colon cancers (80 percent), and to a lesser extent in other cancer types including desmoid, pancreas, stomach, breast and other cancer types. Majority of the *APC* mutations are confined to the 5' carboxy termini

with germline mutations distributed across the region, as opposed to somatic mutation specifically accumulated in the mutation cluster region (MCR), and most of these mutation results in protein truncation (Mori *et al.*, 1992).

The scaffold protein AXIN, another negative regulator of Wnt signaling has also been found to be mutated in various cancer types. Both variants *AXIN1* and *AXIN2* have been found to be mutated in hepatocellular carcinomas (HCC) and hepatoblastomas. The detection of loss of heterozygosity (LOH) associated with *AXIN1* in HCCs confirms the requirement of two-hits, validating its function as a tumour suppressor in these cancers (Taniguchi *et al.*, 2002). Mutations in *AXIN1/2* have also been reported in colorectal cancers (CRCs), and frameshift mutations in the exon 7 of *AXIN2* are particularly associated with a subset of CRCs with defective mismatch repair (MSI-H) (Liu *et al.*, 2000).

Furthermore, the transcriptional regulator *TCF4* have been observed to be frequently mutated in CRC with defective MMR (Duval, Gayet and Zhou, 1999). In addition to *TCF4*, mutations in transcriptional co-activators such as p300/CBP have also been reported in various solid tumours, as well as in certain lymphomas and leukemias (Iyer, Özdag and Caldas, 2004).

1.3.1 N terminal activating mutations in β-catenin – a common occurrence in cancer

The oncogenic potential of β-catenin was initially characterized based on the ability of N-terminally truncated β-catenin to transform NIH3T3 cells, implicated by the formation of transformed foci in a retroviral based complementary DNA (cDNA) library screen that identified 8 other novel oncogenes (Whitehead, Kirk and Kay, 1995). Over the years, small scale sequencing projects carried out by several independent groups have reported mutations in the β-catenin proto oncogene occurring at various frequencies in human tumour samples analysed from different tissue types having dominant activating function, thus validating the criterion of defining β-catenin as a 'proto-oncogene'.

The N terminal domain of β-catenin particularly involving the exon 3 region is highly susceptible to mutation (Saleem *et al.*, 2017). The N terminal interstitial deletions, majority of them involving the exon 3 region, are rare, and have been observed in certain

primary colorectal and hepatocellular carcinomas. A low frequency of β-catenin truncation mutations in colorectal carcinomas and adenomas was reported by Iwao et al. and Murata et al. in two groups of Japanese patients with only around 3 percent of them found to be carrying large interstitial deletions in the *CTNNB1* gene. The mechanism leading to these observed deletion mutations is not clear, but the presence of inverted repeats at both ends of the breakage site indicate a possible involvement of somatic rearrangement (Iwao *et al.*, 1998; Murata *et al.*, 2000).

In addition to the interstitial deletion in the exon 3 region, activating β-catenin mutations (mostly missense) at various residues have been reported in a number of primary human tumours. The exon 3 region consisting the regulatory sequences is known to be the mutational hotspot of *CTNNB1,* and numerous studies based on single strand conformation polymorphism and sequencing of the exon 3 region have been successful in delineating the precise missense substitutions in various cancer types.

Although loss of function mutation in *APC* is the leading cause of colorectal cancers, missense mutations at the phosphorylatable S and T residues have been reported in a subset of tumours with wild type (WT) *APC*. The mutual exclusivity of *APC* and β-catenin mutations is attributed to the common outcome of mediating β-catenin stability, leading to the TCF dependent transcriptional activation (Sparks *et al.*, 1998).

Similar activating β-catenin mutations affecting the GSK3β phosphorylation sites are observed in hepatocellular carcinoma. In addition, mutations surrounding the phosphorylatable S33 residue, particularly the D32 and G34 that are required for ubiquitin mediated proteosomal degradation were found to be commonly mutated (Miyoshi *et al.*, 1998; Omagnolo, Illuart and Nge, 1998; Legoix *et al.*, 1999). A strong correlation between nuclear expression and mutation in the β-catenin gene has been observed in these tumours (Mao *et al.*, 2001). The β-catenin target gene *GLUL*, encoding glutamine synthetase (GS) is often used as a marker for characterization of β-catenin mutations in HCC (Austinat *et al.*, 2008b). In addition, mutations in β-catenin have also been known to increase the production of bile and cholestasis and together with GS expression have been reported to be a good indicator of β-catenin mutations in human HCCs (Audard *et al.*, 2007).

Ovarian tumours especially of the endometroid subtype are particularly susceptible to β-catenin mutations with frequent nuclear accumulation of the mutant protein. The residues D32, S33, G34, S37, T41 and S45 have been reported to be mutated. The combined frequency of these activating mutations is around 30 percent in these primary endometrioid adenocarcinomas. Although variations exists between different subtypes, studies by Gamallo et al., suggests increased nuclear expression of β-catenin to correlate with a favourable prognosis (Gamallo *et al.*, 1999). The differential expression by RNase protection assay confirmed significant increase in the expression of five candidate target genes *MMP7*, *CCND*1, *CX43*, *PPAR-δ* and *ITF2* in a panel of 15 tumour samples with deregulated β-catenin (out of which 13 tumours had mutations in the β-catenin phosphorylatable S and T residues) compared to 17 tumours with WT β-catenin (Zhai *et al.*, 2002). However, the expression of *c-myc* was unaltered between the two groups. The same group also identified *FGF9* as a Wnt target gene that was upregulated in ovarian endometrioid adenocarcinomas with deregulated β-catenin (Hendrix *et al.*, 2006).

The earliest identification of the presence of missense mutations in the exon 3 region of β-catenin oncogene came from sequencing of a cell line from a melanoma patient, confirming the presence of a novel S37F substitution mutation (Robbins *et al.*, 1996). The frequency of β-catenin mutations is very low in melanoma, however, a common form of skin tumour of the pilomatricoma type are known to harbor activating β-catenin mutations at increasingly high rates, present in around 61-75 percent of tumours (Chan *et al.*, 1999; Saleem *et al.*, 2017). The residues S33 S37 and T41 are the most commonly mutated residues with lower incidence of mutations at adjacent residues, including D32 G34, S47 and G48.

Sporadic desmoid tumours are another class of tumours in which increased incidence of β-catenin mutations have been reported. Around 70-90 percent of these tumours harbor mutations in β-catenin with the majority of the mutations confined to T41 (T41A) and S45 (S45F and P) residues (Mullen et al. 2013) (Guellec *et al.*, 2012). In an analysis of a panel of desmoid tumours, Lazar et al. have reported an increased risk of tumour recurrence in patients with S45F mutations, and presence of these mutations could hence act as prognostic tool for therapeutic intervention (Lazar *et al.*, 2008).

Mutations specifically at residue S45 involving serine to phenylalanine and tyrosine residue and in-frame 3 base pair (bp) deletion of residue S45 have been observed in patients with Wilms tumour. Furthermore, these activating β-catenin mutations are frequently associated with mutations in the *WT1* gene (Maiti *et al.*, 2000).

In addition to the tumour types above, mutations in the regulatory β-catenin residues (and adjacent residues) have been observed in cancers including CNS, bone, breast, salivary, kidney, adrenal gland, lung, pancreas and various other tumors in differing frequencies. Although the differences in the tumourigenic potential of these individual mutations has not been clearly established, the general hypothesis is that every activating mutation in-turn governs the turnover and localization of this central mediator of Wnt pathway, making it refractory to regulation and thus activating the Wnt target genes capable of modulating multiple aspects of tumourigenesis. Among the target genes are *c-myc, CCND1*, *MMP*s, *VEGF*, *FGF9*, *uPAR*, *PPARδ*, *AXIN2*, all known to play significant roles in various stages of the tumourigenic process thus exemplifying the implications of deregulated Wnt signaling in cancer.

1.4 Problems in the current on and off model of canonical Wnt signaling

The widely accepted model of canonical Wnt signaling cascade is based on the activity response regulated by the stability of β-catenin protein. However, several evidences indicate the current model of β-catenin regulation to be incomplete. The investigation of the Wnt mediators, especially the analysis of the mutational spectrum of the *APC* tumour suppressor gene, and the subsequent functional analysis of the *Apc* allelic series were among the defining studies highlighting the drawbacks in the Wnt on and off model. The non-random and specific induction of a second hit in the *APC* tumour suppressor gene observed in intestinal tumours was shown to be dependent on the type of existing germline mutation. This selection bias was contrary to the random occurrence of the two hits proposed by Knudson's two hit theory. Although all the mutation combinations resulted in the truncation of APC, these different mutations were proposed to have retained residual β-catenin regulatory activity, albeit at varying strengths, determining the specificity of mutational selection, as opposed to the complete loss of function leading to

a single functional outcome. This mutational selection based on the tumour specific requirement known as the 'just right signaling model' provided a link between an optimal β-catenin activity level and cancer (Albuquerque *et al.*, 2002).

Modelling of *APC* tumour suppressor mutations observed in intestinal tumours and functional analysis in *in vivo* and *in vitro* systems have provided experimental evidence for this proposed model as to how the developing tumours select for an *APC* mutation according to the level of β-catenin activity. Using mouse embryonic stem cells, Kielman et al. have shown that the various inactivating mutations in the *Apc* tumour suppressor gene differed in their differentiation potential, determined by the specific β-catenin dosage conferred by the allelic variants (Kielman *et al.*, 2002). Additional support to this model was provided by *in vivo* studies of mice tumours; the *Apc* allele ($Apc^{+/1572T}$) with lower β-catenin activity resulted in multifocal mammary tumours as opposed to the increased incidence of extra intestinal tumours including cutaneous cyst and desmoid tumours observed in $Apc^{+/1638N}$ with lower multiplicity of intestinal tumours. The reduced multiplicity of intestinal tumours in these $Apc^{+/1638N}$ mice that had an intermediate level of β-catenin activity differed phenotypically from the $Apc^{+/min}$ mice (with a high β-catenin activity) that primarily developed intestinal tumours with higher multiplicity (Gaspar *et al.*, 2009). Furthermore, crossing the *Apc*1638N mice with heterozygous β-catenin knockout mice, and in effect reducing β-catenin signaling, altered the 1638N phenotype to the 1572T like phenotype, resulting in an increased incidence of mammary tumours. This experiment provided evidence to support the claim that different β-catenin signaling levels are the driving force behind these cancer phenotypes (Bakker *et al.*, 2013).

In addition to the mutational bias and differential β-catenin downregulating activity of the *APC* mutant allele, few observations suggests a similar genotype-phenotype correlation directly associated with mutations in the β-catenin proto-oncogene. The preferential selection of S45 mutation in Wilms tumour samples, accounting for about 90 percent of the tumours with β-catenin mutations, suggests a selection bias of specific mutations similar to those observed for *APC* tumour suppressor gene. The functional analysis of the β-catenin mutations by Provost et al, provided evidence of a differential phenotypic response conferred by of the four phosphorylatable residues (Provost *et al.*, 2003). In this study, the stable over expression of alanine variants of S45 and T41 residue were found

to give rise to a more transformed phenotype in comparison to S37 and S33 variants. These results cannot be explained by the current model of regulation of β-catenin stability based on the sequential phosphorylation. In this regard, contrary to the sequential cascade, it has been shown that in colorectal cancer cells with S45 mutation, the phosphorylation of residues T41, S37 and S33 can still take place even in the absence of phosphorylation at residue S45 (Wang, Vogelstein and Kinzler, 2003). Taken together, these evidences strongly suggests the presence of additional regulatory mechanisms in controlling the β-catenin activity that remains to be understood.

1.5 Gene targeting in mouse embryonic stem cells (mESCs) and transgenic technology

In the 1980s, independent studies by Mario Capecchi and Oliver Smithies, provided the first evidence of active homologous recombination (HR) in cultured mammalian cells by their ability to replace or insert exogenous DNA sequence into the endogenous loci of interest (Smithies *et al.*, 1985; Thomas, Folger and Capecchi, 1986). The most promising use of gene targeting was following the successful isolation of mouse embryonic stem cells by Evans and colleagues, and discovering that the reintroduction of ES cells into blastocyst following *in vitro* culture produced germline chimeras, opened exciting prospects for genetic engineering (Bradley *et al.*, 1984). Gene targeting in mESCs followed by the generation of transgenic mice, together provided a perfect tool for studying various genetic conditions.

Initially, targeting approaches based on homologous recombination strategies were tested on selectable genes. Particularly, the hypoxanthine-guanine phosphoribosyltransferase (*Hprt)* locus was a common target that allowed for phenotypic selection using the cytotoxic chemical 6 thioguanaine (Thomas and Capecchi, 1987). Later, polymerase chain reaction (PCR) based screening and enrichment strategies were incorporated that could be extended to non-selectable or even non-expressed genes. One such enrichment technique was the use of a positive negative strategy (PNS) described by Mario Capecchi in 1988, for successful targeting of the *hprt* and *int2* gene in mouse ES cells (Capecchi, 1989). Enrichment for selection of targeted recombinants against random integration was achieved using neomycin (G418 based positive

selection) and herpes simplex virus type-1 – thymidine kinase (HSV1-tk ganciclovir based negative selection) selectable markers (Mansour, Thomas and Capecchi, 1988). Although these conventional gene targeting strategies had earlier proved to be standard tools for genome engineering and numerous knockout transgenic mouse models were successfully generated, the irreversible integration of selectable markers and the inability to control gene expression resulting in lethal phenotypes (due to presence and expression of mutant gene during early stages of development) were the two major drawbacks of this system.

The introduction of conditional targeting using site specific recombinase technology based on the activity of recombinase enzyme of bacterial origin including CRE and flippase (FLP) have helped to overcome some of the drawbacks of earlier conventional targeting strategies. The regulated expression of these enzymes under tissue specific promoters along with inducible systems such as tamoxifen inducible oestrogen-LBD-Cre/flp hybrid or use of inducible promoters such as tetracyclin (Tet) on/off system allowed spatio-temporal control of gene expression (Ryding, Sharp and Mullins, 2001).

1.5.1 Enhanced genome editing using Designer nucleases

The conventional gene targeting and conditional and inducible genome engineering tools are all based on the active HR. The HR mediated repair although provides an error proof system for precise in-situ gene manipulation, the low frequency of recombination events (1 in 10^6 – 10^9 cells) which results in reduced targeting efficiency has been a major drawback of experimental systems based on the HR repair pathway. However, deliberate introduction of double-strand breaks (DSBs) by I-SceI and other homing endonucleases have shown to increase the frequency of HR mediated repair up to a 1000 fold or more (Jasin, 1996). Although increased specificity can be achieved by the relatively long recognition sequences (14-40bp), the applicability of these homing endonucleases as genome engineering tools was significantly reduced due to the complex interaction with the target DNA involving 40-50 amino acid chains, making it difficult to modifying the DNA modular domain of these proteins.

The increased targeting efficiency mediated by endonuclease induced DSB directed repair mechanisms have led to the characterization of similar customizable nucleases, including zinc- finger nucleases (ZFNs), transcription activator like effector nucleases

(TALENs) and CRISPR/Cas systems, providing valuable tools for precise editing of otherwise difficult to manipulate in-vivo and in-vitro systems. ZFNs are engineered nucleases constructed by combining the DNA binding specificity of the eukaryotic transcription factors zinc finger proteins (ZFPs) and the nuclease domain of the TypeIIS restriction enzyme FokI (Kim, Cha and Chandrasegaran, 1996). Based on the sequence of target DNA, the recognition domain of ZFNs can be constructed by modular assembly of an array of ZFPs usually consisting of 3-6 DNA binding Cys2His2 zinc finger domains, and selecting the appropriate fingers that have been developed for identifying a specific nucleotide triplet allows accurate recognition of the region of interest (Wright *et al.*, 2006). However, as FokI functions as a dimeric complex incorporating two adjacent recognition sequences (Vanamee, Santagata and Aggarwal, 2001), the ZFN must be constructed to include two monomeric ZFP arrays recognizing 9/12 bp of target DNA separated by a spacer of 5-6 nucleotides between the two inverted recognition sequences. The two ZFP arrays (Mani et al 2005) allow appropriate recognition and heterodimerization of the nuclease essential for active DSB induction in the spacer region.

Similar to ZFNs, TALENs have a customizable DNA binding domain again fused to the nuclease domain FokI enzyme. The highly conserved TALE are typeIII effectors secreted by bacterial plant pathogens of Xanthomonas sps and fuction as trancriptional activators of a variety of plant genes. The repeat domain of TALE consist of an array of repeats each 33-35 amino acid flanked by additional TALE sequences specifically recognizing a single bp of target DNA, and hence the number of repeats in the array corresponds to the length of the target DNA recognized (Boch and Bonas, 2010). The specificity of DNA recognition of each repeat in an array is conferred by the two hypervariable residues at position 12 and 13 often referred to as the repeat variable di-residue (RVD). Modular assembly using 4 RVDs each with preferential binding to A,G,T,C using various assembly strategies have been described and as FokI functions as a dimer, target specific TALEN pair each fused to a FokI monomer allows induction of DSBs (Boch and Bonas, 2010).

Compared to ZFNs and TALENs which are based on protein-DNA interaction, the CRISPR/Cas system is based on the Watson-Crick base pairing rules henceforth making it a simple yet efficient technique with implications in a wide range of biological applications.

1.5.1.1 Genome engineering using CRISPR-Cas9 system

The term CRISPR was coined by Mojica et al. and Jansen et al. to describe a class consisting of an array of direct repeats separated by short spacer sequences (known as protospacers in bacteriophages) (Jansen *et al.*, 2002). The presence of these direct repeats initially described in 1987 in *E. coli* was later found to be a common feature in the majority of bacterial and archeal strains (Ishino *et al.*, 1987; Mojica, F.J., Díez-Villaseñor, C., Soria, E. and Juez, 2000). Absent in eukaryotic genomes, these direct repeats and spacers were found to be well conserved, but varied in numbers across different strains. Importantly, the newly acquired spacer sequences were found to be similar to, and in-fact derived from extrachromosomal DNA such as phage replicons and plasmids, implicating the presence of a sequence specific nucleic acid based adaptive immunity against bacteriophage (Bolotin *et al.*, 2005;. Mojica *et al*., 2005; Rodolphe *et al.*, 2007). In addition, characterization of CRISPR loci revealed a close association with Cas genes that were almost always found located adjacent to these direct repeats, signifying a functional relatedness between the two (Jansen *et al.*, 2002). Over two decades, genetic and biochemical studies have helped unravel the mechanistic details of the CRISPR Cas mediated bacterial immunity.

The type II CRISPR system from *Streptococcus pyogenes* is the most extensively characterized system that quickly found applicability in genome engineering (Marraffini, 2016). Initially, the direct repeats together with spacers of the typeII CRISPR loci are transcribed into a precursor CRISPR RNA (crRNA) array. After transcription, this pre-crRNA array associates with its auxiliary partner the trans activating RNA (tracrRNA), and directs the RNaseIII dependent processing of the pre crRNA into mature crRNA, containing a 20 nucleotide guide and a part of the repeat sequence (by RNase III) (Deltcheva *et al.*, 2011). The processed guide then associates with Cas9 protein forming an effector complex, and directs the Cas9 to the target locus (protospacer). However, the initial recognition of a protospacer adjacent motif (PAM) is crucial for the subsequent base pairing and Cas9 mediated induction of a DSB (Sternberg *et al.*, 2014).The CRISPR/Cas9 system from *S. pyogenes* has a 5'NGG PAM requirement with the Cas9 induced DSB occurring three base pairs upstream of the PAM motif. The detailed understanding of the mechanism of this bacterial adaptive immune system has helped to engineer programmable Cas9 nucleases that can be used to modulate the endogenous

activity of genes in virtually any organism. Using a two component system consisting of single guide RNA (sgRNA), designed by fusing crRNA and tracrRNA and a mammalian codon optimized Cas9, and only by altering the 20 nucleotide guide sequence, it is now possible to direct the Cas9 to the gene of interest in mammalian cells (Fig 1-5) (Jinek *et al.*, 2012; Cong *et al.*, 2013; Mali *et al.*, 2013).

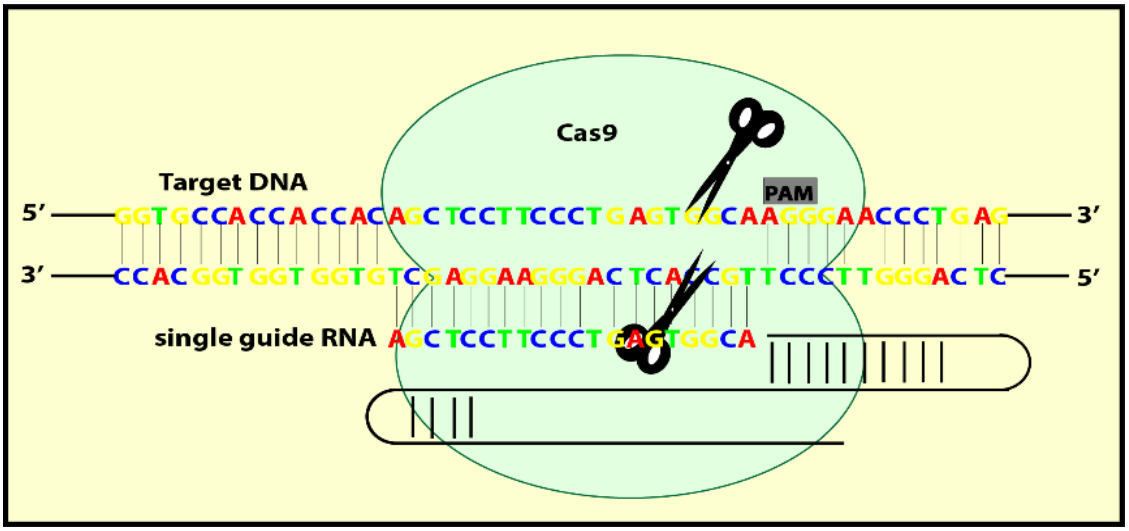


Figure 1-5: Schematic representation of Cas9 nucleases guided by sgRNA. The Cas9 nuclease from *Streptococcus pyogenes* is directed to the target site by the 20 nt guide oligo within the sgRNA followed by Watson Crick base pairing with the target sequence. The Cas9 then induces a double strand break 3 bp upstream of the PAM (5'-NGG).

In addition to the wild type Cas9 nuclease described above, various variants have been described. A Cas9 'nickase' version called Cas9n with a point mutation in the D10 residue (D10A), rendering the RuvC domain inactive, but retaining the ability to induce single strand nicks by the active HNH domain was generated (Jinek *et al.*, 2012). This Cas9n mutant, when targeted along with a pair of appropriately spaced offset guide RNA's, each directed towards opposite strands of the region of interest, induces a DSB while significantly reducing the off-target effects (Ran *et al.*, 2013). The recently described high fidelity spCas9-HF1, with four substitutions (N497A,R661A, Q695A/Q926A) is shown to be capable of reducing off-target effects, while still maintaining a high degree of specificity (Kleinstiver *et al.*, 2016). To overcome the constraints of the stringent requirement for an NGG PAM, for Cas9 recognition of target DNA, different variants with altered PAM specificities have been generated (Kleinstiver *et al.*, 2015). The orthologues of Cas9 from

other bacterial species such as *Staphylococcus aureus*, *Streptococcus thermophile*s have also been characterized and engineered to recognize alternative PAM sequences, to increase the applicability of CRISPR Cas9 system to target a broader spectrum of sequences, and the smaller size of these enzymes provide added advantage of efficient delivery into *in vivo* systems by packaging into the adeno- associated virus (AAV)(Cong *et al.*, 2013; Ran *et al*., 2015; Kleinstiver *et al.*, 2015).

Since its introduction, the RNA guided endonuclease activity of bacterial CRISPR/Cas9 system has been harnessed in combination with the inherent cellular repair pathway of either non-homologous end joining (NHEJ) or homology directed repair (HDR), for the introduction of insertions deletions (indel), epitope tagging, or precise point mutations and has provided an efficient strategy for genome alteration both in in-vivo and in-vitro systems. In addition to gene editing, the CRISPR/Cas9 system can also be used for regulated gene expression. The catalytically inactive Cas9, termed dead Cas9 (dCas9-D10A/H840A mutant), targeted along with guide RNA is an effective strategy for repressing the expression of the known target gene. This strategy termed as CRISPR interference (CRISPRi) has been successfully used for gene silencing (Qi *et al.*, 2013). The dCas9 fused to effector proteins, such as Kruppel associated box (KRAB) can be used to enhance the specificity of efficient transcriptional repression. Similarly, fusion of dCas9 to transcriptional activators, such as VP64 can be used for targeted gene activation (Maeder *et al*., 2014). The CRISPR/Cas9 system has also found applicability in inducing additional epigenome modifications, by coupling dCas9 to various effector proteins for changing the methylation state of the chromatin, and also for histone modifications (Huangfu and Raya, 2017). The wide range of applicability, complemented with the increasing availability of diverse variants and optimization strategies, has made the CRISPR/Cas9 system a robust tool for programming enhanced genome engineering strategies that can be tailored to the specific needs.

1.6 DNA damage and repair response

The loss of integrity of the macromolecular structure of DNA, a consequence of its susceptibility to the various endogenous and exogenous agents capable of inducing wide spread repercussions on the robustness of an organism is a common feature observable throughout its lifetime. The intrinsic susceptibility of glycosyl bonds to hydrolysis resulting

in apurinic sites, sensitivity to hydrolytic deamination (especially the nucleotide bases of cytosine and 5 methyl cytosine), oxidative stress induced by endogenous metabolities capable of generating various base lesions including transversions, conversions of pyrimidine to ring saturated forms, helical distortions, non-enzymatic S-adenosylmethionine induced DNA methylation, are few of the many DNA lesions caused due to endogenous DNA damaging agents (Withers *et al.*, 1993). Physical exogenous agents, including ionizing radiation are capable of inducing single and double strand breaks and exposure to ultra violet rays is capable of generating pyrimidine dimers with increased frequency. In addition, chemical agents such as chemotherapeutic drugs acting as either alkylating agents or by forming intra and inter covalent linkages contribute to the added genotoxic burden. These exogenous and endogenous DNA damaging agents, combined with the errors generated during the intrinsic replicative process, result in increased incidence of replicative stress leading to toxic cellular effects. However, each of these DNA damages trigger specific DNA damage response (DDR) events to restore the integrity and prevent the catastrophic effects of the unrepaired genotoxic lesions. Various DDR pathways have been described including, the well-studied mismatch repair, nucleotide exchange repair, base excision repair (BER), DNA double strand break repair pathways. These DNA repair pathways play a crucial role in maintaining the genomic integrity of an organism.

1.6.1 DSB induced repair mechanisms

Of the various DNA lesions, DSBs caused by nicks in complementary DNA strands are known to be highly genotoxic. The induction of DSB in DNA by extra or intra cellular agents triggers the activation of either one of the two major cellular repair pathways, namely: the error prone NHEJ, or the high fidelity HDR, both of which help maintain the structural integrity of the chromosomes (Pastink, Eeken and Lohman, 2001). The Ku mediated canonical NHEJ offers a higher mechanistic flexibility, the independent enzymatic activity of the nucleases, polymerases and ligases, in combination with iterative processing, contribute to the diverse array of junctional outcomes, including insertions, deletions, inversions and direct repeats at the break point. The NHEJ pathway, in addition, can operate in any stage of the cell cycle without the requirement of a homologous repair template making it a more frequent potentiator of DSB repair. However, homologous recombination, the most common form of HDR is functional during

the late S/G2 phase of the cell cycle due to the presence of sister chromatid that provides a preferred homologous template for HR mediated repair (Lieber, 2010). The resection of DNA ends determines the induction of repair by either NHEJ or HR pathways. In the case of NHEJ, the binding of the Ku protein prevents excessive end resection whereas resection of the 5' end at break point resulting in extended 3' single strand DNA (ssDNA) overhangs mediated by the MRN core complex is necessary for HR directed repair. The recombinase protein Rad51 assembly onto ssDNA, leads to the formation of a nucleoprotein filament that invades the homologous DNA forming the D loop structure. The D loop intermediate is then resolved by either synthesis dependent strand annealing (SDSA), or second strand capture, resulting in double Holliday junction that can be resolved by cross over and non-cross over events (San Fillipo, Sung and Klein, 2008).

The RNA guided endonuclease activity of bacterial CRISPR/Cas9 system harnessed in combination with the inherent cellular repair pathway of either NHEJ or HDR, for the introduction of insertions, deletions or precise point mutation has provided an efficient strategy for genome alteration (Fig 1-6).

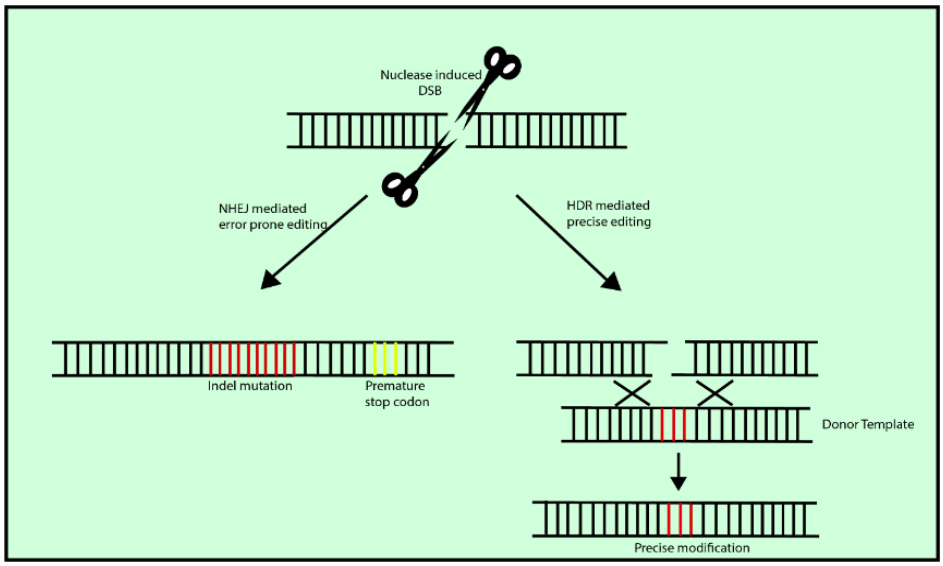


Figure 1-6: Repair pathways induced by double strand breaks. DSBs induced by nucleases can either be repaired by NHEJ or HDR. The NHEJ pathway is an error prone mechanism that can introduce indels creating a premature stop codon that may result in gene knockout. Alternatively, a repair template can be supplied to introduce precise editing mediated by HDR repair pathway.

1.7 Aim of my thesis

The current model of β-catenin activity is based on the stabilization of the protein, which in turn is coupled to the kinase dependent sequential phosphorylation of the serine and threonine residues. According to this model, the phosphorylation of the S45 residue by CK1 acts as a priming event, which leads to the GSK3 mediated phosphorylation of the T41, S37 and S33 residues. Once the last two residues have been phosphorylated, a docking site for the E3 ubiquitin ligase is formed, and β-catenin is marked for degradation. Since this is a sequential cascade, mutations at any of these residues result in the same outcome, i.e. the stabilization of β-catenin, and ectopic activation of the pathway. While this type of regulation mechanism suggests an equal selection of these mutations in tumours, this is not the case in Wilms tumours. In these tumours, mutations in the β-catenin oncogene are largely biased towards one specific residue, S45, accounting for about 90 percent of tumours with β-catenin mutations. This preferential selection of the S45 residue implicates the presence of an allele specific phenotype, perhaps offering an advantage for this tumour type. Such selection over one specific residue, challenges the accepted mechanism, and raises the question as to whether or not there are phenotypic differences that these mutations confer.

A link between an optimal β-catenin activity level and cancer has been previously described in intestinal tumours with a "just right signaling model" (Albuquerque *et al.*, 2002). Furthermore, subsequent functional analysis of the *Apc* allelic series by both *in vitro* and *in vivo* approaches have provided evidence to support the claim that different β-catenin signaling levels are the driving force behind the observed cancer phenotypes (Kielman *et al.*, 2002; Gaspar *et al.*, 2009; Bakker *et al.*, 2013).

The evolution of cancers have long been known to follow the Darwinian principles of selection. According to this selection paradigm, mutations in cancers are selected for their ability to confer the optimal phenotypic advantage to tumour cells. The specific selection of the S45 mutation not only imply the presence of such selective force in Wilms' tumours, but suggests that this could also be observed in other tumour types with β-catenin mutations. The widely accepted current model of β-catenin regulation cannot explain how these mutations would lead to different phenotypes, and there must exist other, yet to be identified, players in this complex pathway. The weakness of this model

has already been exposed by Wang et al, using an S45-mutant colorectal cancer cell line to show that T41, S33 and S37 phosphorylation still occurs in the absence of phosphorylation at the S45 residue (Wang, Vogelstein and Kinzler, 2003).

The Wnt/β-catenin mediated pathway, in-spite of being one of the most extensively studied signal transduction cascades, still remains a Pandora's Box, and detailed characterization of β-catenin, the central switch of the pathway, is especially necessary. Over the years, *in vivo* and *in vitro* modelling of diseases have not only revealed the functional correlation of the genetic consequences, but have also provided deeper insights into the mechanism of some of the key regulatory aspects governing the signaling circuitry.

With this in mind, the main aim of my project was to develop an *in vitro* system to generate and analyse all of the mutations observed in the *CTNNB1* gene. The compelling evidence coming from cancer genetics, and from the literature, suggest an urgent need for this, and great advances in genome editing technology in recent times have provided an invaluable tool with which to dissect these mutations. Understanding the allele specific consequences of each of the different mutations observed in the *CTNNB1* gene across cancer types will provide a better perspective of the functional and regulatory aspects governing β-catenin activity in the context of Wnt signaling.

Chapter 2 Analysis of β-catenin mutational spectrum

2.1 Introduction

The revolutionary genetic and evolutionary principles of nineteenth century biologists Mendel and Darwin, remain the basis for our understanding of various biological processes, including the multistep process of tumourigenesis. The concepts and principles of these great visionaries, and their intricate observations, till date, remain to be of utmost importance. However, many observations and ideas of these visionaries, and their contemporaries remained descriptive, due to the lack of appropriate tools to prove their hypotheses. The technological advancements in the 20th century, especially in the field of molecular biology, following the seminal discovery of the double helical DNA structure by Watson and Crick, hugely complemented these theoretical principles, providing the tools to deconstruct the molecular framework and mechanistic details of the cellular processing unit.

The sequencing techniques, including the chemical sequencing introduced by Maxam and Gilbert and dideoxy/chain termination method by Fredrick Sanger, played a huge role in deciphering the genome at the single nucleotide level (Maxam and Gilbert, 1977; Sanger, Nicklen and Coulson, 1977). The availability of the entire genomic sequence for a number of organisms following the Human Genome Project, has provided the much needed reference sequence that greatly facilitated the detection of the precise genetic alterations underlying various anomalies, including cancer. The massively parallel and high throughput sequencing techniques have helped in identifying insertions, deletions, translocations, point mutations and other recently described phenomena such as chromothripsis and kataegis (Stephens *et al.*, 2011; Nik-Zainal *et al.*, 2012). Furthermore, sequencing techniques developed for uncovering the epigenetic and transcriptional regulation of gene expression, provide a newer dimension to our understanding of the cancer genome landscape (Meldrum, Doyle and Tothill, 2011) .

Taking leverage of these technological advancements, various large scale sequencing projects have been initiated to screen patient tumour samples from different cancer types, for the genetic, epigenetic and transcriptional variation. A number of completed and ongoing projects including the Cancer genome project (CGP) by Welcome trust Sanger Institute UK, The Cancer Genome Atlas (TCGA) a joint venture by the National Institutes of Health (NIH) and the National Cancer Institute (NCI), The International Cancer

Genome Consortium (ICGC), a voluntary and collaborative effort from various countries, have all generated massive amounts of data that have been made publicly available (Forbes *et al.*, 2009; Campo *et al.*, 2010; Weinstein *et al.*, 2013). In addition, various user friendly web based repositories including Catalogue of Somatic Mutations in Cancer (COSMIC), cBioPortal, Genomic Data Commons (GDC) dataportal and many such interactive platforms provide a comprehensive compilation from these large scale and various other smaller independent sequencing projects (Forbes *et al.*, 2008; Cerami *et al.*, 2014; Jensen *et al.*, 2017) .

With regard to β-catenin, various small scale independent sequencing projects have reported mutations in the exon 3 region of *CTNNB1* in a variety of solid tumours. As expected, the four phosphorylatable serine and threonine residues that are known to play a significant role in stabilizing the β-catenin protein were found to be the common target of mutagenesis in different cancer types. In addition, sequencing studies have also revealed mutations at various other residues in the exon 3 region (Polakis, 1999, 2000). However, there were no reports on the inclusive analysis of frequency and distribution of mutations in the β-catenin gene across various cancer types. Prior to analysis of the genotype phenotype correlation of β-catenin mutations, an overview of the mutational distribution across cancer types was required. The availability of large scale cancer genome data for various cancer types has made this investigation of mutational spectrum of β-catenin oncogene possible. Hence, this chapter will provide an overview of the mutational spectrum of β-catenin across cancer, compiled from the COSMIC database.

2.2 Results

2.2.1 Analysis of mutational spectrum of *CTNNB1* across cancer types

COSMIC is a database of somatic mutations frequently observed in human cancers that has been extracted from the literature, and from various other large scale cancer genome sequencing projects, including CGP and TCGA (Forbes *et al.*, 2010). Mutational data from cancer types, each having a minimum of 10 samples harbouring mutations in the *CTNNB1* gene was compiled from COSMIC.

The initial observation of the distribution of various types of mutation in *CTNNB1* from the COSMIC database showed an increased incidence of missense mutations. Although various mechanisms of genetic alteration exists, it has been documented that in a typical human tumour, around 90 percent of these mutations in known oncogenes are a result of missense changes (Vogelstein *et al.*, 2013).This was also observed for the mutation distribution of *CTNNB1* from the COSMIC database, with around 89 percent of them being missense mutation (Fig 2-1). Given that the majority of the mutations observed across the *CTNNB1* gene were single nucleotide changes, I considered to analyse only these mutations across various cancer types.

Mutation type	Mutant samples (Percentage)
Nonsense substitution	0.65
Missense substitution	88.8
Synonymous substitution	1.63
Inframe insertion	0.13
Frameshift insertion	0.03
Inframe deletion	5.98
Frameshift deletion	0.34
Complex mutation	0.47
Other	2.69

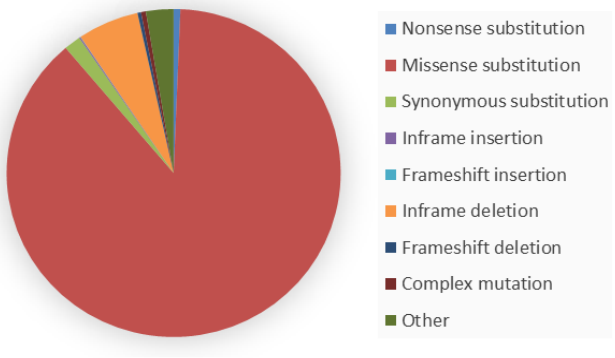


Figure 2-1: Distribution of various types of mutation in the *CTNNB1* gene. Table and pie chart showing the frequency of the different types of mutation observed in *CTNNB1* gene. Source: COSMIC database.

The initial tabulation and comparison of frequency distribution between cancers revealed mutations in a wide range of target residues across the *CTNNB1* gene. The serine and threonine residues which are phosphorylated by the destruction complex, were among the most frequently mutated residues, contributing to around 70 percent of mutations observed in the *CTNNB*1 gene (Fig 2-2). Of the 4560 *CTNNB1* mutations recorded across various tumour types from COSMIC database, the highest number of mutations were observed at the T41 residue. Mutations at the T41 residue was present in 1135 tumour samples, and this was closely followed by mutations at the S45 residue that was seen in 1092 tumour samples. Next, the two phosphorylatable serine residues S37 and S33, which are known to be the key sites for recognition by the ubiquitin ligase complex

occupy the third and fourth position, with 515 and 492 tumour samples bearing mutation at the S37 and S33 residues, respectively. The presence of mutations at the regulatable serine and threonine residues in over 2/3rds of the tumour samples, reflects the susceptibility of these residues to mutagenesis, and underlines the functional significance of phosphorylatable serine and threonine residues in maintaining the stability of the β-catenin protein.

Mutated Residue	mutant samples (Percentage)
T41	24.89
S45	23.94
S37	11.29
S33	10.78
D32	10.02
G34	6.95
I35	1.42
H36	1
P44	0.43
T40	0.35
K335	0.32
K49	0.3
A43	0.3
S47	0.28
G245	0.28
G48	0.28
Other	7.08

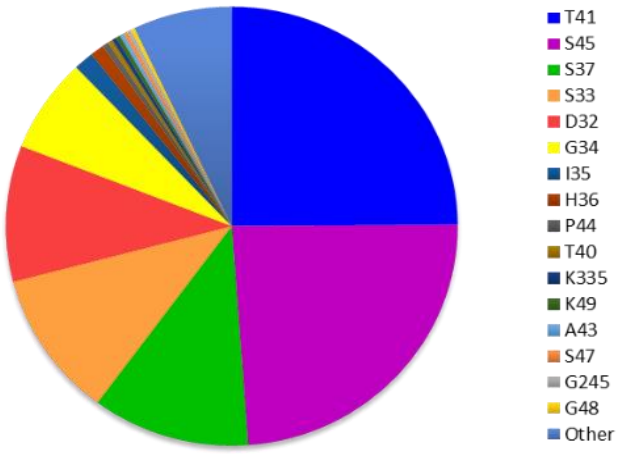


Figure 2-2: Distribution of mutation at various residues in across the *CTNNB1* gene. Table and pie chart showing the frequency of the different mutated residues in the *CTNNB1* gene compiled from COSMIC database.

However, the β-catenin mutations were not restricted to these sites. A large number of other residues adjacent to the serine and threonine residues, were also found to be commonly mutated. Particularly, the two residues D32 and G34, that are present immediately adjacent to the phosphorylatable S33 residue were found to be mutated at a relatively high frequency. The D32 mutations were seen in 452 tumour samples, and 317 tumour samples harboured a mutation at the G34 residue. Both D32 and G34 residues are known to be a part of the β-catenin degron motif, and the observed mutations might vary the specificities required for interaction with the E3 ubiquitin ligase complex. Furthermore, except for mutations at K335 and G245 residues, the rest of top

16 mutated residues were present between residues D32 and G50 in the exon 3 region. The residues I35 and H36 (that are also a part of the degron motif) were mutated in a considerable number of tumour samples. The mutations at residues P44, T40, K49, A43, S47 and G48 were observed in a variety of tumour types, but at a lower frequency. The presence of these very many allelic variants of β-catenin across various tumour types, observed due to mutations specifically confined to the region between L31 and G50, makes this region a mutational hotspot of β-catenin.

2.2.2 Analysis of the mutation pattern at specific residues across different tumour types

Analysis of the mutations at the hotspot region revealed a preferential selection of mutations at different residues among different cancer types (Fig 2-3). For example, the cancer types including the central nervous system (CNS) and pancreas are largely biased towards mutations at residues S37, S33, D32 and G34, contributing to an overall 97 and 90 percent of the mutations observed in these tissue types, respectively. Whereas, tumours of the soft tissue, specifically selected for mutations at the T41 and S45 residue. The mutations at the T41 and S45 residues contributed to 58 and 39 percent of the mutations observed in soft tissue tumours, together accounting for 97 percent of the mutations in these tumours. The tumours of the soft tissue constitute the highest frequency of β-catenin mutations, with 1325 mutated tumour samples, and the major bias towards selection of T41 and S45 tumour samples exemplifies the existence of a preferential selection of mutation among different tumour types. In addition, other examples worth mentioning are the tumours of the adrenal gland and kidney that specifically select for mutation at residue S45. This specific selection of a single residue attributing to around 79 and 77 percent of the observed mutation in the adrenal gland and kidney, indicate the existence of genotype-phenotype correlations among these β-catenin mutations.

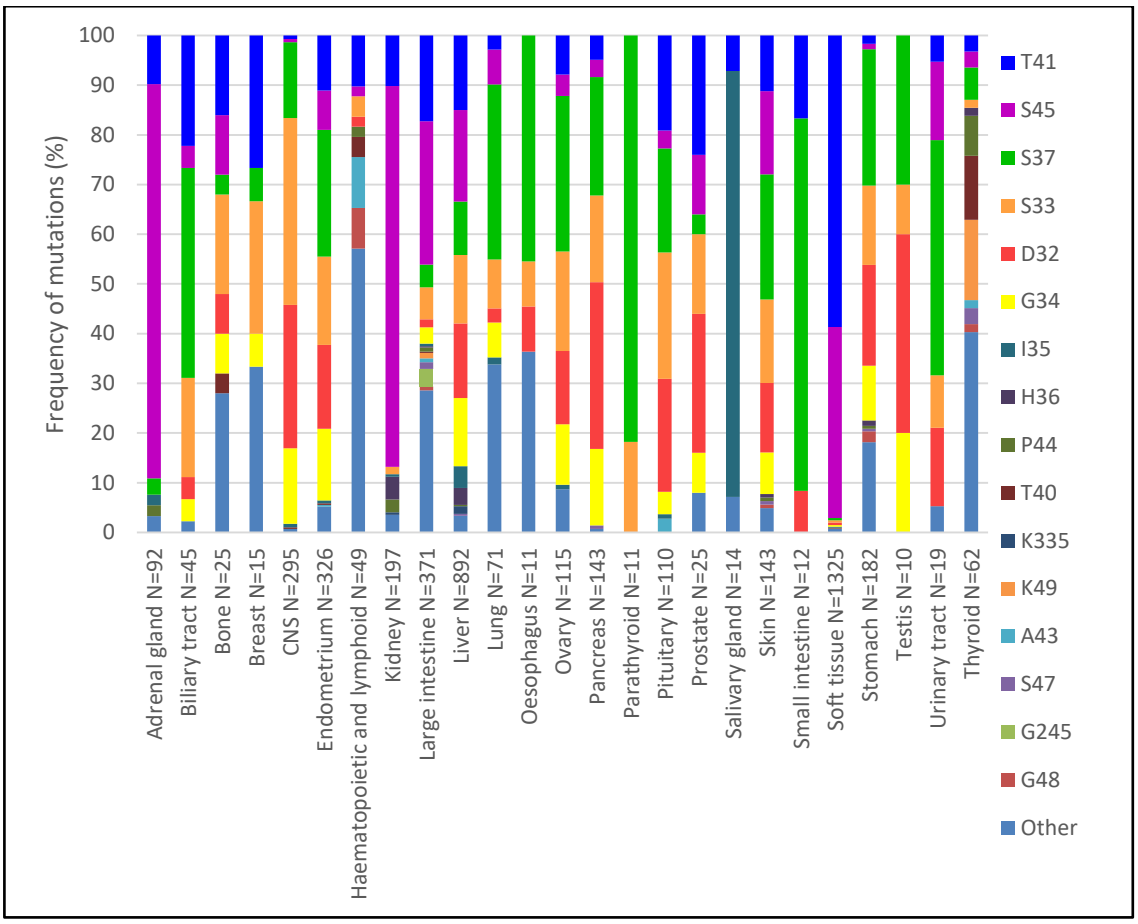


Figure 2-3: Graph representing the frequency distribution of the top 16+ other mutated residues across cancer types. The graphical representation of the *CTNNB1* mutation data compiled from COSMIC database showing a preferential selection of different residues among different cancer types.

2.2.3 Analysis of the mutational pattern of the amino acid variation across different cancer types

Over the years, mutation data for various genes have focused mainly on the 'residue' being mutated, and little or no emphasis is given to the 'amino acid variant' that the residue is being mutated to. The overexpression studies for functional analysis of the β-catenin mutants including S33, S37, T41 and S45 are mostly based on the analysis of one or very few amino acid variants of these residues, and all these mutations are often generalized to produce the active form of β-catenin (Morin *et al.*, 1997; Provost *et al.*, 2003; Baba *et al.*, 2006; Wege *et al.*, 2011).However, given the variation in the

biochemical properties of the 20 amino acids, and differences in their bonding specificities for inter and intra molecular interactions, each of these amino acids may vary in their functional attributes. This might result in preferential selection of specific amino acids at a particular residue among different tumour types. In addition to assessing the mutational pattern of the various residues across the exon 3 region, I decided to further analyse the amino acid variation at the individual residues across the different cancer types.

Assessment of the COSMIC database revealed that, indeed for a given residue, different tumours showed preferential selection for the amino acid substitution, adding an additional layer of complexity. For example, among the 20 available amino acid variants, mutations at the T41 residue were largely confined to substitution to alanine and to a lesser extent isoleucine. In addition to T41A and T41I substitutions, few tumours also selected for T41S, T41N and T41P substitutions. However, except for these five amino acid variants, none of the other 15 amino acids were selected for among the various tumor samples. Among these T41 missense variants, the soft tissue tumours (99 percent) specifically selected T41A substitution, whereas tumours of the pituitary gland (95 percent) largely selected for T41I substitutions (Fig 2-4). The mutations in the S45 residue in kidney cancers were substituted for by multiple amino acid variants including S45Y, S45P, S45F, S45C and also by S45 deletion, whereas the S45 mutation observed in soft tissue tumours specifically selected for S45F and S45P variants (Fig 2-5). This specific selection of different amino acid variants, imposes an additional criterion that must be taken into account when assessing the genotype phenotype correlation.

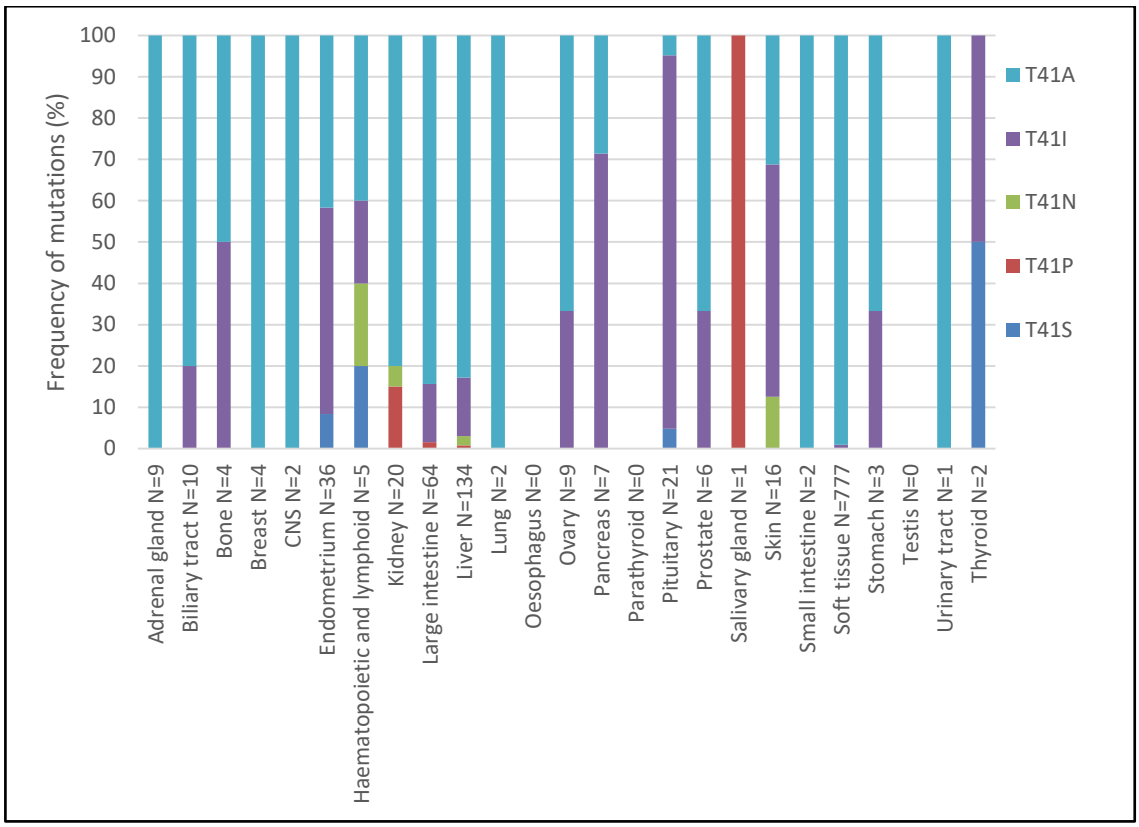


Figure 2-4: Graph representing the frequency distribution of different amino acid substitution across cancer types at the residue T41. The graphical representation of the *CTNNB1* mutation data compiled from COSMIC database showing a preferential selection of different amino acid substitution among different cancer types at the residue T41.

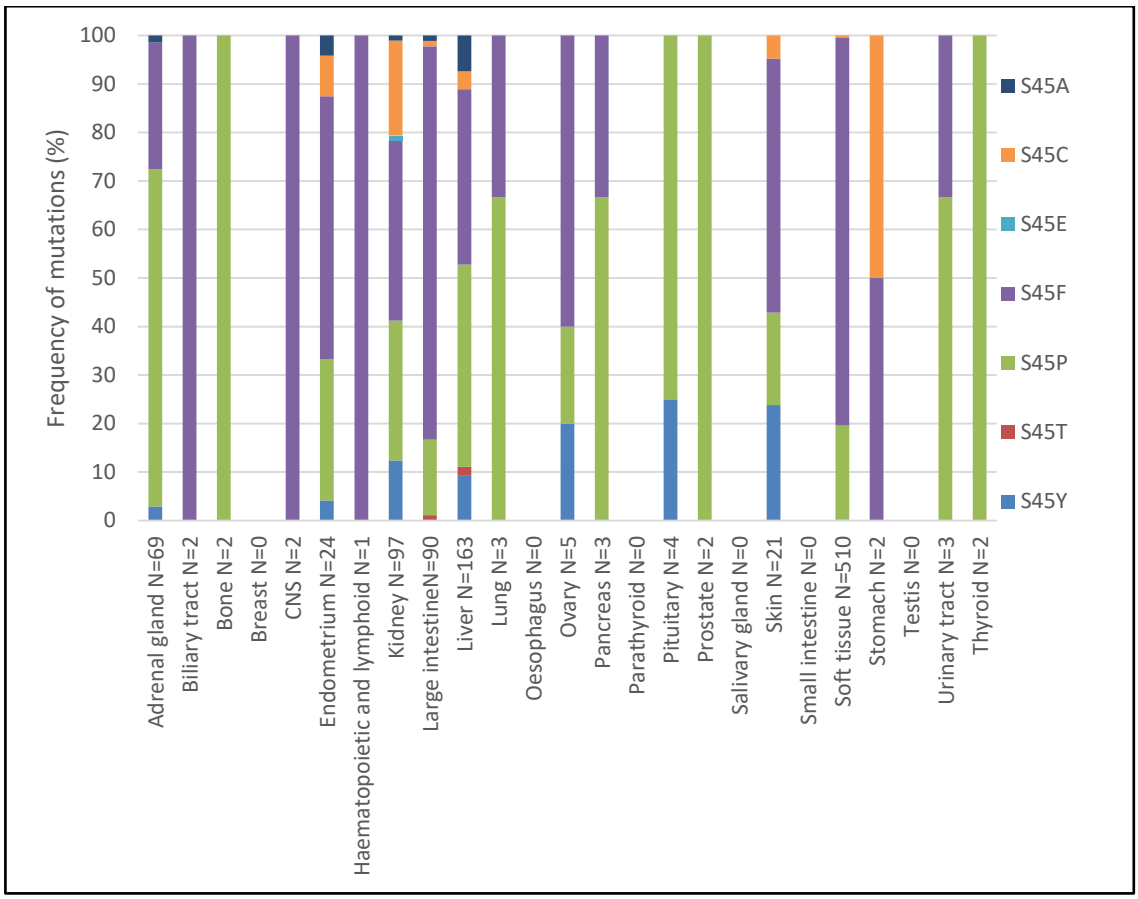


Figure 2-5 : Graph representing the frequency distribution of different amino acid substitution across cancer types at the residue S45. The graphical representation of the *CTNNB1* mutation data compiled from COSMIC database showing a preferential selection of different amino acid substitution among different cancer types at the residue S45.

2.2.4 Assessment of the statistical significance of the observed β-catenin mutations across different tumour types.

Before commencing analysis of the biological significance of these mutations, it was important to determine if indeed these differences were statistically significant. For the purpose of further analysis, I mainly questioned: a) Which of the residues would be good candidates for further investigation, b) Which of the amino acid substitutions relating to the selected residues would be good candidates for further investigation. Statistical assessment was done by Dr. Helen Brown (senior statistician, The Roslin Institute). Statistical analysis was done assuming a null hypothesis of no differences in the expected proportion of mutations across different cancer types to test; a) Whether or not there are significant differences in the proportion of mutation (of a specific residue / amino acid substitution) across all the tumour types, and b) Which sites have an overall significant difference in mutation (of a specific residue / amino acid substitution) between tumour types.

The top six mutations with an increased frequency among cancer types (T41, S45, S37, D32, S33 and G34) also differed in statistically significant ($p<0.05$) proportions across all cancer types (Appendix 1A). Also a significant difference was observed in the spectrum of mutations between tumour types (Appendix 1C). For example, tumour types such as CNS, pancreas, stomach and endometrium, described earlier were significantly enriched for mutations in D32, G34, S33 and S37 residues, whereas kidney, soft tissue and Large intestine showed a significant preferential selection for mutation at the S45 residue. Mutations at the T41 residue were significantly higher in soft tissue tumours. In addition, mutations at various other residues in the hot spot region including I35, H36, T40, A43, P44, G48 and K49, although observed in fewer tumour types, were found in significantly enriched proportions in the tumours in which they were present.

Furthermore, analysis of the top six frequently mutated residues revealed statistically significant differences in the amino acid substitutions among cancer types (Appendix 1B). The T41 missense variants that were observed at differing frequencies, also differed in statistically significant proportions. The pituitary tumours were significantly enriched for T41I substitution, whereas the tumours of the soft tissue presented a significant enrichment for T41A substitution (Appendix 1D). Another mutation noteworthy of mention

are those observed in tumours of the salivary gland, the majority of the mutations observed in these tumors are T>C transition mutations at the I35 residue, resulting in a specific I35T substitution, that were again present in statistically significant proportions. The Liver tumours are another tumour type in which the I35 mutations are statistically significant. However, the I35 mutations in the liver largely selected for I35S substitution and not a single I35T variant was observed.

This preferential selection of different residues and different amino acid substitutions among cancer types, implicate the probable existence of a fundamental difference between these mutations, thus validating the importance of understanding the genotype phenotype correlation among β-catenin mutations.

2.3 Discussion

Oncogenic conversion and deviation from the normal state requires the continuous acquisition of various genetic and epigenetic alterations, making the cancer genome landscape an evolving hub of mutations capable of providing selective advantages at multiple stages of the tumourigenic process. The cellular proto-oncogene required for regulated maintenance of the 'master clock of the cellular circuitry' - The cells intrinsic growth control machinery, is one of the highly susceptible targets for gaining the characteristic cellular advantages attributed to tumours. In the 1970s, viral integration was thought to be the major cause of oncogenic activation, however, since then a myriad of genetic mechanisms for activation of an oncogene have been described, and can be mainly classified into two major groups based on the elicitation of changes in either expression levels (regulatory) or structure of an oncoprotein (Pierotti et al. 2003). The quantitative changes resulting from multiple mechanisms including translocations and viral integration of oncogenes, placing them under ubiquitously expressing promoters, or by gene amplification, lead to constitutive gene expression. Furthermore, genetic alterations including insertions, deletions and missense mutations, can lead to subtle changes in the protein structure, and in addition large inversions and translocations are capable of producing a hybrid protein with altered functional role (Botezatu, 2016). Besides these genetic changes, epigenetic mechanisms including changes in DNA methylation pattern are capable of producing an altered state of oncogene expression (Cheung *et al.*, 2009).

Detailed analysis of each of these genetic aberrations is of absolute importance not only in uncovering the mechanistic significance of the cellular transforming events, but also for targeted therapeutic interventions, and great strides in this direction have been made in the recent years. The advancements in sequencing platforms have greatly complemented the cytogenetic studies in underpinning the nuances of the mutational landscapes not limited by the scale of analysis, and account for the availability of tumour data at an unprecedented rate.

Mutations in the proto-oncogene *CTNNB1,* which encodes for β-catenin have been reported in various tumours, however, there is a lack of comprehensive analysis of the β-catenin mutational pattern in different cancer types. Taking advantage of the available tumour data from large scale sequencing projects, I selected one such web based repository, the COSMIC database, to analyse the mutational spectrum of *CTNNB1* across various cancer types.

Non-random distribution of gain of function mutations localized to a particular domain referred to as mutational hotspots, exist in *RAS*, *PIK3CA*, *IDH1* and many other well characterized oncogenes, leading to changes in the canonical confirmation of the protein and activation (Miller *et al.*, 2015; Baeissa *et al.*, 2017). Small scale sequencing studies of human tumours have reported the exon 3 region of *CTNNB1* to be the mutational hotspot of β-catenin (Polakis, 1999, 2000). Our mutation data of the *CTNNB1* gene across various cancer types revealed a similar pattern, with the majority of mutations localized near the regulatory residues between L31 and G50. The phosphorylatable sites were among the most frequently mutated residues, however, mutations were not restricted to the regulatory serine and threonine residues but were also observed at the surrounding residues making the entire stretch of the exon 3 region a susceptible target for mutagenesis. The frequency of mutation was lower at the residues surrounding the regulatable serine and threonine sites. Although observed at a lower incidence, these mutations may have a crucial role in the tumours in which they are present, and hence assessing the phenotypic consequences of each of the allelic variants is necessary.

Until recently, there were no reports on the detailed analysis of the β-catenin mutational spectrum in different types of cancers. However, during the course of my PhD, Çelen et al. published a similar analysis of the β-catenin mutational pattern of the top six residues

across different cancer types from COSMIC database (Çelen *et al.*, 2015). By performing hierarchical clustering, they were able to group the cancer types into two main clusters; cluster1 included tumour types that were biased towards mutation at residues D32, S33 and S37, and with very low frequency of T41 and S45 mutations, and cluster 2 was biased towards mutation at T41 and S45 residues. A similar clustering or preferential selection of specific residues among different tumour types was observed in our study as well. These results suggest that the specific selection mutation confined to one or few residues, are a common feature in several cancer types harbouring β-catenin mutations. However, based on the current dogma of β-catenin activity, and considering the equal probability of occurrence of every observed mutation among different tissues, one would hypothesize a random distribution of mutations at the phosphorylatable serine and threonine residues across tumour types. This specific selection of mutations among different types of cancers, might imply an existence of fundamental difference between these mutations.

Furthermore, as pointed out earlier, little emphasis is placed upon the analysis and functional importance of the different amino acid variants, and the majority of the phenotypic analysis of the mutant alleles are based on a single amino acid variant. The study by Çelen et al, included a brief account on the amino variants observed at the residues I35 and H36. However, the detailed analysis of amino acid variants across the frequently mutated residues was not performed. Our analysis of the COSMIC database, not only includes the mutational spectrum of different cancer types, but also provides additional information on the preferential selection of different amino acid substitutions, again indicating a selection bias based on functional significance governed by Darwinian principles.

In addition to the Darwinian principle of selection of cancer specific mutations based on functional significance, various evidence indicate the presence of differing mutational processes contributing to the background mutational rate specific for each of the cancer types, to be another important factor that may lead to observed tissue specific mutational bias. The differences in the background mutational rate among different cancer types are attributed to the differences in sensitivity to endogenous and exogenous agents, and mutation inducing mechanisms operative in different tissue types (Pfeifer, 2010;

Alexandrov *et al.*, 2013). In this chapter, a basic approach considering no mutational bias and assuming equal substitution rates among different cancer types was used while calculating the statistical significance of the tissue specific differences in the observed spectrum of mutations. However, owing to the importance of the tissue specific mutational pattern in determining the overall mutational rates, a detailed bioinformatics analysis was later undertaken inclusive of the background substitution rates and will be discussed in chapter 4.

It remains to be understood whether the mutations seen in β-catenin proto-oncogene are selected for either based on functional significance, or by the existence of different mutation inducing mechanisms in different tissue types, or a combination of both these processes. However, it is important to understand the functional consequence of these observed mutations, with regard to its role in the tumorigenic process. Moreover, the large scale mutational data is of little significance, unless complemented with experimental approaches to uncover the underlying genotype-phenotype correlations, which presently remain unestablished for majority of the genes.

Chapter 3 Optimization of strategies and tools for generation of heterozygous endogenous β-catenin mutants using CRISPR/Cas9 technique.

3.1 Introduction

The detailed analysis of the COSMIC database revealed a preferential selection of β-catenin mutations among different cancer types, suggesting a possible fundamental difference between these mutations. In order to investigate allele specific consequences of the various mutated residues in the β-catenin hotspot region (L31-G50 which cover 93 percent of the total number of missense mutations observed across the different types of cancers), and to explore the functional significance of the individual residues across the region of interest, I decided to adopt two complementary approaches.

In the first approach, I wanted to address our research question by hypothesizing that the differences in signaling activity can be a selective pressure that could lead to the observed variation in mutations. Recently, it has been shown that “saturation editing” combined with CRISPR/Cas9 provides a prospective way to analyse the functional significance of every nucleotide/amino acid residue within a short stretch of the genome (Findlay *et al.*, 2014). Our analysis of *CTNNB1* COSMIC data showed that the majority of the mutations are clustered around the conserved phosphorylatable serine and threonine residues, making this region a suitable candidate to perform saturation editing. Introducing all possible codon substitutions surrounding the phosphorylatable residues L31-G50 in a pool of mES cells, which carry a β-catenin activity reporter, allowed us to quantify the corresponding β-catenin activity for each substitution at single cell resolution. This unbiased approach also allowed us to compare the cancerous and non-cancerous mutations in the same region for their effect on β-catenin activity.

Secondly, I generated mutant mES cell lines using CRISPR/Cas9 technology with “multiplex targeting”, and analysed them *in vitro* for β-catenin activity and the expression of the known target genes. For this, I selected the top six statistically significant mutations with all the significant amino acid substitution (4-5 amino acid substitution for each residue) from the COSMIC database analysis. By analyzing the most frequently mutated residues in stable cell lines, I tried to understand the reason for this preferential selection and its significance in different tumour types.

These two approaches adopted to address our research question, multiplex targeting (in E14 mES cell line) and saturation editing (in TCF/Lef:H2B-GFP β-catenin activity reporter

mES cell line), were both challenging tasks and required a strategic and robust experimental set up. Both approaches required a similar system for the generation of endogenous mutations, and many strategies were common for both. Hence, in this chapter, a detailed view of the various optimization strategies adopted will be provided, highlighting their importance in either multiplex targeting, saturation editing, or both.

3.2 Results

3.2.1 mCherry pX458 CRISPR nuclease vector construction and validation of expression

Given the large scale nature of our experimental approach, with the requirement of the generation of multiple variants, this project would not be feasible using conventional targeting approaches. Therefore, I decided to use the CRISPR/Cas9 system due to its simplicity in adapting for the target of interest, and increased editing efficiency with the aid of endonucleases. Among the various different systems available, I choose to use the plasmid based system that contains two main components; the Cas9 and the sgRNA with 20bp user specific sequence to define the genomic target. A universal CRISPR nuclease plasmid (pSpCas9(BB)-2A-GFP- also called pX458) was generated in Feng Zhang's lab that contains all the required components; Cas9, an oligo cloning site to insert the target specific 20 bp sequence, and the remainder of the sgRNA scaffold. This vector also has a GFP selection cassette, which allows selecting the transfected cells, increasing the targeting efficiency (Ran *et al.*, 2013), and I expected to need such enrichment to reach the required efficiency for my project. As my reporter cell line was also based on GFP expression, for quantification of β-catenin activity in the saturation editing approach, it would not be possible to distinguish CRISPR transfected cells from the cells with β-catenin activity. Therefore, to be able to sort CRISPR transfected cells, I replaced the GFP cassette in pSpCas9(BB)-2A-GFP with mCherry, using Gibson assembly. This cloning system was developed by Dr. Daniel Gibson, and it is based on having overlapping DNA fragments with an enzyme mix containing an exonuclease, DNA polymerase and DNA ligase in a single isothermal step (Gibson *et al.*, 2009). Using this method, the PCR amplified mCherry fragment (amplified using primers having overlap with the vector) was cloned into the CRISPR nuclease (pSpCas9(BB)-2A-(pX458)), that had been initially digested with EcoRI to remove the GFP cassette (Fig 3-1).

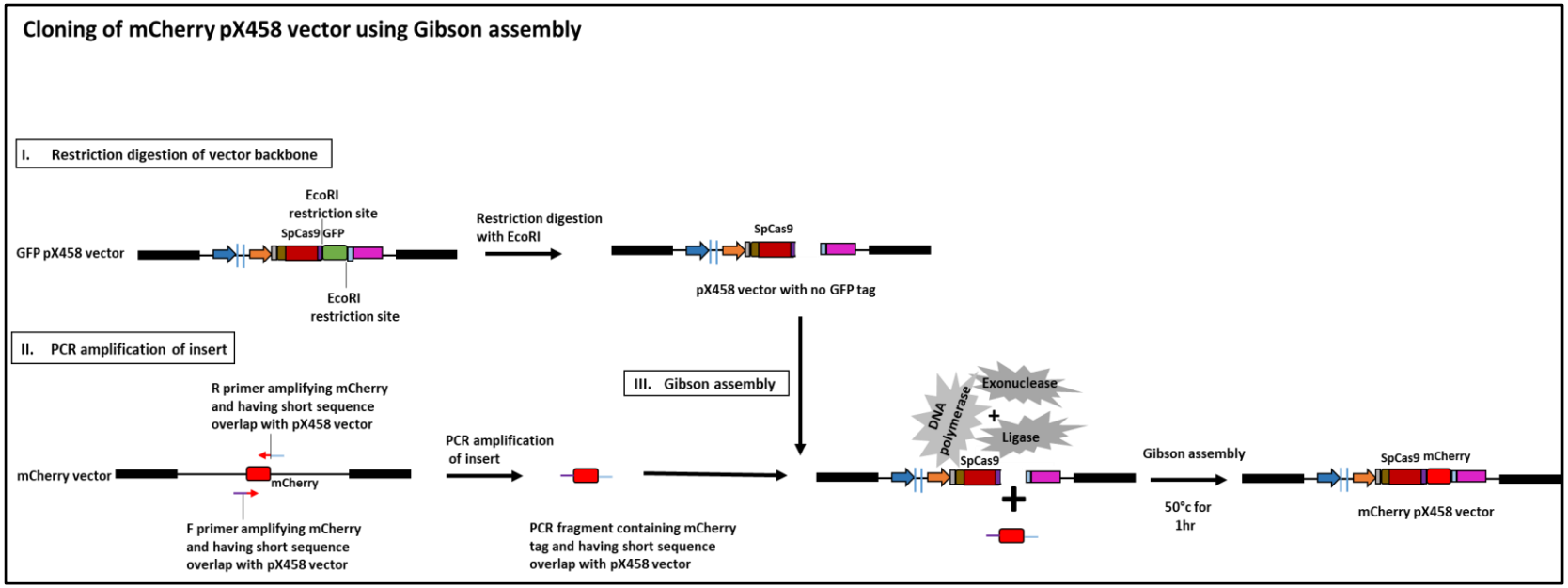


Figure 3-1: Cloning of mCherry pX458 vector using Gibson assembly. The GFP pX458 vector was initially digested with EcoRI to remove the GFP tag. The mCherry tag was amplified using primers having sequence overlap with the pX458 vector. The overlapping backbone and insert fragments were cloned by Gibson assembly.

Colony PCR using primers specific for mCherry was performed to select for positive clones (Fig 3-2). Next, sequence verification revealed the correct insertion of the mCherry fragment in the vector.

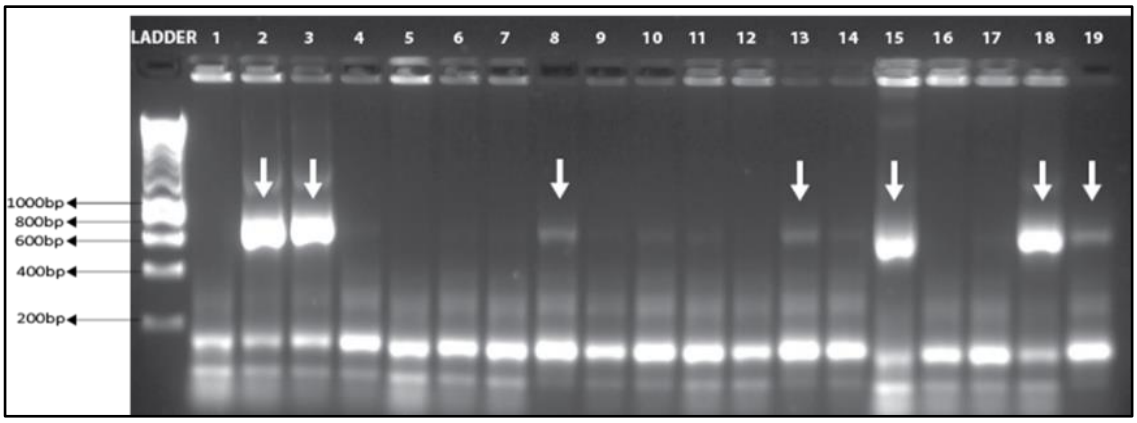


Figure 3-2: Colony PCR of mCherry pX458 vector. Agarose gel electrophoresis image of the Gibson assembly transformants screened by performing colony PCR using mCherry specific primers. PCR positive clones of expected band size (809bp) are indicated by arrows. Hyperladder 100bp.

To validate the expression of the mCherry reporter, and also to compare its expression to that of the existing GFP CRISPR vector, both vectors were transfected independently into E14 cells, and analysed by flow cytometry (Fig 3-3). Wild type E14 cells were used as a control to gate the GFP/mCherry negative population. Firstly, to gate the viable population from the dead cells, a dot plot of Forward (FSC) vs side scattering (SSC) which distinguishes cells based on the size (FSC) and the granularity or cellular complexity (SSC) was plotted. The P1 gate represents the viable population of E14 cells (3-1A and 3-1E). The P1 gate was applied to the rest of the dot plots for analysis of viable cells alone. The P3 gate in the FITC-A vs SSC-A dot plot consists of GFP positive population which was ~20 percent of the total P1 (viable population) in the GFP pX458 transfected cells, validating the expression of GFP reporter, and was negligible in the mCherry pX458 transfected cells as expected (Fig 3-3B and C). The P2 gate consist of mCherry population which was 17 percent of the total P1 population in the mCherry transfected cells, and was negligible in the GFP pX458 transfected cells (Fig 3-3F and G). Flow cytometric analysis thus confirmed the expression of mCherry, and also the transfection efficiency was comparable to that of the GFP vector.

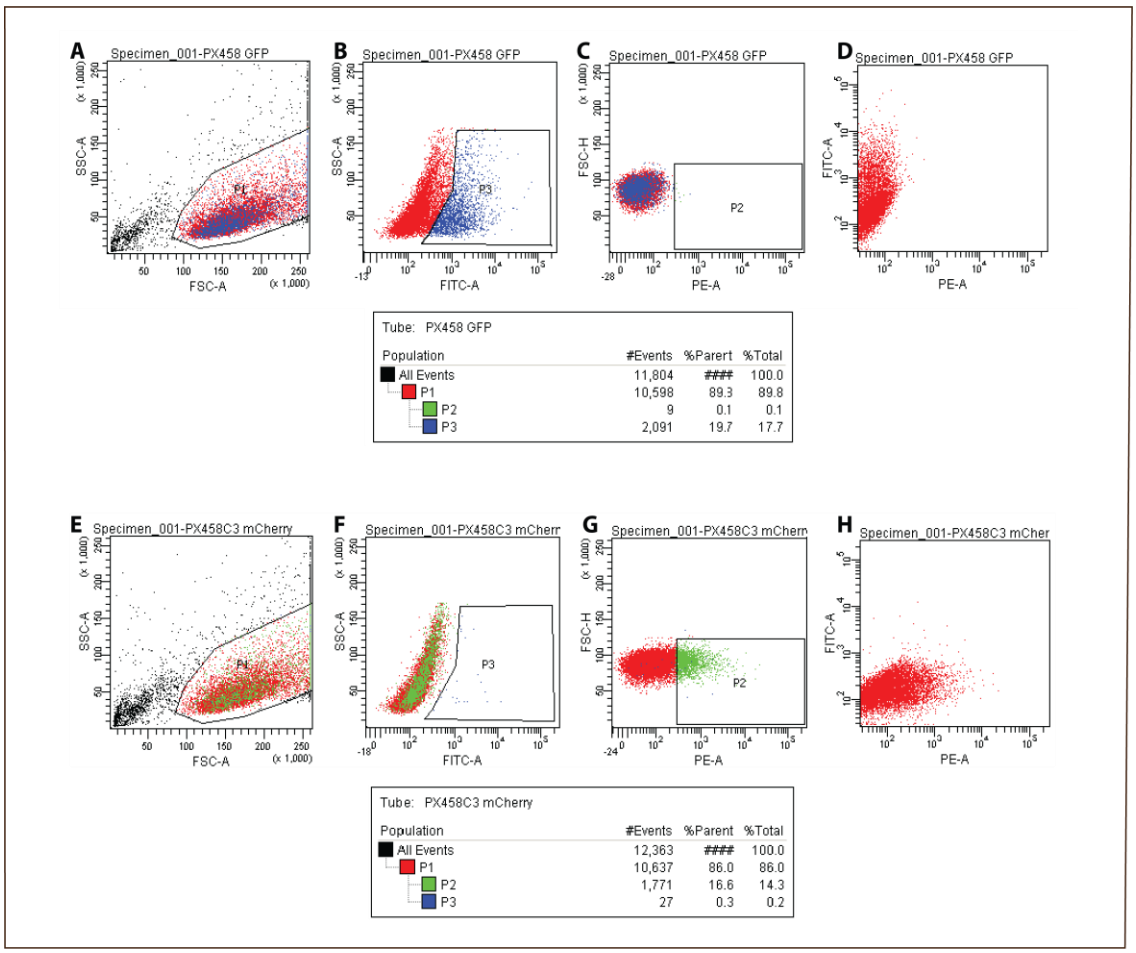


Figure 3-3: Validation of GFP and mCherry reporter expression in GFP and mCherry pX458, respectively. Flow cytometric analysis of the expression and transfection efficiency of GFP and mCherry pX458.

3.2.2 Design and assembly of various guides targeting the exon 3 region of β-catenin in GFP px458 and mCherry Px458

Now that I had the mCherry pX458 vector, the next step was to design CRISPR guide sequences. For the purpose of saturation editing, a single guide that cuts in the centre of this region was preferred. However, as every guide may vary in their efficiency to induce DSB, six guides targeting this region covering the *Ctnnb1* mutational hotspot (residue L31-G50) were designed to test and select the optimal guide (Fig 3-4). The guides were designed using the online CRISPR design tool provided by Zhang lab (http://crispr.mit.edu/). All six guides were assembled in both mCherry pX458 and GFP

pX458 using the protocol published by Ran et al. (Ran *et al.*, 2013). Briefly, a 20 bp guide RNA for the target region was ordered as oligos and ligated in to the backbone vector in a single reaction of digestion-ligation using BbsI restriction enzyme and quick ligase. The reaction was then transformed into Stbl3 competent cells and 3 colonies were randomly selected for verification for correct insertion by sequencing.

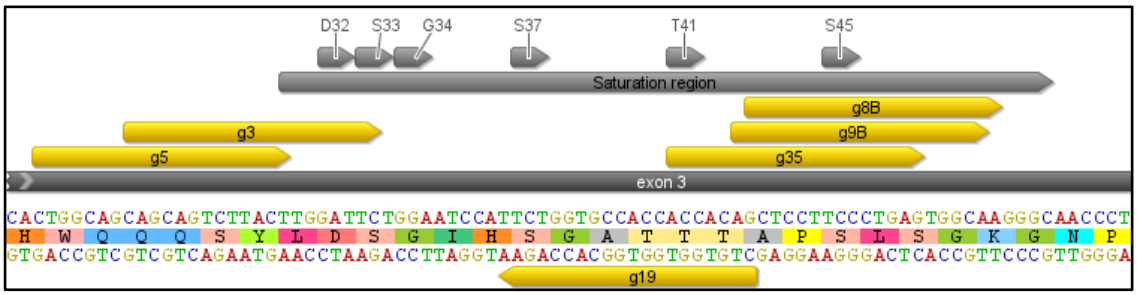


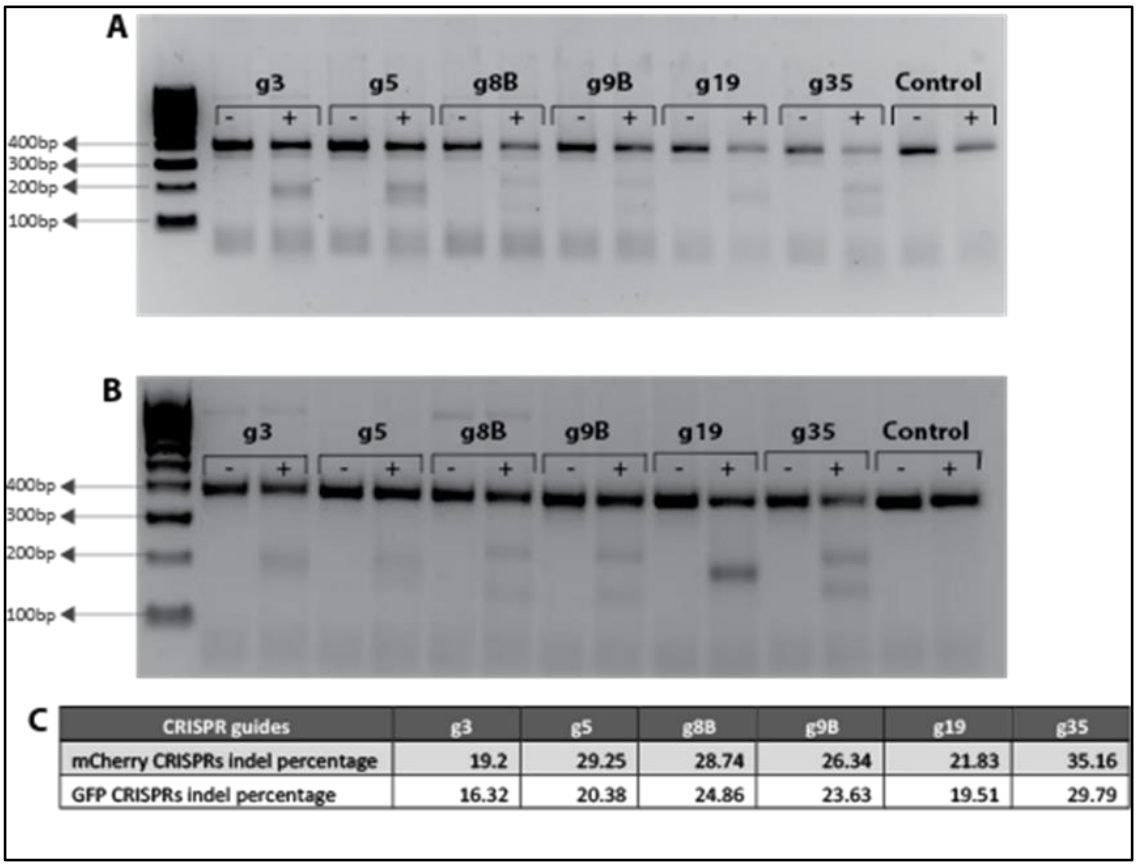
Figure 3-4: Guides targeting exon 3 region of β-catenin. Sequence view of a part of exon 3 region of β-catenin with the six selected guides that were assembled into mCherry pX458 and GFP pX458.

3.2.3 Comparison of the editing efficiency (Indel frequency) of the mCherry vs GFP PX458 guides

The efficiency of the 6 guides assembled in both mCherry pX458 and GFP pX458 to induce double strand breaks was analysed by performing a T7 endonuclease I (T7E1) assay. T7E1 specifically recognizes and cleaves mismatches in heteroduplex DNA, and hence is used to identify the presence of mutations. The presence of mutations in the CRISPR transfected cells is in turn indicative of the CRISPR editing efficiency, and T7 assay is routinely used as a preliminary screening strategy for the detection of CRISPR editing efficiency. For identifying mutations by T7 assay, initially the PCR product covering the region of interest is denatured and reannealed to allow heteroduplex formation in the presence of mismatches. On digestion with T7EI that recognizes and cleaves the mismatches, presence of mismatches are visible as extra bands on agarose gel.

For the purpose of analysing the editing efficiency of the designed CRISPRs, all the guides assembled in the GFP pX458 and mCherry pX458 were transfected into E14 cells and FACS sorted 24 hours post transfection. As this is a transient transfection, I had

previously tried sorting of transfected cells at various time points and the reporter signal was lost over the time course, and I found that 24 hours post transfection results in optimal reporter signal in our experimental system (data not shown).The isolated genomic DNA from the sorted population was then used to amplify a 400bp region of exon 3. Next, the PCR product was denatured and reannealed to allow heteroduplex formation, followed by digestion with T7EI. The presence of mismatches were visible as extra bands on agarose gel (Fig 3-5A and B). The intensity of the bands were quantified using ImageJ software, and the indel percentage was calculated using the ratio of the cleaved bands to the total PCR product (Fig 3-5C). The T7 endonuclease results were comparable in both GFP pX458 and mCherry pX458, with slightly better editing efficiency in mCherry pX458 vectors. Among the six tested guides it was not possible to modify the PAM of guides g5, g35 and g8B. Hence, I selected the remaining three guides (g5, g19 and g9B) to test further for their HDR efficiency. The guide g19 was especially a good candidate for being in the middle of the region to be edited. In addition to the three guides tested here, I also included another guide g6 (2) that had been previously designed and tested in the lab for my further analysis.



C

CRISPR guides	g3	g5	g8B	g9B	g19	g35
mCherry CRISPRs indel percentage	19.2	29.25	28.74	26.34	21.83	35.16
GFP CRISPRs indel percentage	16.32	20.38	24.86	23.63	19.51	29.79

Figure 3-5: T7 endonuclease I assay. (A) Agarose gel of T7 endonuclease I assay of guides in mCherry pX458. (B) Agarose gel of T7 endonuclease I assay of guides in GFP pX458. (C) Tabular column showing the indel percentage of mCherry and GFP pX458. The two lanes for each of the 6 guides (g3, g5, g8B, g9B, g19, g35) represents undigested (-) and T7 endonuclease I digested (+) PCR products. The CRISPR pX458 (either mCherry or GFP, respectively) backbone vector only transfected cells (without the guides inserted) were used as controls (c) for T7 assay.

3.2.4 Generation of PAM mutant cell line

Studies have suggested β-catenin mutations to be activating, and having a dominant effect. The mutation in a single allele is sufficient for the manifestation of the tumourigenic potential of this oncogene. Hence, studying the heterozygous condition would be more physiologically relevant, therefore I decided to generate heterozygous mutants; i.e. with one WT β-catenin allele and one allele harbouring the desired mutation. The initial strategy to achieve this, was to generate a cell line with a heterozygous synonymous mutation in the PAM region of the selected guide. The PAM sequence is necessary for

the recognition and subsequent Cas9 mediated cleavage of the target site, and an intact PAM in the targeted clones will result in repeated re-cutting and may lead to additional indels, and these NHEJ events would affect the β-catenin activity of that cell. If we could generate an allele that is not responsive to the editing guide, this would ensure that in every cell the activity is derived from one WT and one mutant allele, making the comparison between different cells more reliable. Hence, I initially strategized to generate a TCF/Lef:H2B-GFP clone with a heterozygous synonymous mutation in the PAM region.

In this approach, an ssODN repair template with synonymous mutations in the PAM region for four of the selected guides (g5, g19, g6 (2) and g9B) was designed (referred as 4mutPAM oligo) (Fig 3-6A). However, as the TCF reporter cells were still being derived from the TCF/Lef:H2B-GFP reporter mice in the lab of Kat Hadjantonakis (Memorial Sloan-Kettering Cancer Center), I tested the targeting efficiency of the 4mutPAM in E14 cells along with g19 CRISPR that induces a DSB at the centre of the region of interest. The reason for mutating the PAM sites of all these 4 guides was to allow flexibility in the future as to which guides I can use in these cells. Although I had tested the editing efficiency of these guides, I did not know which one would be best for HDR efficiency. We had already known from our concurrent work that although 2 guides may have similar editing efficiency, the HDR efficiency can vary greatly. Furthermore, trying to introduce 4 mutations at the same time, using a guide in the middle, would also give us information about the positional effect of the edit to the guides.

146 clones were tested for correct integration by sequencing. The results showed that the editing efficiency of g19 CRISPR was 70 percent with 102 clones repaired by NHEJ events. The majority of the indels were very close to the cutting site. A single insertion of T nucleotide immediately next to the g19 cutting site was very common, especially among the homozygous NHEJ events, and 7 out of 15 clones had this T insertion. The remaining homozygous NHEJ events were all deletion of varying lengths. Various different indel events were observed among the heterozygous clones repaired by NHEJ, all close to the CRISPR cutting site. Although a very good rate of editing was observed, the targeting resulted in a very poor frequency of HDR events (Fig 3-6B). Only a single clone was homozygous for all the four PAMs and with a HDR of 0.7 percent, and I could not identify

any clones targeted correctly in the heterozygous state with a clean WT allele at any of the PAM sites I attempted to mutate.

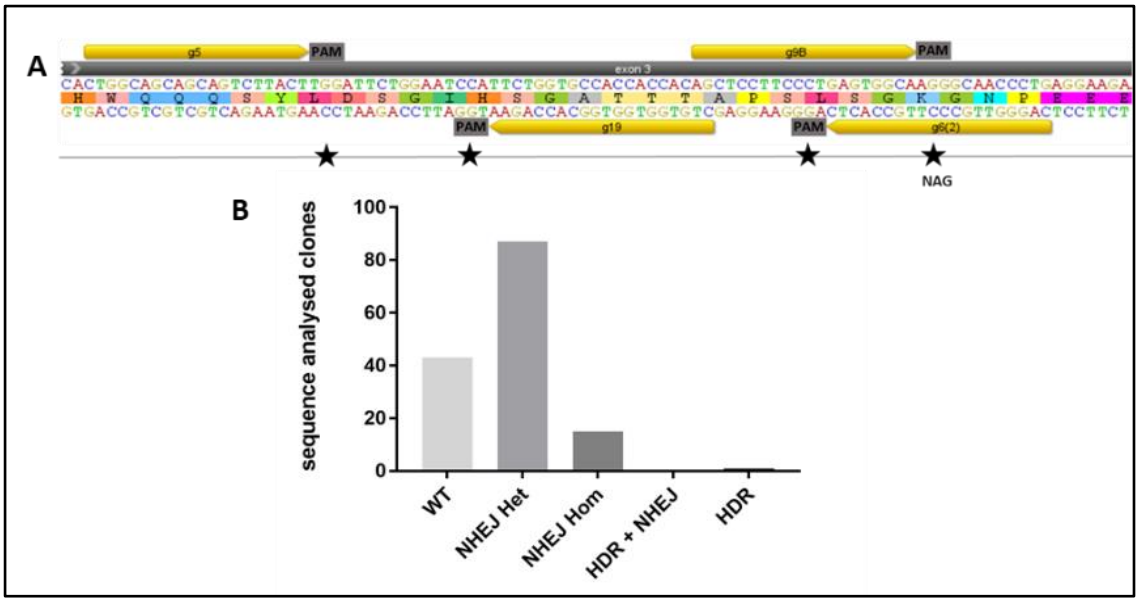


Figure 3-6: ES cell targeting with ssODN repair template using nucleofection method. The initial targeting performed using an ssODN (4mutPAM oligo) and g19 CRISPR vector by nucleofection method: (A) ssODN repair template with synonymous mutations at PAM sites of 4 guides. (B) Graph representing the number of sequence analysed Wild type (WT) and mutant (NHEJ Hom, NHEJ Het, HDR+NHEJ and HDR only) clones.

3.2.5 Optimisation of HDR efficiency

3.2.5.1 Small molecule enhancer

The initial targeting experiment in E14 cells using g19 CRISPR and 4mutPAM oligo using nucleofection resulted in a very low percentage of HDR. Both our experimental approaches, generating 5-6 different mutants in a single step via multiplex targeting and targeting the 20 residue region for saturation editing, required a much higher efficiency than this. Hence, I turned my focus on how to further optimize the HDR efficiency. It has been shown that use of small molecule compounds had a potential to increase HDR mediated DSB repair mechanism, also with reduced cellular toxicity. One such compound SCR7, which is an inhibitor of DNA ligase IV - a key enzyme of canonical NHEJ pathway, was shown to promote the frequency of HDR events in both *in-vivo* and *in-vitro* systems

(Chu *et al.*, 2015; Maruyama *et al.*, 2015). To test this drug, the 4mutPAM oligo (designed for saturation editing experiment), an S33Y oligo and ΔS45 oligo were used in combination with either g19 or g9B CRISPR in various attempts.

Targeting of E14 cells was repeated with an ssODN with S33Y point mutation and silent PAM mutation (NGG-NGT) (Fig 3-7A), using the nucleofection method, and the cells were incubated in 1μM SCR7 containing media for 24hours followed by FACS sorting. The analysis of sequencing data for S33Y targeting resulted in a HDR efficiency of 1.2 percent, which was a two fold increase from the previous targeting experiment (Fig 3-7B).

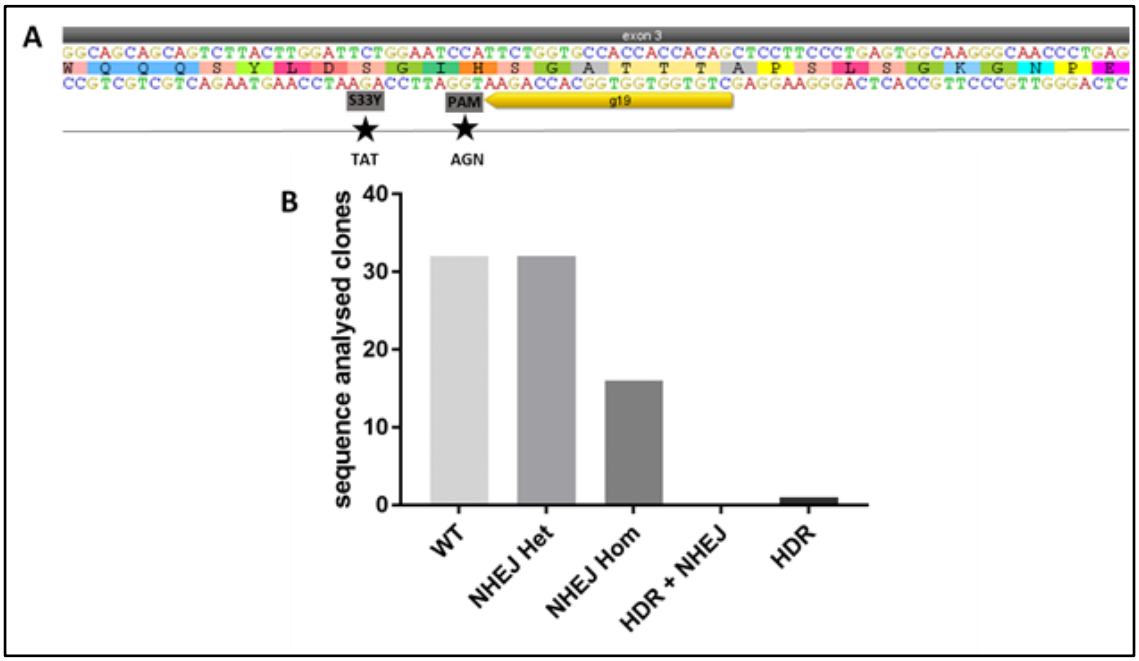


Figure 3-7: E14 targeting using DNA ligase IV inhibitor SCR7. The use of SCR7 in clones targeted by nucleofection method. Fig (A) ssODN repair template with S33Y mutation and synonymous mutations at PAM (NGG-NGA). (B) Graph representing the number of sequence analysed Wild type (WT) and mutant (NHEJ Hom, NHEJ Het, HDR+NHEJ and HDR only) clones.

In addition to SCR7, another small molecule compound, L755507 was identified in a reporter based screening method as having an ability to promote HDR events upon induction of DSB by CRISPR/CAS9 system (Yu *et al.*, 2015). To test if this compound could promote HDR at our targeting site, again 4mut PAM oligo (designed for saturation

editing experiment) along with g19 CRISPR was nucleofected and treated with 5 µM L755507 for 24 hours followed by FACS sorting. However, the analysis of sequencing results of the targeted clones showed no improvement in the HDR percentage. Only a single clone out of 146 clones analysed had a mutation in g19 PAM with 0.7 percent HDR efficiency, similar to the targeting observed with nucleofection alone in the absence of the compound (Fig 3-8).

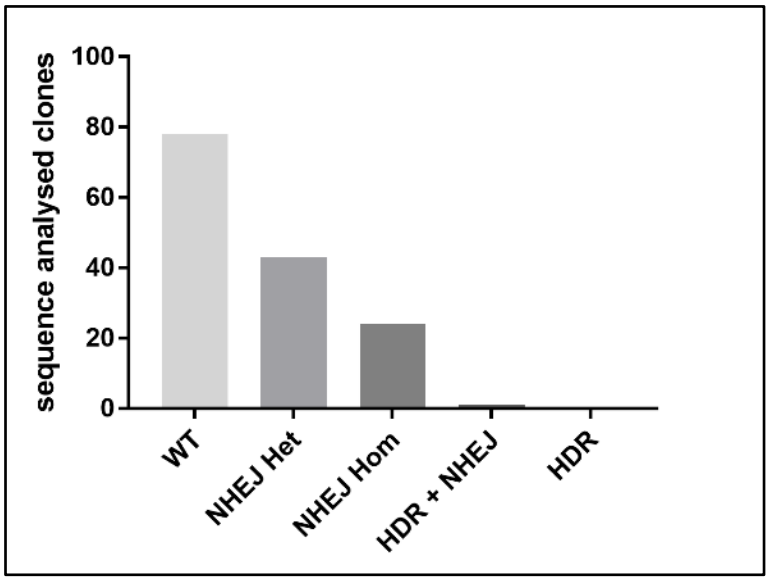


Figure 3-8: E14 targeting using small molecule compound L755507. The use of L755507 in E14 cells targeted by nucleofection method. Graph representing the number of sequence analysed Wild type (WT) and mutant (NHEJ Hom, NHEJ Het, HDR+NHEJ and HDR only) clones.

Even with the use of small molecule compounds, the HDR events at our target site was considerably low, and hence it was necessary to attempt different approaches of optimization. In the next targeting experiment, a different transfection technique was used. Instead of nucleofection, E14s were transfected using lipofectamine with suspension cells. The targeting was performed using S33Y oligo and g19 CRISPR along with small molecule compound L755507. The analysis of sequencing data revealed a drastic improvement in the HDR frequency. The percentage of total HDR events was 15.6, and except for one clone, all the clones with insertion of PAM mutation also had the S33Y mutation. However, the majority of these clones had indels on the other allele and

only 2 HDR clones (1S33Y het with PAM het and 1 S33Y hom with PAM hom) out of 115 (1.7 percent) had no additional NHEJ events (Fig 3-9).

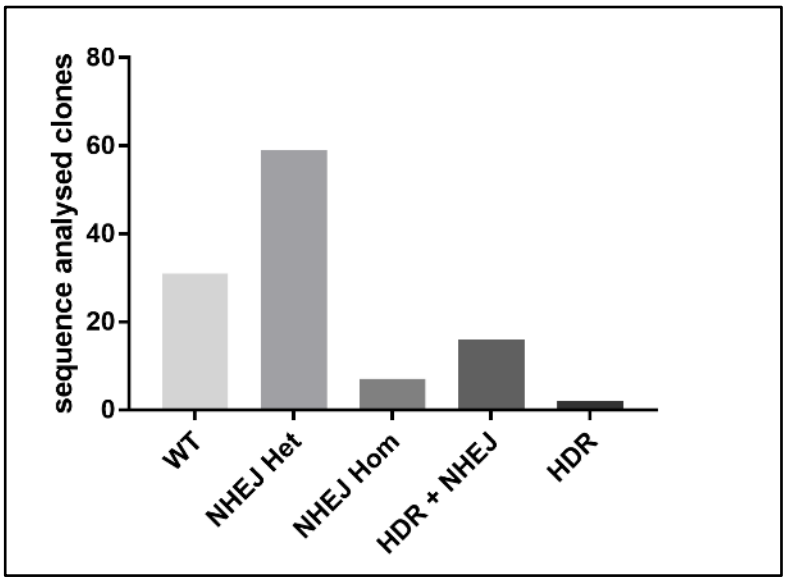


Figure 3-9: E14 targeting by lipofection in combination with small molecule compound L755507. The use of L755507 in cells targeted by lipofection method. Graph representing the number of sequence analysed Wild type (WT) and mutant (NHEJ Hom, NHEJ Het, HDR+NHEJ and HDR only) clones.

3.2.5.2 E14 targeting using ssODN template with additional mutation in the 'seed sequence'

As mentioned previously, the introduction of a synonymous mutation in the PAM sequence greatly reduces the chance of re-cutting by Cas9. However, for certain guides, introduction of a synonymous mutation in the NGG PAM is not possible, or can only be converted to NAG PAM in the targeting template. It has been documented that in addition to NGG PAM, SpCas9 has the ability to cleave NAG PAM, albeit with 1/5th the efficiency as compared to NGG PAM, thus resulting in additional indels observed along with HDR events (Hsu *et al.*, 2013). To test whether an additional mutation in the proximal region of the CRISPR binding site reduces the re-cutting of NAG PAM, an ssODN was designed with silent mutation in the 19bp of the CRISPR binding site along with PAM mutation in g9B CRISPR and S45 deletion (Fig 3-10A). To test this, E14 cells were transfected with the ssODN and g9B CRISPR using lipofection and L755507 treatment. Sequencing

analysis revealed 28 percent of clones with HDR and out of the 13 clones that were repaired by HDR, 7 clones were homozygous for all the three insertions and without any NHEJ events (Fig 3-10B).

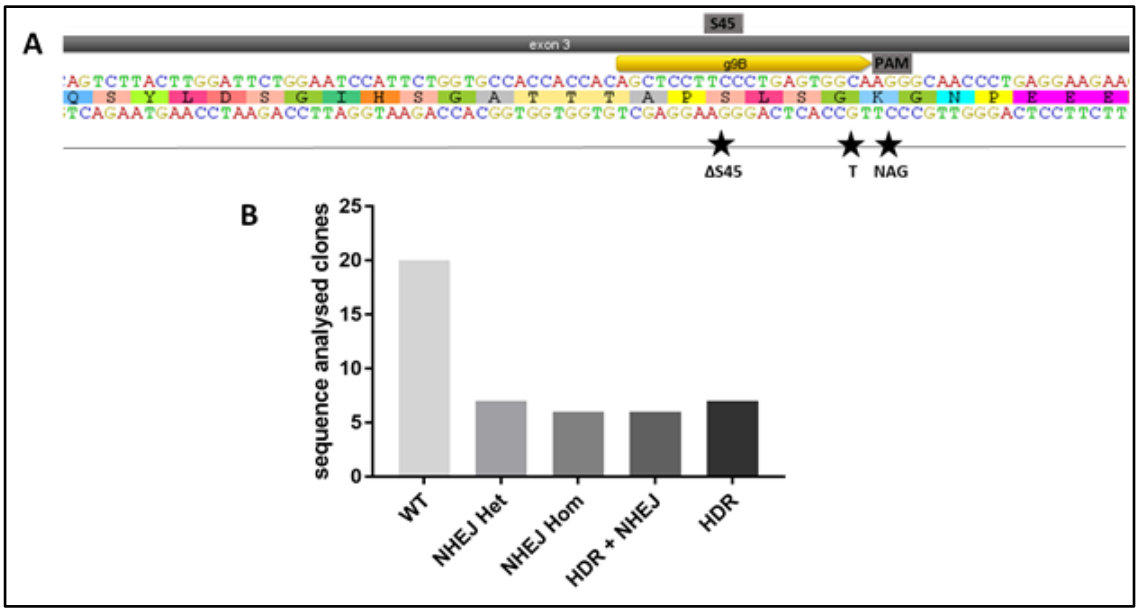
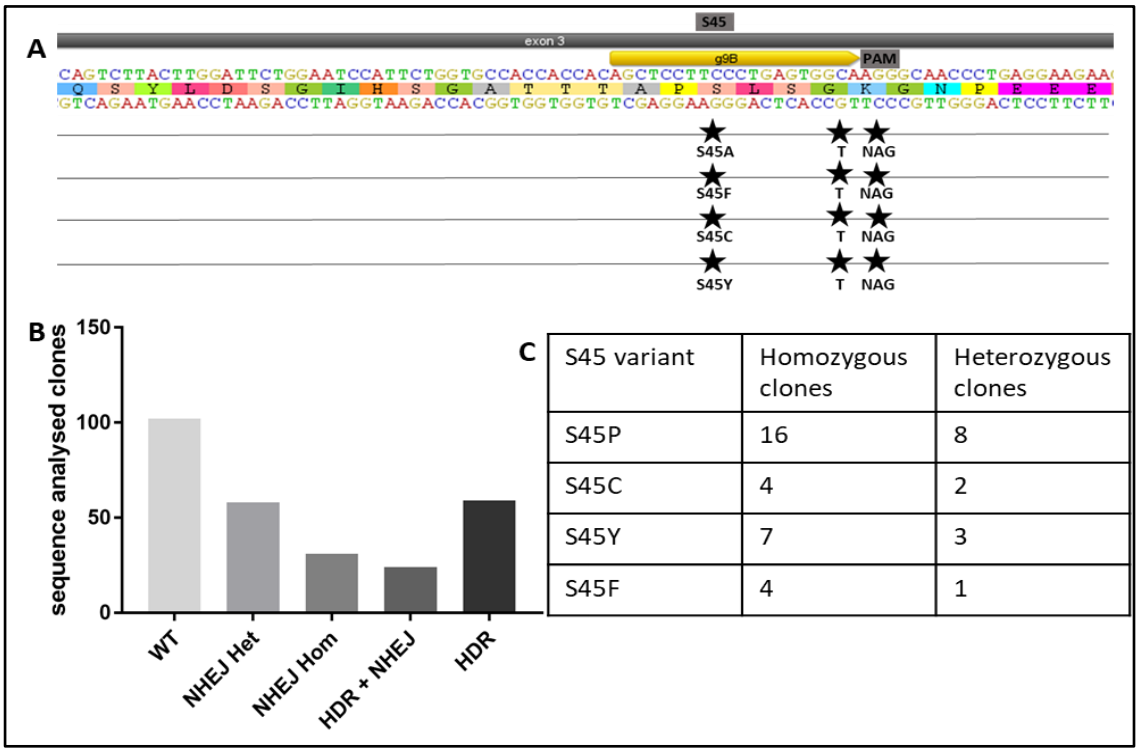


Figure 3-10: E14 targeting by using a repair template with additional mutation in the seed sequence. The incorporation of an additional mutation in the 19th bp of CRISPR binding site to reduce re-cutting of the NAG PAM. (A) ssODN repair template with S45 deletion, synonymous mutation in the PAM (NGG-NAG) and an additional synonymous mutation in the 19th bp of CRISPR binding site. (B) Graph representing the number of sequence analysed Wild type (WT) and mutant (NHEJ Hom, NHEJ Het, HDR+NHEJ and HDR only clones.

3.2.5.3 Multiplex targeting using ssODN as repair template

Applying the above strategies, ssODNs with a synonymous mutation in the PAM region and an additional synonymous mutation in the 19th bp of g9B CRISPR binding site and the specific mutation (S45 C P Y F) in S45 residue were designed to test how efficient multiplex targeting was (Fig 3-11A). The S45 ssODN variants were multiplexed along with g9B CRISPR, and transfected into E14 using lipofectamine and L755507, in a single transfection. Sequencing analysis revealed an overall HDR percentage of 30, in this first attempt of multiplex targeting. However, among the HDR only clones, the frequency of homozygous mutants were relatively higher in comparison to the heterozygous mutants (Fig 3-11C). Nevertheless, the initial round of multiplex targeting using ssODN was very

efficient and we were able to generate all the required S45 mutants. Interestingly among the analysed clones there were no compound heterozygous mutant clones which might be due to the mitotic recombination events triggered by DSB induction, wherein a DSB in the second allele may induce reciprocal crossover with its homologous chromosomes that has already been repaired by HDR using the exogenously introduced repair template.



S45 variant	Homozygous clones	Heterozygous clones
S45P	16	8
S45C	4	2
S45Y	7	3
S45F	4	1

Figure 3-11: S45 Multiplex targeting using ssODN as repair template. Multiplex targeting of S45 performed in E14 with g9B CRISPR (A) ssODN repair templates of S45 variants with synonymous mutation in the PAM (NGG-NAG) and an additional synonymous mutation in the 19th bp of CRISPR binding site: (B) Graph representing the number of sequence analysed Wild type (WT) and mutant (NHEJ Hom, NHEJ Het, HDR+NHEJ and HDR only clones. (C) Table representing the number of homozygous and heterozygous clones acquired for each of the four S45 variants.

Although the initial multiplex targeting of residue S45 using g9B CRISPR that induces a DSB 7bp away was successful, the distance between the cutting site and mutation might have a significant effect. Based on all our targeting experiments, and testing various

guides, we found only two efficient CRISPRs (g9B and g19) in the region of interest that were also at a reasonable distance from the desired mutations. The other guides including g5 and g6 (2) when used for targeting, had a very low editing efficiency (data not shown). Hence, for the purpose of both multiplex targeting and saturation editing, it was important to test if we could use these two CRISPRs to generate all mutations in the hotspot, using ssODN as template. To test this, we next performed multiplex targeting of residue T41 using g19 CRISPR that cuts 10bp away from residue T41 (Fig 3-12A).

E14 cells were transfected using lipofectamine and L755507 along with g19 CRISPR and ssODN variants harbouring mutations at T41, and a synonymous substitution at g19 PAM. Sequence analysis revealed a reasonable HDR efficiency (17 percent). However, out of the 12 clones repaired by HDR, only two clones (<3 percent) harboured mutations at both T41 and the PAM region, and the remaining 10 clones had incorporated the PAM mutation but not the T41 mutation (Fig 3-12B). This indicated that the distance between the cut-site had a significant effect on the introduction of mutations when using ssODN as repair template, and this was a major drawback for both our experimental approaches.

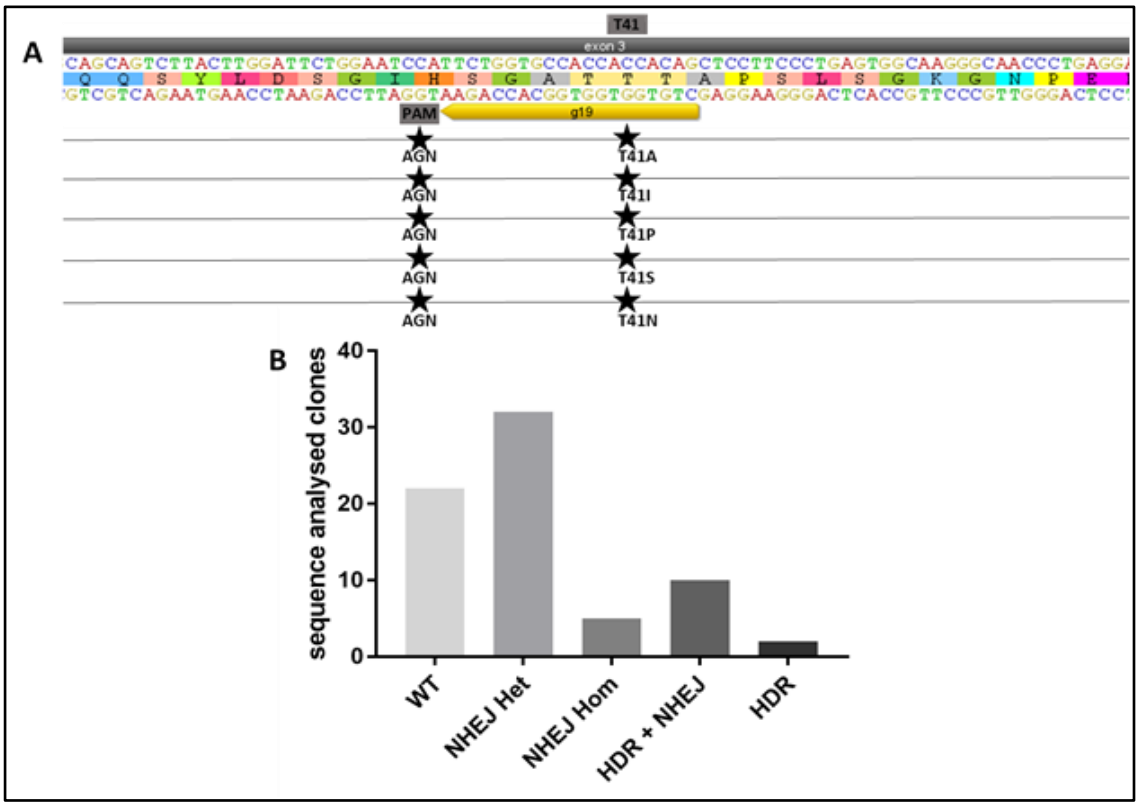


Figure 3-12: T41 multiplex using ssODN. Multiplex targeting of T41 performed in E14 with g19 CRISPR (A) ssODN repair templates of T41 variants with synonymous mutation in the PAM (NGG-NGA) (B) Graph representing the number of sequence analysed Wild type (WT) and mutant (NHEJ Hom, NHEJ Het, HDR+NHEJ and HDR only clones.

3.2.6 E14 targeting using vector (TV) with 1Kb homology arms

Through the targeting's I performed and the other work done in the lab, it became clear that, although using ssODN as a template for HDR was efficient if the mutation is very close to the cutting site, the efficiency decreased greatly after approximately 8 bp distance between the mutation and the cutting site. As I was trying to target a considerably larger area than this, I decided to test using a vector with long homology arms as an HDR template. Previously, various CRISPR mediated HDR based targeted gene editing studies have reported efficient targeting using vectors with homology arms 500-1000 nucleotide in length (Merkle *et al.*, 2015; Ratz *et al.*, 2015). To test if we can edit with good efficiency using vectors, a targeting vector with 1KB homology arm was constructed.

3.2.6.1 Cloning of 1Kb Homology arm TV

To test the efficiency of targeting with a vector as an HDR template in the most time-efficient way, we changed our focus to the S45 deletion. The reason for this was that, from a different project we already had an E14 cell line with a homozygous S45 deletion. These cells were also inserted with a V5 tag immediately after the start codon in the exon 2 region in the endogenous *Ctnnb1* gene on both alleles. The V5 tag and S45 deletion were introduced using CRISPR/Cas9 mediated gene editing. Briefly, E14 cells were transfected using two CRISPRs scg3 (that cuts in exon2 region) and g9B (that cuts in the exon3 region) along with HDR vector template with V5 tag and S45 deletion. Using the homozygous clone obtained from this targeting, we could amplify both 5' and 3' arms, and the mutation in one fragment (2.5kb in total), and clone it directly to a vector backbone. The genomic DNA isolated from these S45 mutant cells were used as a template to amplify the region covering 1kb homology arm on the 5' side and 3' end of the region of interest. The amplified PCR product was cloned into TOPO 4 vector (Fig 3-13). The transformed clones were screened by colony PCR (Fig 3-12B), few PCR positive clones were sequenced and one correct vector was selected (to be used as HDR template) for targeting (Fig 3-14).

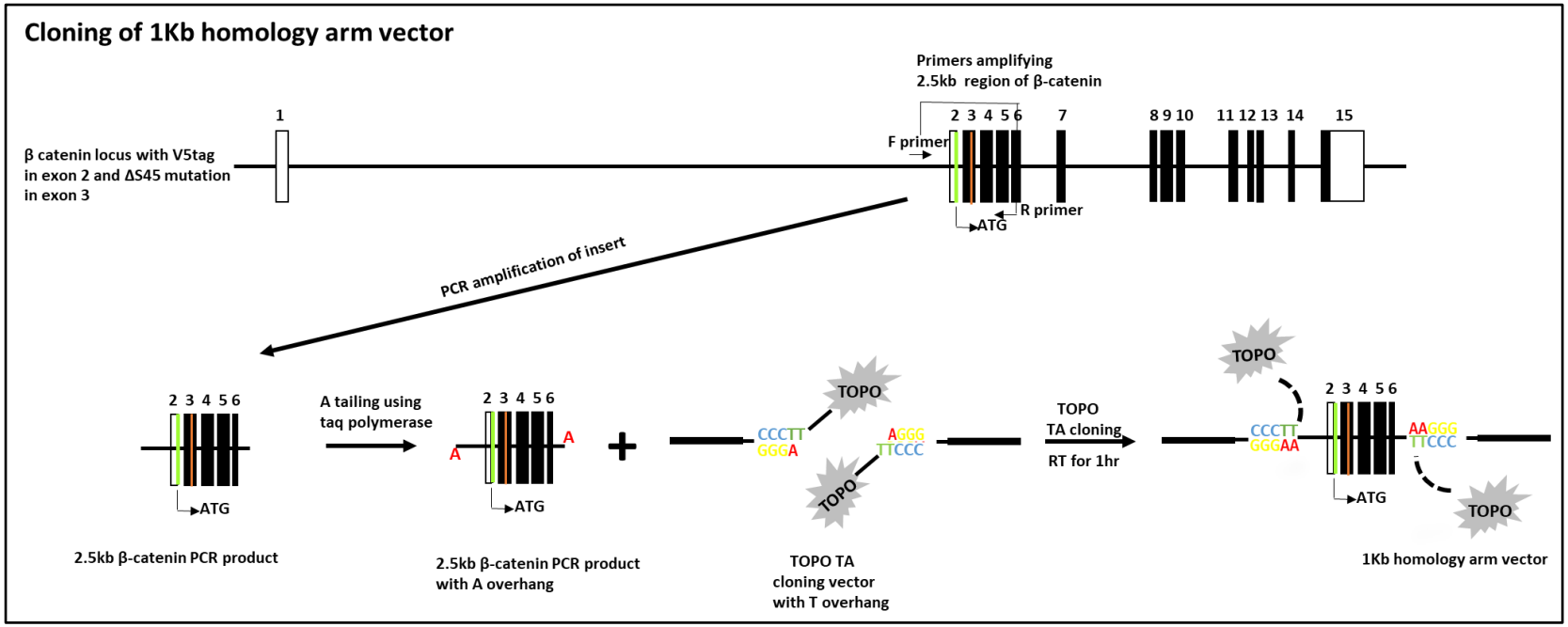


Figure 3-13: Schematic representation of cloning of 1Kb homology arm TV. The 2.5kb β-catenin region consisting of V5 tag immediately after the start codon in the exon 2 and ΔS45 mutation in exon 3 was amplified using genomic DNA from V5 and ΔS45 homozygous mutant cells. Following amplification, A overhangs were added to the 3'end of the PCR product using taq polymerase and subsequently cloned into TOPO TA vector.

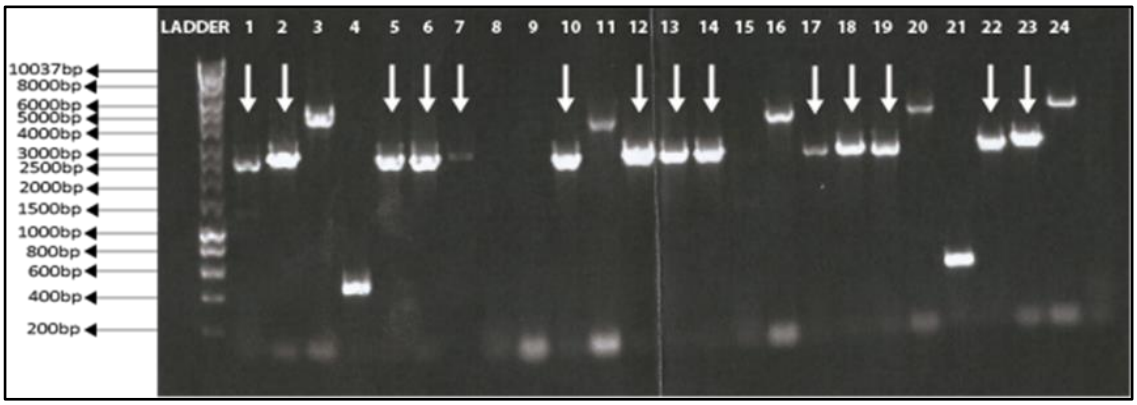


Figure 3-14: Colony PCR of 1Kb homology arm TV. Agarose gel electrophoresis image of the TOPO cloning transformants screened by performing colony PCR using β-catenin specific primers (1Kb F/R). PCR positive clones of expected size (2453bp) are indicated by arrows. Hyperladder 1Kb.

3.2.6.2 Targeting using 1Kb Homology arm vector

E14 cells were transfected with the targeting vector along with two CRISPRS; scg3 CRISPR (that cuts immediately next to the start codon) and g6 CRISPR (cuts 19bp downstream of S45 residue) (Fig 3-15). Using two CRISPRs, the entire region between start codon and exon 3 could be deleted, and HDR mediated repair using the 1Kb homology arm vector would allow insertion of both V5 tag and S45 mutation.

The sequencing data revealed HDR events in 12 clones out of 89 (13.5 percent) clones analysed (Fig 3-16). The insertion of the V5 tag abolishes the binding site for the scg3 guide, therefore protecting it against re-cutting and NHEJ. The PAM region (NGG) for g6 could only be changed to NAG which still had some cutting ability, albeit less efficient. All the clones with g6 PAM mutation inserted were also either homozygous or heterozygous for S45 deletion, but with additional NHEJ in most of the clones, which shows the importance of the correct PAM mutation to prevent further editing.

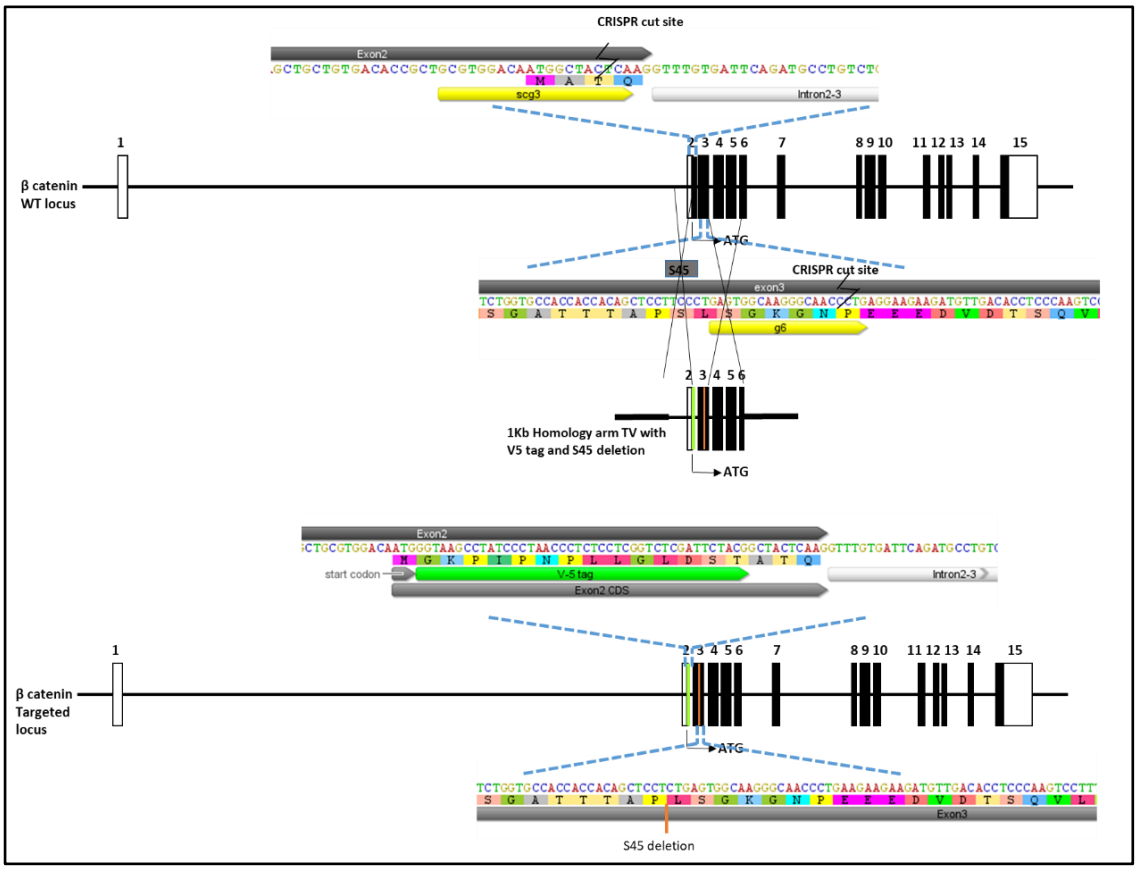


Figure 3-15: Strategy used for 1Kb homology arm vector targeting. Schematic representation of β-catenin WT locus, 1Kb homology arm vector and β-catenin targeted locus. The scg3 CRISPR (near start codon) and g6 CRISPR (downstream of S45) were used to delete the region of interest. The 1Kb homology arm vector with V5 tag and S45 deletion was used as template for HDR.

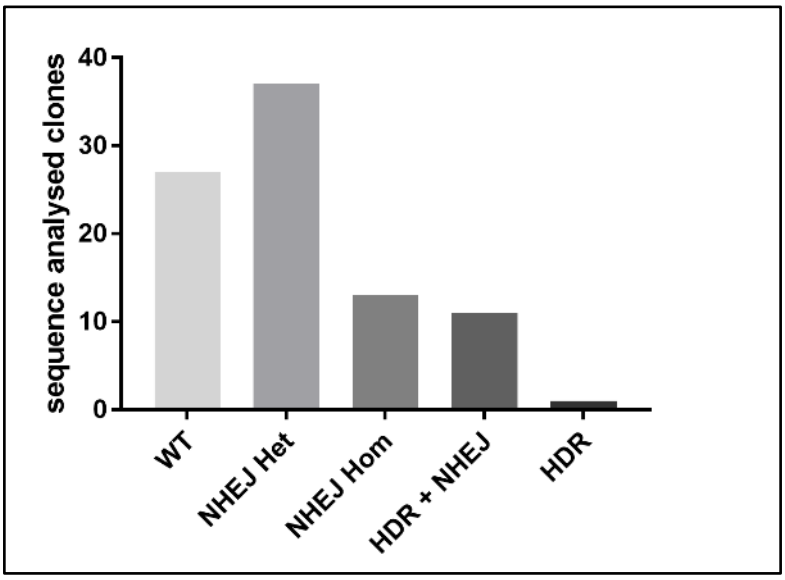


Figure 3-16: E14 targeting using 1Kb homology arm vector. The use of 1Kb homology arm targeting vector to compare the efficiency vs ssODN. Graph representing the number sequence analysed Wild type (WT) and mutant (NHEJ Hom, NHEJ Het, HDR+NHEJ and HDR only) clones.

3.2.7 Generation of β-catenin KO cell line with puDeltatk selection cassette

Our repeated attempts of generating a heterozygous PAM mutant cell line remained unsuccessful. It was very important to generate clean heterozygous mutants, but the inability to generate a PAM heterozygous mutant cell line, and the caveats of using ssODNs as HDR templates for introducing mutations throughout the 20 amino acid region, proved to be two major drawbacks, for both multiplex targeting and saturation editing.

To overcome the above drawbacks, I decided to make the following two changes:

1) Use vectors as an HDR template – that could overcome the caveats of using ssODNs as repair templates.
2) Instead of using heterozygous PAM mutant cell line, generate a β-catenin KO cell line with puDeltatk counter selection cassette – that would provide a system for generating clean heterozygous mutants.

Conventional targeting approaches based on positive negative selection strategy have been successfully adopted in mESCs. The negative selection marker HSV1-tk, in combination with various positive selection markers, such as antibiotic resistance gene have been widely used (Schwartz *et al.*, 1991; Karreman, 1998). The HSV1-tk is capable of phosphorylating nucleoside analogs ganciclovir or FIAU, and incorporation of these modified variants during the replicative phase of DNA, terminates the elongation process, finally leading to cell death, and hence conferring sensitivity to tk. However, the original HSV1-tk caused sterility in male rodents and was a major setback for its use in transgenic approaches (Braun *et al.*, 1990). To overcome this drawback, tk was modified at the carboxy terminal, to produce a truncated version that allowed successful transmission through the male germline. Based on this a novel version of a counter selection cassette, by combining the puromycin N acetyltransferase and ∆tk, was described by Chen and Bradley. This puro HSV1∆tk fusion protein, under the control of PGK promoter and polyA signal from bGh, conferred resistance to puromycin and sensitivity to ganciclovir/FIAU (Chen and Bradley, 2000).

To overcome the drawback of generating clean heterozygous mutants, and given the advantages of positive negative selection, we decided to use this counter selection strategy to generate heterozygous β-catenin KO cell lines in both E14 and TCF cells. The β-catenin allele in these cell lines would be replaced by puDeltatk counter selection cassette using CRISPR/Cas9. Various β-catenin knockout conditional mouse models with loxp sites flanking either exon 3 and exon 6 or exon 2 and exon 6 of the *Ctnnb1* gene, have been widely used to study the role of β-catenin during development and disease (Brault *et al.*, 2001; Huelsken *et al.*, 2001). Taking this into consideration, heterozygous β-catenin KO cell lines were generated with a deletion between the exon 2 and exon 6 regions of the protein.

Briefly, one of the *Ctnnb1* alleles between intron 1 and intron 6 was knocked out and replaced by the puDeltatk selection cassette that allows puromycin based positive selection of the correctly targeted clone. In the second round of targeting, the puDeltatk allele of the β-catenin KO cell line was replaced by β-catenin allele with the desired mutation using FIAU negative selection, and this provided us with a robust system for

generating heterozygous mutants for both saturation editing and multiplex targeting approaches.

3.2.7.1 Cloning of puDeltatk targeting vector

For cloning of the puDeltatk targeting vector with β-catenin homology arm, initially a 5.5Kb of WT β-catenin region amplified from WT E14 DNA was cloned into PCR 4-TOPO vector, using the TOPO TA cloning strategy (Fig 3-17), and sequence verified.

Next, a Gibson assembly was designed for cloning the puDeltatk targeting vector (Fig 3-18). Using the 5.5Kb β-catenin TOPO vector as template, primers were designed to amplify the entire vector excluding the region between intron1 and intron 6 of β-catenin. This would constitute the 5' and 3' β-catenin homology arms of the vector. The puDeltatk region was amplified using primers having overlap with the β-catenin region. The two amplicons were then gel extracted and cloned using Gibson assembly. The transformed clones were screened by restriction digestion using XhoI and XbaI enzymes (Fig 3-19). Correctly digested clones were sent for sequencing, and one correct clone was selected to be used as HDR template for targeting.

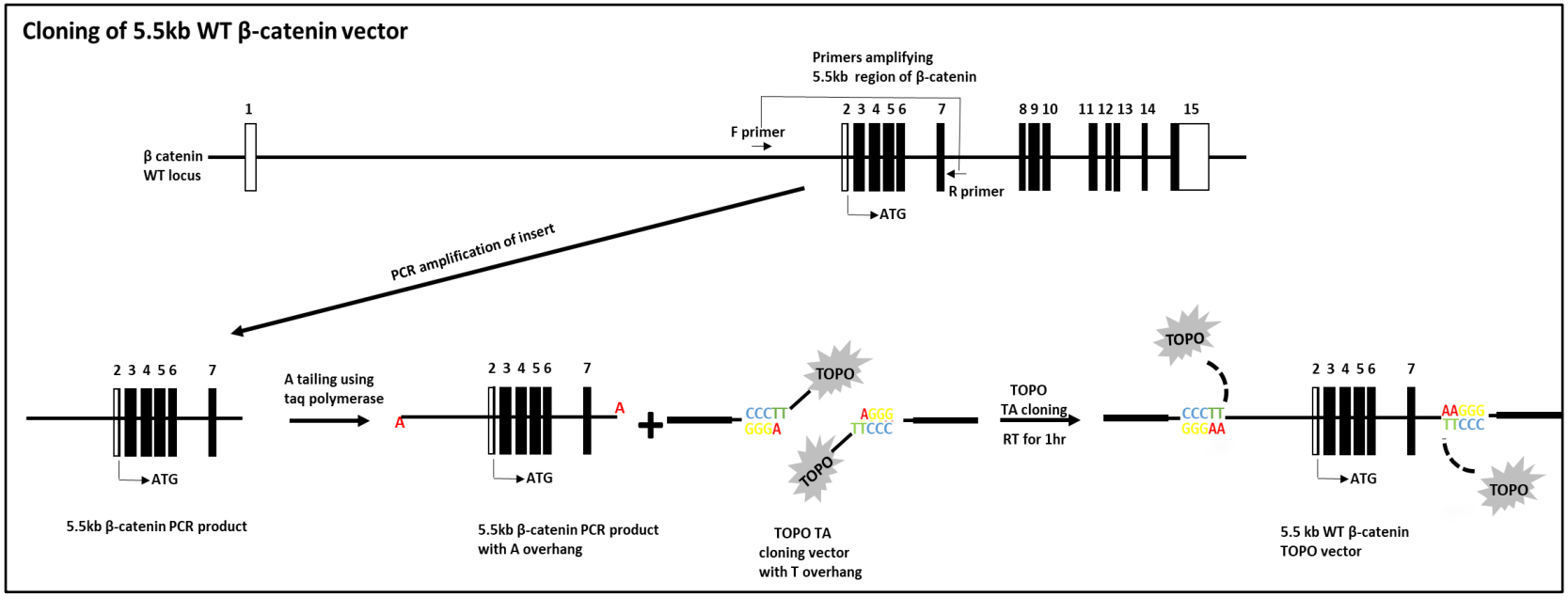


Figure 3-17: Schematic Representation of cloning of 5.5kb WT β-catenin TOPO vector. The 5.5kb WT β-catenin region was amplified using genomic DNA from mESCs. Following amplification, A overhangs were added to the 3'end of the PCR product using taq polymerase and subsequently cloned into TOPO TA vector.

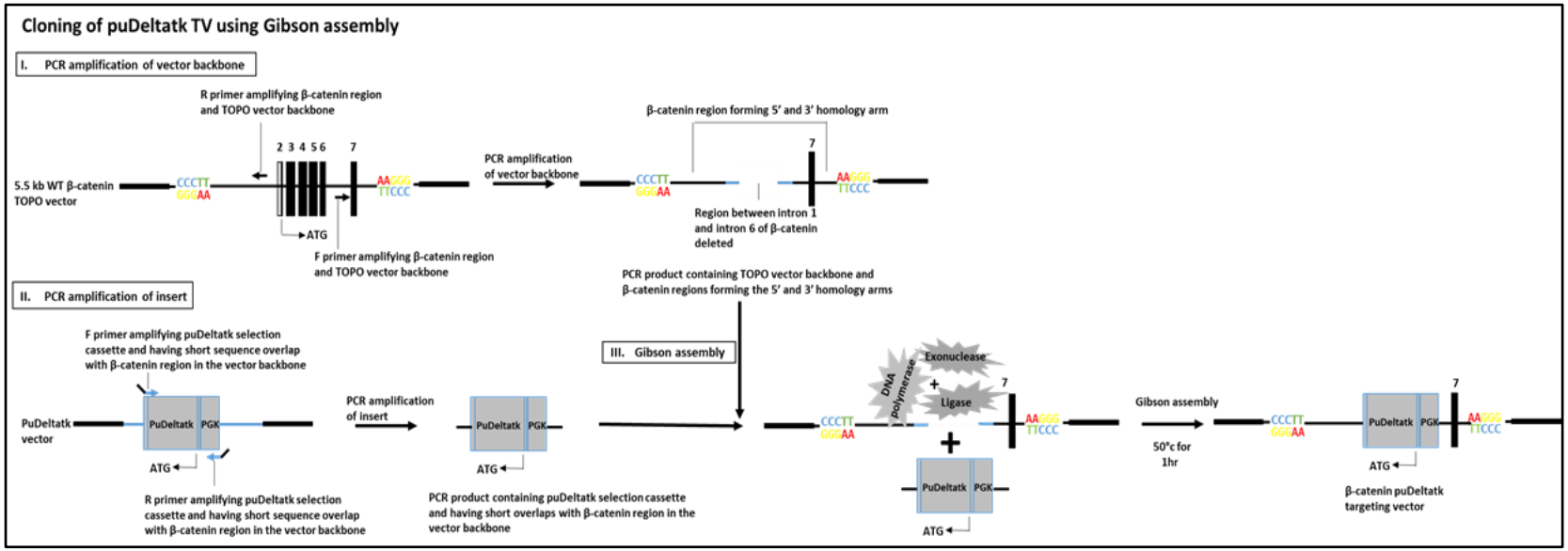


Figure 3-18: Schematic representation of cloning of β-catenin puDeltatk TV. The β-catenin homology arms along with the vector backbone were amplified from 5.5kb WT β-catenin TOPO vector. The puDeltatk selection cassette (insert) was amplified using primers having a short overlap with the β-catenin region in the vector backbone. The overlapping backbone and insert fragments were cloned by Gibson assembly.

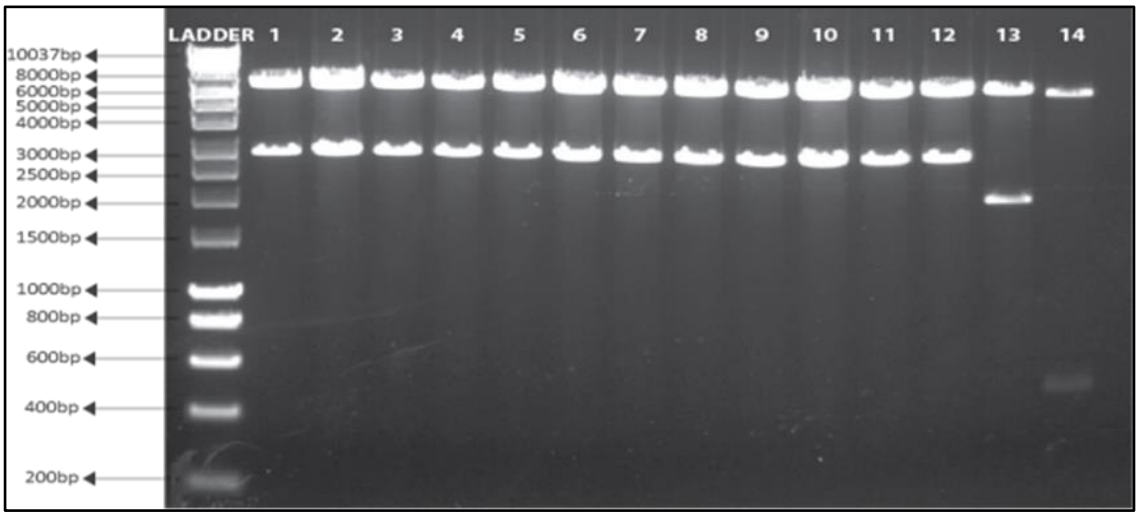


Figure 3-19: Restriction Digestion of puDeltatk targeting vector. Agarose gel electrophoresis image of Restriction digestion of miniprep DNA from Gibson assembly transformed clones (Lane 1-12 /12 clones) and controls (Lane 13 and 14 clone 13 and 14) using enzymes XhoI and XbaI (2668 and 5623). Control 13 PGK vector and control 14 5.5Kb WT TOPO vector. Hyper ladder 1Kb.

3.2.7.2 Designing and cloning of CRISPR guides in intron 1 and intron 6 of β-catenin

In order to knock out the β-catenin allele, two CRISPR guides were designed in each of the intron 1 and intron 6 regions of β-catenin, and were cloned into mCherry pX458 vector (Fig 3-20A and B) and verified by sequencing.

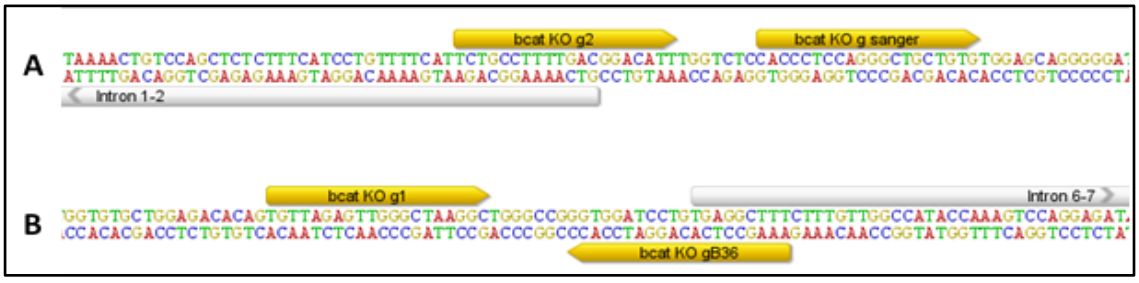


Figure 3-20: CRISPR guides to Knock-out WT β-catenin. Two guides each targeting the intron 1 and intron 6 of β-catenin were designed and cloned into mCherry pX458. (A) Sequence view of guides targeting intron 1 of β-catenin. (B) Sequence view of guides targeting intron 6 of β-catenin.

3.2.7.3 Targeting of mESCs to generate heterozygous β-catenin KO cell line

Given the advantages of positive negative selection for the generation of clean heterozygous β-catenin mutants, we decided to use this strategy for both multiplex targeting and the saturation assay. The mESCs, E14 and TCF cells were transfected with the puDeltatk targeting vector and intron 1 and intron 6 specific CRISPR guides (Fig 3-21A) using lipofectamine and L755507. Next day post transfection, the cells were trypsinized and plated at various densities in 10cm dishes, 8 hours after which the media was replaced with fresh media substituted with positive selection antibiotic puromycin. The clones were picked from both E14 and TCF transfections. A PCR based screening method was designed to identify the correctly targeted clones and specific primers were designed for identifying; A) 5' arm F primer from outside the vector homology arms - to avoid amplification of random integration and R primer in β-catenin region – to identify 5' end of WT β-catenin allele; B) 3' arm R primer from outside the vector homology arms and F primer in β-catenin region – to identify 3' end of WT β-catenin allele; C) 5' arm F primer from outside the vector homology arms and R primer in puDeltatk region – to identify 5' end of β-catenin KO allele with puDeltatk; D) 3' arm R primer from outside the vector homology arms and F primer in puDeltatk region – to identify 3' end of β-catenin KO allele with puDeltatk. PCRs were performed for DNA from several clones from both E14 and TCF targeting and correctly targeted clones having bands in all 4 PCRs were sequenced (Fig 3-21B). The majority of the clones had incorporated the puDeltatk selection cassette in one of the alleles. However, among the sequenced clones, I was unable to find clones with clean WT allele (due to the biallellic nature of CRISPR activity) and since the indels were in the intron region of β-catenin, the chances of the protein being affected were very low. Hence one clone each from both E14 and TCF with correctly targeted puDeltatk allele and having minimum indels in the intron region were selected for further experiments.

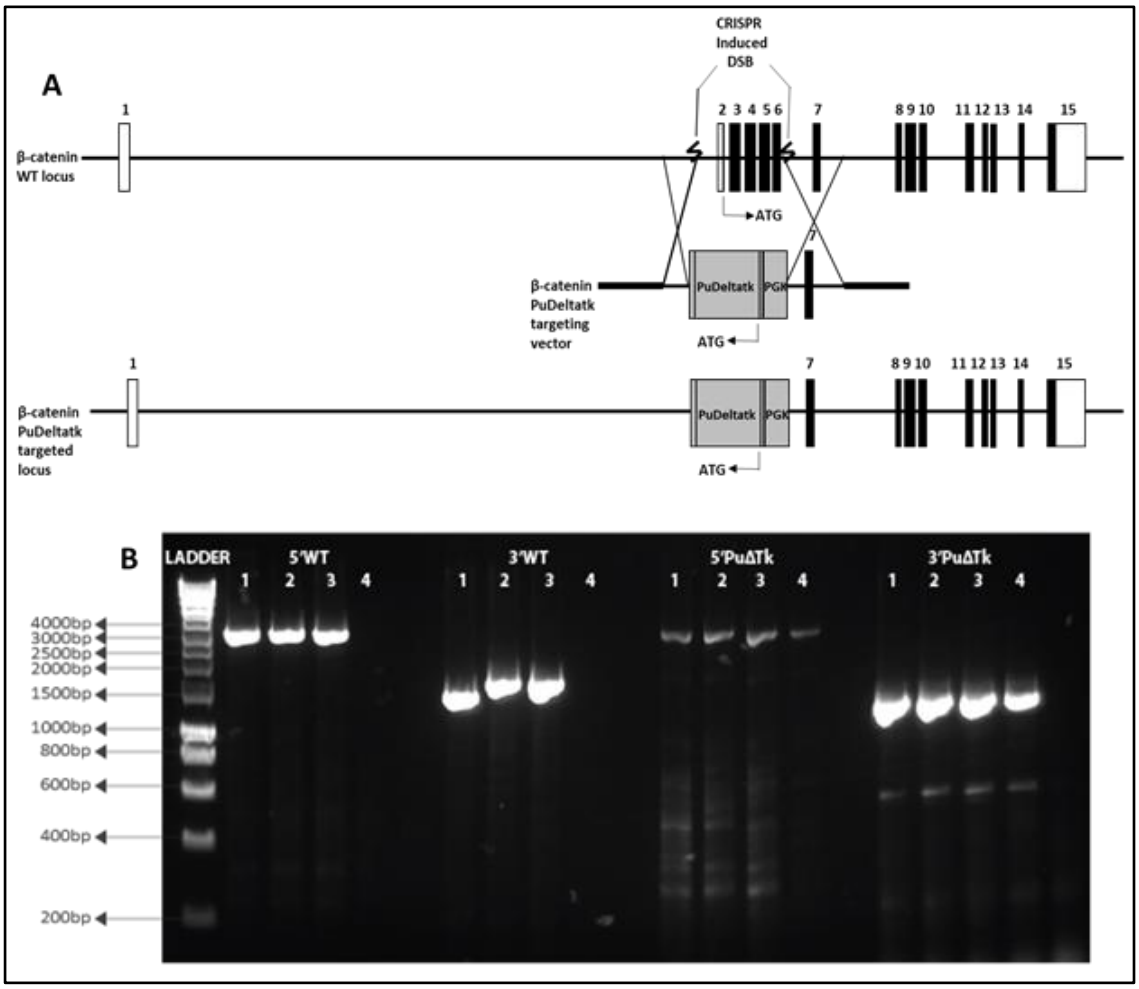


Figure 3-21: Targeting strategy for generation of heterozygous β-catenin KO cell line. Schematic representation of β-catenin WT locus, puDeltatk targeting vector and puDeltatk targeted allele after CRISPR mediated HDR. Coding exons are black boxes and introns are solid lines. (B) PCR screening of β-catenin KO cell lines: Agarose gel electrophoresis image of 5' WT, 3' WT, 5' puDeltatk and 3' puDeltatk of PCR performed for four representative clones. Clones 1,2,3 were positive for all 4 PCRs whereas clone 4 did not show any band for 5'WT or 3' WT PCRs. Similar PCR was done on several clones from both E14 and TCF targeting and few of the rightly targeted clones having bands in all 4 PCRs were sequenced. Hyper ladder 1Kb.

In evidence with homozygous β-catenin deletion having been reported to result in lethal phenotypes *in vivo*, and increased cell death in *in vitro* culture systems, I observed a similar effect of the loss of β-catenin on the viability of ES cells (Haegel *et al.*, 1995) (Raggioli *et al.*, 2014). The majority of the targeted cells of both E14 and TCF

backgrounds were heterozygous for β-catenin. In addition, heterozygous β-catenin knockout E14 cells that were being cultured in normal ES media were especially sensitive to loss of β-catenin, and β-catenin expression from a single allele seemed to be insufficient in maintaining the integrity of these cells, evident by increased differentiation and cell death, and hence needed the support of 2i for continual maintenance in culture. However, no loss of viability was observed in heterozygous β-catenin KO TCF cells that were already being cultured in media supplemented with 2i. The need for β-catenin in maintaining the integrity of ES cells, increased the efficiency in generation of successful heterozygous β-catenin KO cell line in our targeting experiment.

3.2.7.4 Analysis of β-catenin expression in β-catenin KO cell line

The sequencing results confirmed the KO of one of the β-catenin allele and replacement with puDeltatk selection cassette. However, to validate the functional expression of the β-catenin WT allele, the total protein was separated by performing SDS polyacrylamide gel electrophoresis followed by Western blot of the protein, using pan β-catenin antibody and Proliferating Cell Nuclear Antigen (PCNA) as loading control (Fig 3-22A). The β-catenin KO cell lines of both E14 and TCF cells expressed β-catenin. However, both WT and β-catenin KO cell lines showed similar quantities of protein expression.

In addition to analysis of protein expression by western blot, the β-catenin activity in heterozygous β-catenin TCF KO cells was analysed by flow cytometry. The WT TCF cells had higher GFP intensity in comparison to β-catenin heterozygous KO TCF cell line indicating the reduction in β-catenin activity in the β-catenin heterozygous KO cell line as expected (Fig 3-22B).

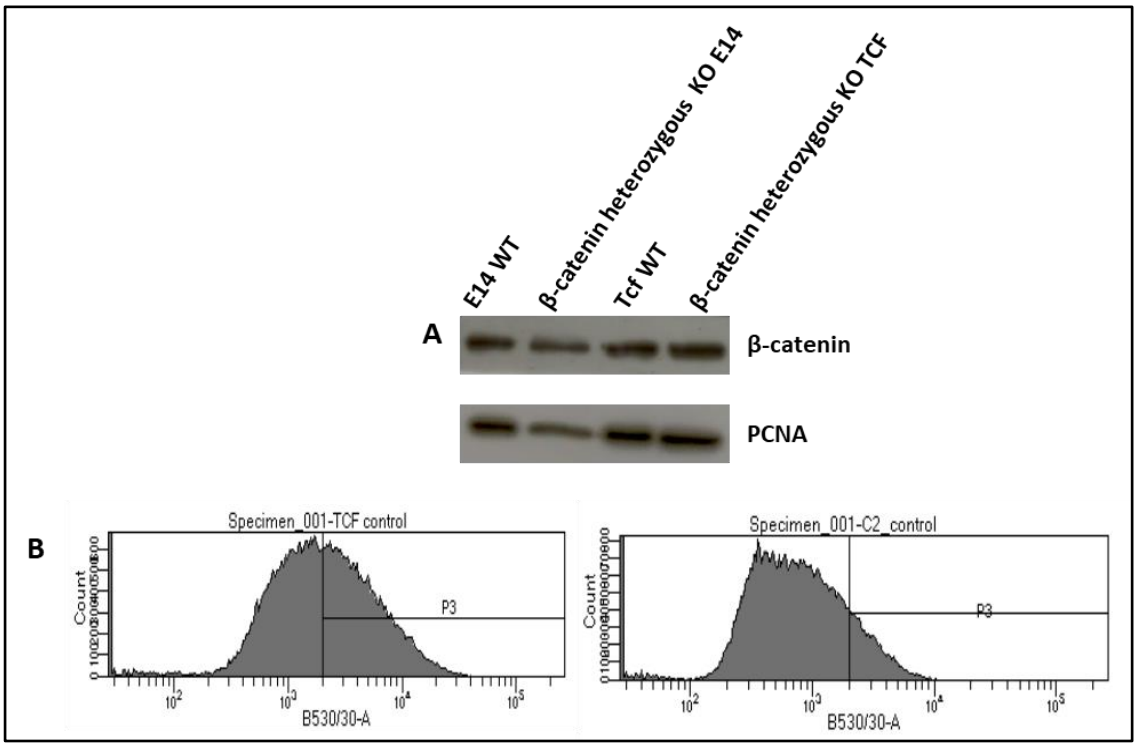


Figure 3-22: Analysis of β-catenin expression in heterozygous KO cell lines. (A) Western blot of β-catenin WT and heterozygous KO cell lines: Analysis of protein expression of WT and heterozygous KO E14 and TCF cell lines using pan β-catenin antibody indicating β-catenin expression in both WT and β-catenin heterozygous KO cell lines. PCNA was used as loading control. (B) Flow cytometry analysis of β-catenin activity in WT and heterozygous KO TCF cell line: Histograms of WT TCF (TCF control) and heterozygous KO TCF cell line (C2). The GFP positive cells are represented in the P3 gate. The heterozygous KO cell line showing lower number of GFP+ cells indicating reduced β-catenin activity in comparison to the WT TCF control. The GFP+ cells were gated using E14 as negative control.

3.2.8 Cloning of β-catenin vector with BbsI Restriction site

To overcome the drawbacks of using an ssODN as HDR template, our second strategy was to generate targeting vectors. However, to individually generate 28 vectors for the purpose of multiplex targeting and 400 vectors for saturation editing, would have been a laborious and time consuming process. To clone all the targeting vectors in an efficient way, I decided to take advantage of the Golden gate cloning strategy. Golden gate cloning is based on the Type IIS restriction enzymes that cut outside their recognition site and allows seamless cloning, with the advantage of performing restriction digestion and

ligation in a single step reaction (Engler, Kandzia and Marillonnet, 2008). Type IIS restriction enzymes that leave 4bp overhang including BbsI have shown to produce the best result. Owing to the advantages of this system, we decided to clone a β-catenin backbone vector by inserting a pair of type IIS restriction enzyme BbsI in the region of our interest (Fig 3-23). This β-catenin Golden gate vector would later be used as a destination vector to clone all the targeting vectors required for both multiplex targeting and Saturation editing in a single step (Fig 3-24).

A Gibson assembly was designed to clone the β-catenin golden gate vector using the 5.5Kb WT β-catenin TOPO vector as template. However, this template already had a BbsI site in the 5' end of β-catenin region, and hence primers were designed such that they included a synonymous mutation in BbsI recognition site. In addition, to be able to identify the TVs with inserts while cloning the multiple vectors, the region of our interest was deleted and replaced with two BbsI sites in the opposite orientation that would allow size based separation of positive vectors.

Using this strategy, both the vector backbone and insert were amplified using 5.5Kb WT β-catenin vector. The PCR products were then gel eluted and cloned using Gibson assembly protocol. The transformed clones were screened by double digestion using BbsI and NotI restriction enzymes (Fig 3-25). Following the preliminary screening, correctly digested clones were sequenced and one correct vector was selected for future use.

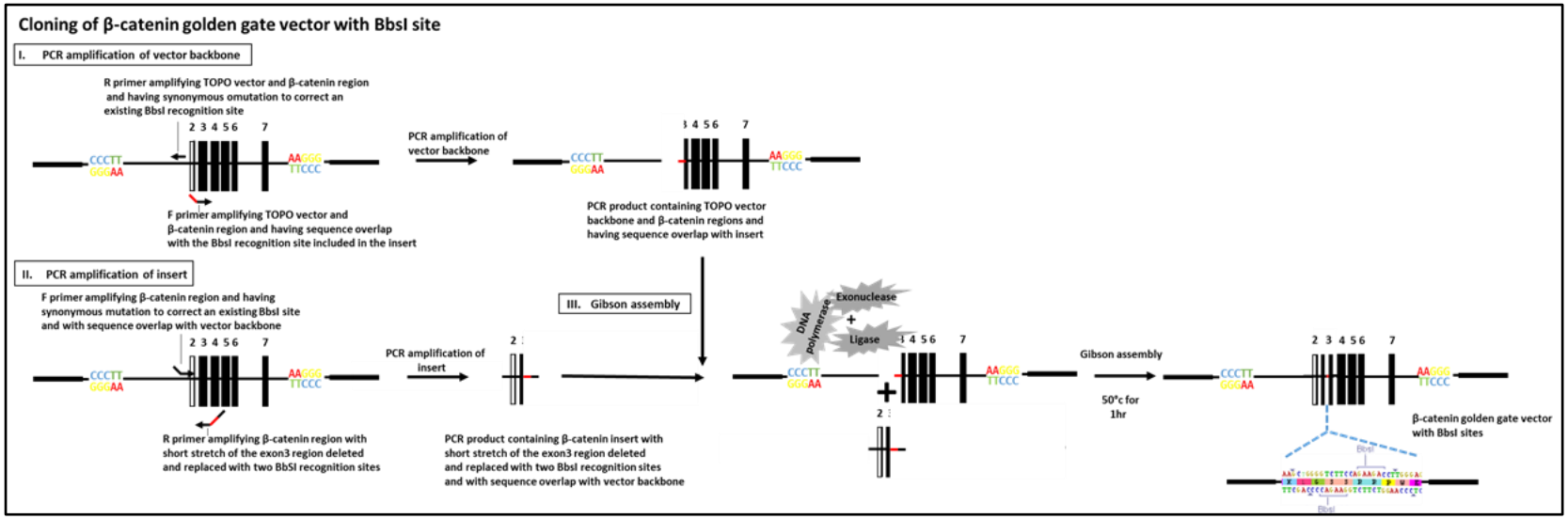


Figure 3-23: Schematic representation of cloning of β-catenin golden gate vector with BbsI sites. The 5.5Kb WT β-catenin TOPO vector was used to amplify both the vector backbone and the insert (with BbsI sites and deleting the region of interest) having short sequence overlap with each other. The overlapping backbone and insert fragments were cloned by Gibson assembly.

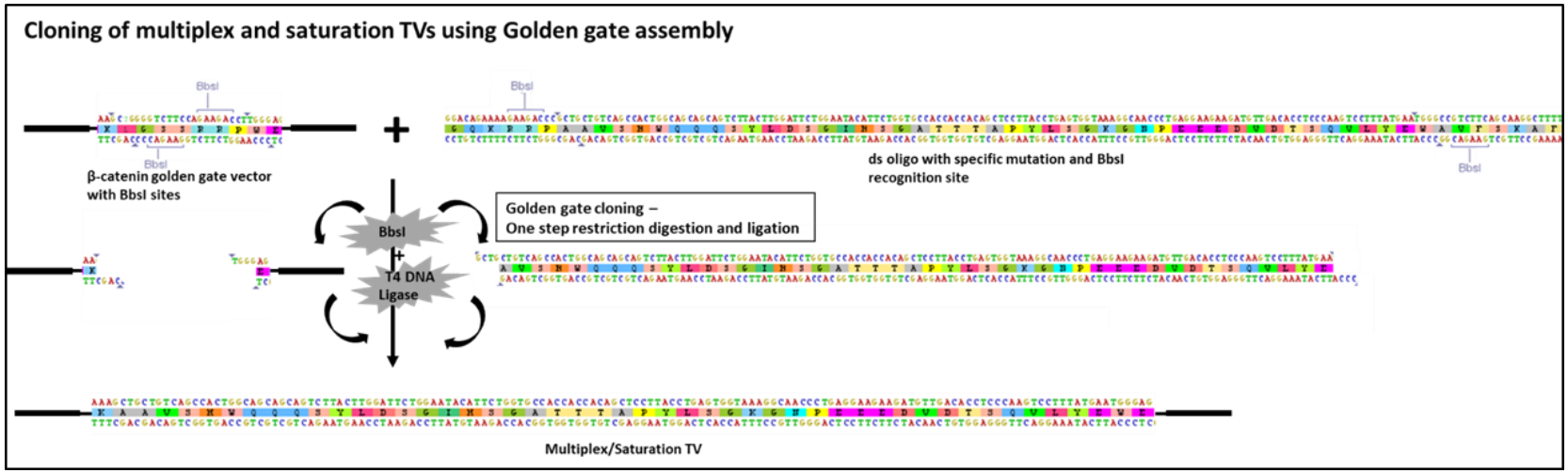


Figure 3-24: Schematic representation of cloning of multiplex and saturation TV using golden gate assembly. The β-catenin golden gate vector with BbsI sites was used as a destination vector to clone the multiplex and saturation TV in a single step restriction digestion and ligation reaction.

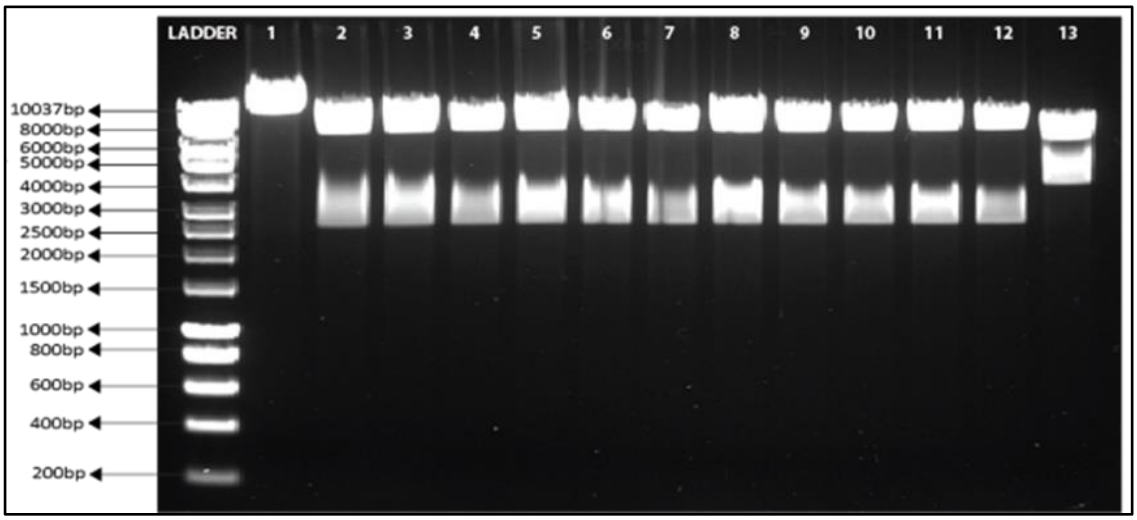


Figure 3-25: Restriction digestion of β-catenin Golden gate vector. (A) Map of β-catenin golden gate vector. The β-catenin golden gate vector with BbsI restriction sites was cloned to be used as a backbone vector for cloning the TVs for saturation editing and multiplex targeting. (B) Agarose gel electrophoresis image of Restriction digestion of miniprep DNA from Gibson assembly transformed clones (Lane 1-12) and control (Lane 13) using enzymes BbsI and NotI(2482bp and 6820bp). 5.5Kb WT TOPO vector was used as control. Hyperladder 1Kb.

3.2.9 Preliminary functional analysis of β-catenin mutation

Another potential problem arose when I analysed the wild type β-catenin reporter cell line (refer Fig 3-22B) using flow cytometry. The data showed that the β-catenin activity was not uniform in all cells within the population. This could have generated a challenge in identifying the potential difference in β-catenin activity in various different mutants. Therefore, before the generation of mutant clones using multiplex targeting and saturation assay for interpreting the functional significance of β-catenin mutation, it was important to conduct a preliminary analysis. The major significance of this experimental approach was to test whether or not the difference in β-catenin activity between different mutants would be apparent when analysed in a pool of cells, and that the differences in activity level was indeed the effect of mutation and not a variation observed due to clonal heterogeneity of mESCs. For this purpose, I strategized to use two mutations, S33Y and ΔS45 mutations, that were being extensively studied in our lab by my mentor and senior post doc Derya Ozdemir. The aim was to target the β-catenin reporter cells with either S33Y or ΔS45 HDR vectors and analyse them in pools, rather than generating clonal cell

lines. To ensure all the cells in the pool analysed were mutant cells, I used fluorescent tagging and FACS sorting. The targeting vectors required for this experiment was already available in the lab. To generate a pool of heterozygous S33Y cells, I targeted the TCF/Lef:H2B-GFP β-catenin reporter cells with wild type β-catenin vector tagged with RFP and S33Y vector tagged with BFP. By transfecting the reporter cells with these two vectors and selecting for the cells which were double positive for RFP and BFP, I made sure to have a heterozygous population (Fig 3-26). This transfection was also done using WT-RFP and ΔS45-BFP. As a control, I transfected the cells with WT-RFP and WT-BFP. For the targeting, the 2 guide CRISPR system discussed above was used, and this time the guides had puromycin selection instead of a fluorescent protein. All these 3 pools of cells were analysed by flowcytometry using GFP reporter as a β-catenin activity readout. Analysis by flow cytometry indicated a shift in β-catenin activity by both ΔS45 and S33Y mutant pool, when compared to WT (Fig 3-28A). The greater shift observed in S33Y peak indicated a higher increase in β-catenin activity when compared to ΔS45 mutant pool, showing that the mutants differed in their ability to activate β-catenin regardless of the clonal variation.

In order to analyse the mutants in hemizygous state, I repeated the experiment in β-catenin heterozygous KO TCF cell line (Fig 3-27). The cells were transfected using either S33Y BFP, ΔS45 BFP or WT BFP along with two CRISPR guides scg3 and g9B, keeping the puDeltatk allele intact. Analysis by flow cytometry of the sorted hemizygous cells also showed the similar trend, S33Y giving a higher β-catenin signal than ΔS45 (Fig 3-28B). The differences between the populations was more prominent in the hemizygous condition when compared to heterozygous condition, probably due to the activity coming from the WT allele in the heterozygous state.

This experiment showed that, although there seem to be some clonal heterogeneity, evident from the width of bell shaped peak of the pooled populations, there exists a clear difference in the effect of the two mutations on β-catenin activity. In conclusion, the differences in the activity conferred by the two mutations in this experiment provided a preliminary validation of our hypothesis of the existing genotype-phenotype correlation among β-catenin mutation.

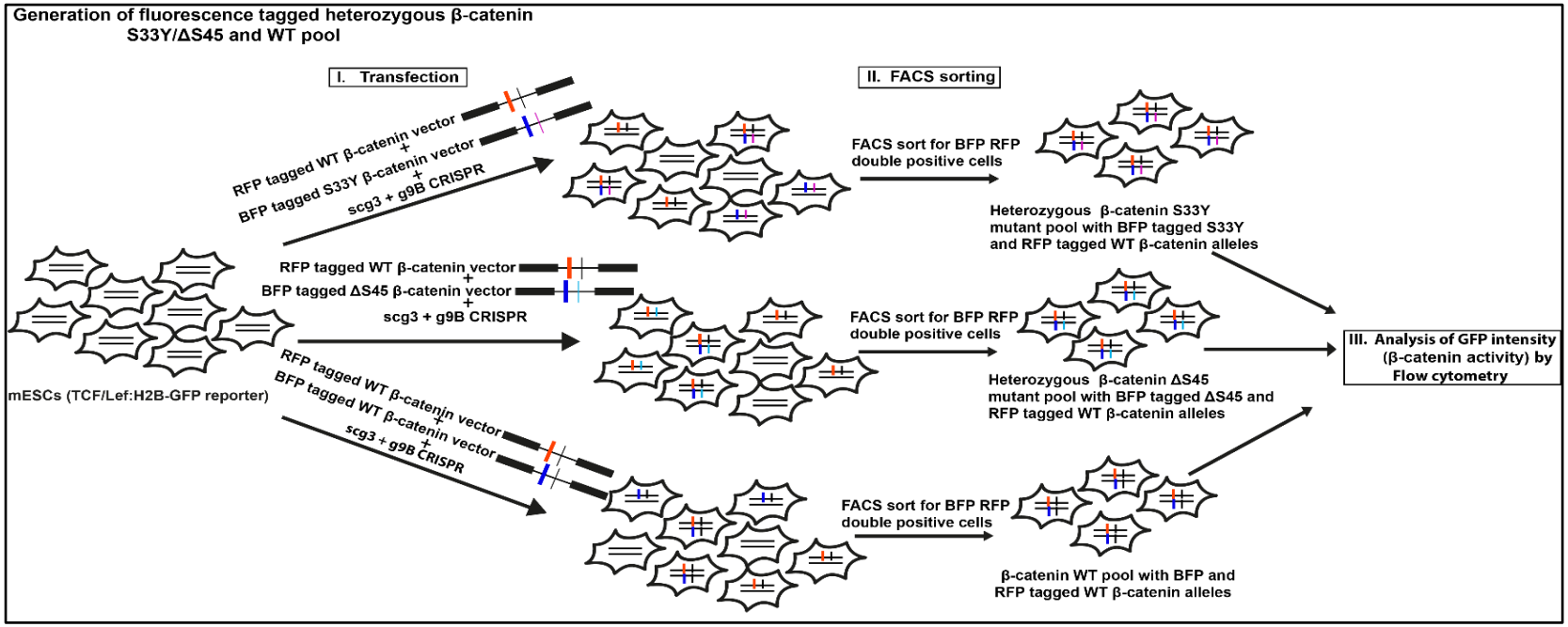


Figure 3-26: Schematic representation of generation of fluorescence tagged heterozygous β-catenin S33Y/ΔS45 and WT pool. The TCF cells were transfected with RFP tagged WT β-catenin vector and BFP tagged S33Y/Δ S45/ WT β-catenin vector along with the CRISPR guides. The BFP RFP double positive cells were sorted from each of the three targeted population. The sorted cells were then analysed for GFP intensity by flow cytometry.

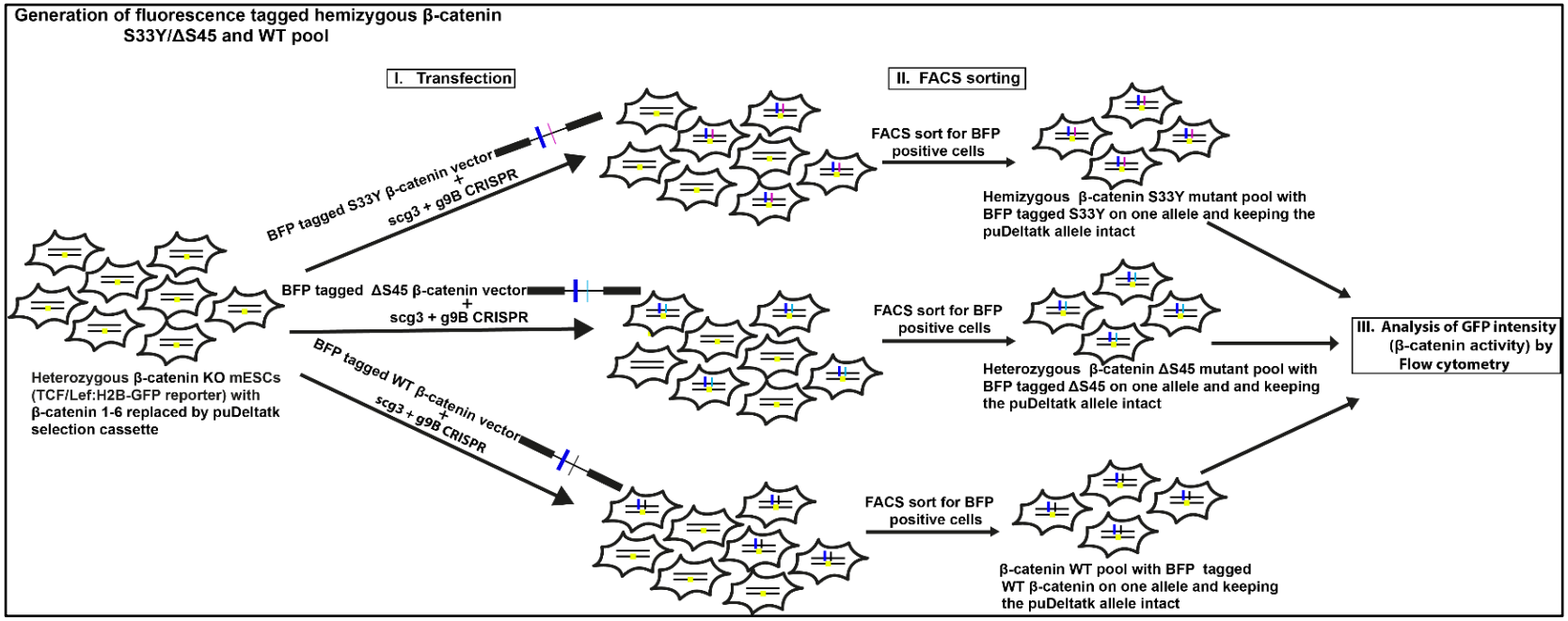


Figure 3-27: Schematic representation of generation of fluorescence tagged hemizygous β-catenin S33Y/ΔS45 and WT pool. The heterozygous β-catenin KO TCF cells were transfected with BFP tagged S33Y/ΔS45/WT β-catenin vector along with the CRISPR guides. The BFP positive cells were sorted from each of the three targeted population. The sorted cells were then analysed for GFP intensity by flow cytometry.

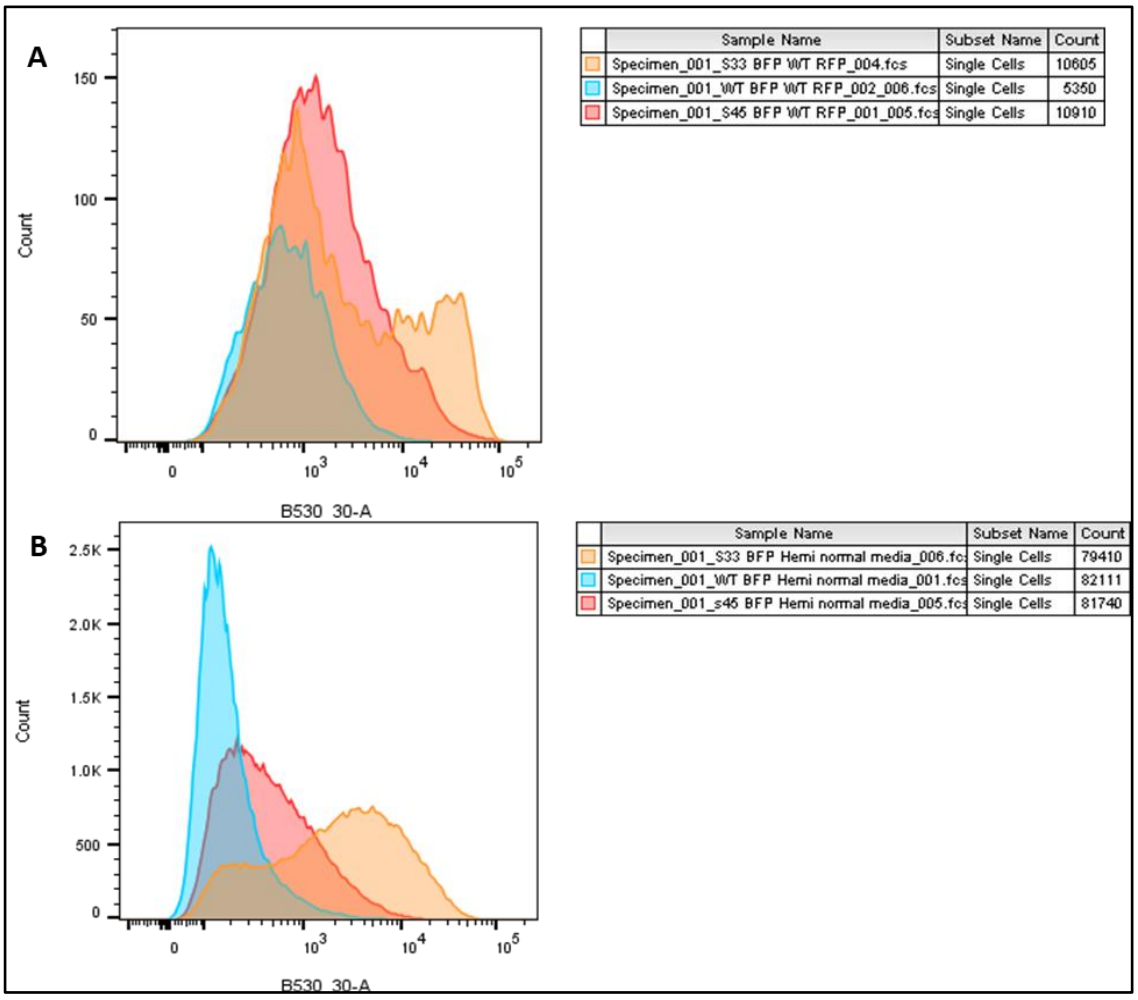


Figure 3-28: Flow cytometric analysis of pooled mutant S33Y, ΔS45 and WT TCF clones. (A) Histogram of heterozygous S33Y, ΔS45 and WT pooled samples. (B) Histogram of hemizygous S33Y, ΔS45 and WT pooled samples. The differences in the observed shift between S33Y, ΔS45 and WT peaks indicating the existence of genotype phenotype variation with respect to β-catenin activity.

3.3 Discussion

Over the years, most of the experimental genetic analysis on the activation of oncogenes or repression of tumour suppressor activity have relied on either cDNA based overexpression, or knockdown studies mediated by RNAi strategy, both these approaches have their own limitations. In addition, the conventional gene targeting strategies based on HR, although widely used for transgenic technology, falls short in its

extended applicability to large scale genome engineering projects. The recently developed gene editing techniques, predominantly the CRISPR/Cas9 system, is reported to be capable of mediating endogenous genetic manipulations in a range of *in vivo* and *in vitro* systems, and these RNA guided nucleases have revolutionized the field of genetic engineering with implications in a variety of fields, including cancer biology (Sánchez-Rivera and Jacks, 2015).

Our goal of understanding the functional significance of the many mutations observed in β-catenin would have otherwise been labor intensive and time consuming. However, the ease and versatility of the CRISPR/Cas9 system has provided a perfect platform for our investigation. Further, the CRISPR/Cas9 system in combination with the two approaches, saturation screening and multiplex targeting, will provide a better perspective of genotype-phenotype correlations of β-catenin mutations. I decided to use mESCs, E14 for multiplex targeting and TCF/Lef:H2B-GFP reporter cell line for saturation editing approaches.

The mESCs were selected for the purpose of *in vitro* analysis of genotype-phenotype correlations in β-catenin. ES cells have several advantages - The HDR efficiency of ES cells is better in comparison to that observed in somatic cells (Kass *et al.*, 2013); mESCs have been studied extensively and are one of the better characterized in-vitro systems; *in vivo* differentiation being one of the functional readouts of β-catenin activity, the mutant alleles can be readily used to generate teratomas, and the endogenous mutant cell lines of interest can be used directly to generate mouse models. However, this system also has certain disadvantages. The Wnt/ β-catenin signaling is one of the major pathways governing the 'stemness' of ES cells, and hence β-catenin signaling activity resulting from specific mutations may have an impact on self-renewal and differentiation potential of ES cells (Merrill, 2012). The preferential selection of specific mutations among different cancers point towards a tissue specific selection. However, ES cells may fail to select for mutations at particular residues or the mutations that affect the viability of the ES cells may simply be lost for further analysis. In spite of these caveats, the potential advantages conferred by these cells provided a strong reason for choosing mESCs for *in vitro* modelling of β-catenin mutations. The understanding of β-catenin activity using mESCs will be a starting point and will provide a good perspective of the genotype-phenotype

correlation of the observed β-catenin mutations, laying a strong foundation for future analysis.

For the purpose of saturation editing and multiplex targeting, design and validation of the CRISPR guides to induce strand breaks was necessary. Since different guides may vary in their cutting ability, six different guides spanning the region of interest were designed and cloned into both the mCherry pX458 and GFP pX458 vectors. The mCherry and GFP cassette in the pX458 transfected cells allows selection of transfected cells.

The analysis of indel percentages of these guides varied between 17 to 35 percent in the T7 assay. Although T7 endonuclease I assay is known to be a sensitive approach (Vouillot, Thélie and Pollet, 2015), we did not observe a correlation in the indel percentages analysed by T7 assay and the number of targeted mutant clones analysed by sequencing. The g19 CRISPR which resulted in an indel percentage of 22 percent in the T7 assay showed a mutation of greater than 70 percent in the clones analysed by sequencing and the g5 CRISPR with a higher indel percentage (30 percent) in T7 assay resulted in less than 5 percent of mutation in the clones analysed by sequencing (data not shown). These results suggest a need for a more sensitive approach to predict the efficiency of guides to induce DSBs until then it is difficult to select guides based on the T7 assay alone.

Although the CRISPR/Cas9 system is a simple and robust technique for genome editing, the efficiency of this RNA guided endonuclease for HDR mediated precise gene manipulation is significantly limited by the increased frequency of NHEJ events (Maruyama *et al.*, 2015). This was observed even in our study, whilst most of the CRISPRs were efficient in induction of DSB (as indicated by NHEJ events), the frequency of HDR still remained poor. The initial targeting experiments for generation of a heterozygous PAM mutant allele resulted in very low HDR events and we did not succeed in generating a heterozygous PAM mutant clone in our initial attempt. The low HDR events observed at our target site would result in reduced efficiency for our multiplex targeting approach. For the purpose of multiplex targeting, we required a higher HDR frequency, and hence further optimization was required to improve the HDR events. In this regard, various approaches to reduce the inherent NHEJ mechanism and to increase the HDR frequency have provided promising results. One such compound SCR7

interacts with the DNA binding site of DNA ligase IV, a key enzyme in canonical NHEJ pathway, was shown to significantly reduce the NHEJ and subsequently increased the HDR events (Chu *et al.*, 2015; Maruyama *et al.*, 2015). The use of this compound, although increased the HDR events to a small extent, but the frequency of HDR was still reasonably low for our purpose of multiplex targeting. However, the drug was used at the same concentration (1μM) that was shown to be optimal in MelJuSo, DC2.4 cells and I did not optimize the concentration for our cell line, which might be the reason for the lack of significant increase in HDR frequency in our experimental set up. Further optimization of the drug concentration in E14 cells would be required to draw conclusions on the efficacy of this DNA ligase IV inhibitor in increasing HDR events at our target site.

In addition to SCR7, another small molecule compound, L755507, was identified in a reporter based screening method as having an ability to promote HDR events upon induction of DSB by the CRISPR/CAS9 system. The screen involved around 4000 small molecule compounds, whose biological activity had been well characterized, and among all the compounds L755507, a β adrenergic receptor agonist, yielded a 3 fold increase in large insertions and 9 fold increase in point mutations (Yu *et al.*, 2015). The use of this small molecule compound in combination with lipofection resulted in a drastic increase in the HDR frequency. Although the authors used the electroporation method for L755507, our results with the same drug previously used with nucleofection did not result in increase of HDR. Our main aim was to improve the HDR frequency and as these experiments are time consuming, in most of our targeting experiments I had pooled different strategies for optimization. Since the combination of L755507 with lipofection had yielded similar HDR frequency in repeated targeting experiments, I decided to continue to use this approach. However, an experimental set up with appropriate controls for comparison of nucleofection vs lipofection in the presence and absence of the drug would provide conclusive results.

Previous gene editing studies have shown the successful use of short single stranded oligodinucleotides (ssODN) especially for insertion of point mutations and with an efficiency comparable to that of using targeting vectors with long homology arm (Soldner *et al.*, 2011). In addition, ssODNs with a length of 70bp have been shown to achieve an optimal frequency of HDR (Yang *et al.*, 2013). The CRISPR based targeting experiments

for introduction of short inserts and point mutations using ssODN have been done successfully in our lab. However, the only paper on saturation screening uses a plasmid with long homology arms, and hence it was decided to test the efficiency of using a targeting vector as HDR template (Findlay *et al.*, 2014). The ΔS45 targeting vector with 1Kb arm resulted in a good frequency of HDR percentage.

It is well established that, the insertion of a synonymous mutation in the NGG PAM while designing a repair template reduces the guide mediated recognition and re-cutting by Cas9. However, in certain cases it is not possible to introduce a silent substitution in the PAM sequence for many guides, or can only be substituted to an NAG PAM in the targeting template. It has been reported that in addition to NGG PAM, SpCas9 has the ability to cleave NAG PAM, albeit with 1/5th the efficiency compared to NGG PAM, thus resulting in additional indels observed along with HDR events (Hsu *et al.*, 2013). To overcome the re-cutting of NAG PAM, a template was designed with introduction of an additional silent mutation in the 19thbp of the g9B CRISPR binding site. Various studies on off- target effect have proposed the region consisting of 8-10 nucleotides proximal to the PAM sequence (known as the seed sequence) with requirement of high specificity and mismatches in this region are less tolerated and resulted in reduced Cas9 activity (Hsu *et al.*, 2013; Anderson *et al.*, 2015). In agreement with this, our final strategic approach of designing a template with introduction of an additional silent mutation in the seed sequence greatly reduced the chances of re-cutting by Cas9. This strategy of introduction of a synonymous mutation in one or a couple of nucleotide positions in the seed sequence prevented re-cutting of the NAG PAM to a great extent and yielded a large number HDR only clones.

The initial multiplex targeting of residue S45, incorporating the various optimized strategies resulted in successful generation of both homozygous and heterozygous mutants. However, multiplex targeting at the T41 residue did not result in successful generation of the required T41 variants/mutants although the HDR efficiency was reasonable and 17 percent of the clones had incorporated the PAM mutation but only 3 percent of the clones had both T41 and PAM mutations. This may be attributed to the distance between CRISPR cut site and the mutation being incorporated. However, the previous S33Y targeting with g19 CRISPR that cuts 12bp away had yielded a HDR

percentage of 15.6 and all the clones, except one clone, had incorporated both PAM and S33Y mutations with the distance between PAM and S33Y being 6bp. In the case of the T41 residue, although the distance between the cut site and T41 is 10bp, the distance between PAM and T41 was 15bp. From these results, it can be speculated that the presence of two close mutations have better chances of both being incorporated when compared to T41 mutations where the distance between PAM and T41 mutation is slightly large, hence the PAM is being repaired but without the T41 residue mutation. These results can be attributed to the process of template switching, a common feature in SDSA (Synthesis dependent strand annealing, a repair mechanism widely used when ssODN acts as repair template). In support with our observation Paix et al have shown that the presence of mutation every 3 or 6bp reduces template switching in comparison to mutations 12bp apart where increased template switching was observed (Paix *et al.*, 2017).

In addition, other factors may contribute to the observed variation in HDR frequency even within a short stretch of locus, for instance purine and pyrimidine rich regions have known to favour HDR, and the region surrounding S45 being extremely purine rich and with g9B inducing a DSB very close to this region, probably results in higher HDR percentage observed. Reports have suggested the heterozygous mutation is favored when increased distance between CRISPR cut site and homozygous when the CRISPR cut site is closer (Paquet *et al.*, 2016), similarly the distance between the S45 and g9B being the shortest may have contributed to the increased homozygous mutants in S45 targeting as compared to either S33Y targeting or T41 multiplex, both having higher number of heterozygous mutants. These results implicate the existence of multiple factors that contribute to the observed differences in HDR based editing rates, especially when using ssODN as repair templates, thus highlighting the importance of carefully designed templates and guides for need based enhancement of targeting efficiency.

However, with only two efficient CRISPRs in the region of interest, and the increased variability in HDR outcome when using ssODN as repair templates, proved to be the major limitations, creating bottlenecks in our experimental progress for both multiplex targeting and saturation editing. Although few of the drawbacks of using ssODN as repair template could be overcome by using vector based HDR template, and a good HDR

efficiency was observed with 1Kb homology arm vectors, it was still difficult to generate clean heterozygous mutants, without NHEJ on the other allele. Moreover, our strategy of generation of PAM mutant cell line remained unsuccessful, requiring the need for a new strategy. Using a haploid cell line, or knocking out one of the WT allele in diploids, and studying the mutations in a hemizygous condition would have been two other possibilities, that could substitute for PAM mutant cell line. However, I decided to continue to analyse heterozygous mutants by generating a heterozygous β-catenin KO cell line with puDeltatk counter selection cassette. This positive negative counter selection strategy, along with the use of vectors as HDR template was our best chance at generating clean heterozygous mutants. For this purpose, β-catenin KO heterozygous cell lines were generated in both E14 and TCF cells. In addition, a β-catenin destination vector with BbsI site was cloned that could be used as backbone for efficient Golden gate based cloning of all the vectors for both multiplex targeting and saturation editing.

As proof of principle, to test whether or not different mutations resulted in differences in β-catenin activity, a preliminary experiment performed using pooled S45 and S33Y mutants in TCF cell line showed differences in activity of the two mutants, providing initial evidence of the genotype- phenotype correlation.

In conclusion, every targeting experiment as discussed in this chapter, although time consuming, provided valuable insights, and helped us design the best possible strategies and tools for enhancing the targeting efficiency at the region of our interest. Furthermore, the preliminary validation of our hypothesis provided a strong basis for our complementary approaches of multiplex targeting and saturation editing.

Chapter 4 Saturation editing of β-Catenin hotspot region

4.1 Introduction

The directed evolution experiments are frequently adopted in the field of protein/enzyme engineering. These experiments are based on the creation of a mutant library encompassing a diverse array of amino acid variants, followed by iterative selection and high throughput screening for assessing the fitness of the biocatalyst, providing an efficient system for selection of mutants with the desired or novel catalytic activities (Cobb, Chao and Zhao, 2013). Not limited to protein engineering, these strategies have also found applicability in evolutionary analysis of proteins, deciphering metabolic pathways, and are further expandable to genome wide level. Adapting similar approaches, recently it has been shown that saturation editing combined with the CRISPR/Cas9 provides a prospective way to analyse the functional significance of every nucleotide/amino acid residue within a short stretch of the endogenous loci of interest (Findlay *et al.*, 2014). Given the diversity in the mutational landscapes of the cancer genome, CRISPR/Cas9 mediated saturation editing offers a simple tool for induction of multiple genetic variations to create a mutant library in a single large scale assay. Moreover, the availability of high throughput sequencing platforms provide perfect complementary screening approaches for such large scale strategies, and together with a detectable phenotypic assay, provide a robust means of determining the precise functional consequence of every variant in the generated mutant library.

Given that a majority of the cancer mutations in the β-catenin oncogene are clustered around the conserved phosphorylatable serine and threonine residues, with a defined mutational hotspot in the exon 3 region, makes this region of *Ctnnb1* a suitable candidate to perform saturation editing. Functional analysis of β-catenin activity based on the TCF/Lef reporter system, allowing quantification of β-catenin activity at single cell resolution through FACS, provided an elegant screening approach for phenotypic assesment. The mES cell line with the TCF/Lef:H2B-GFP reporter required for this experiment was derived for us in the lab of Kat Hadjantonakis (Memorial Sloan-Kettering Cancer Center), from the blastocysts of a well-established reporter mouse model (Ferrer-Vaquer *et al.*, 2010). As discussed before, in order to mimic the β-catenin mutations in cancer as closely as possible, the targeting strategy was based on generating heterozygous mutations. 400 substitutions, targeting a 20 amino acid region (residues L31-G50), were randomly generated in 200 million reporter cells, which was then divided

into 6 segments according to the intensity of the reporter activity. Deep sequencing of each individual segment, permit correlating the specific β-catenin activity range to the precise mutation the cells carried (Fig 4-1). This chapter will discuss the experimental set up and the analysis of the saturation editing.

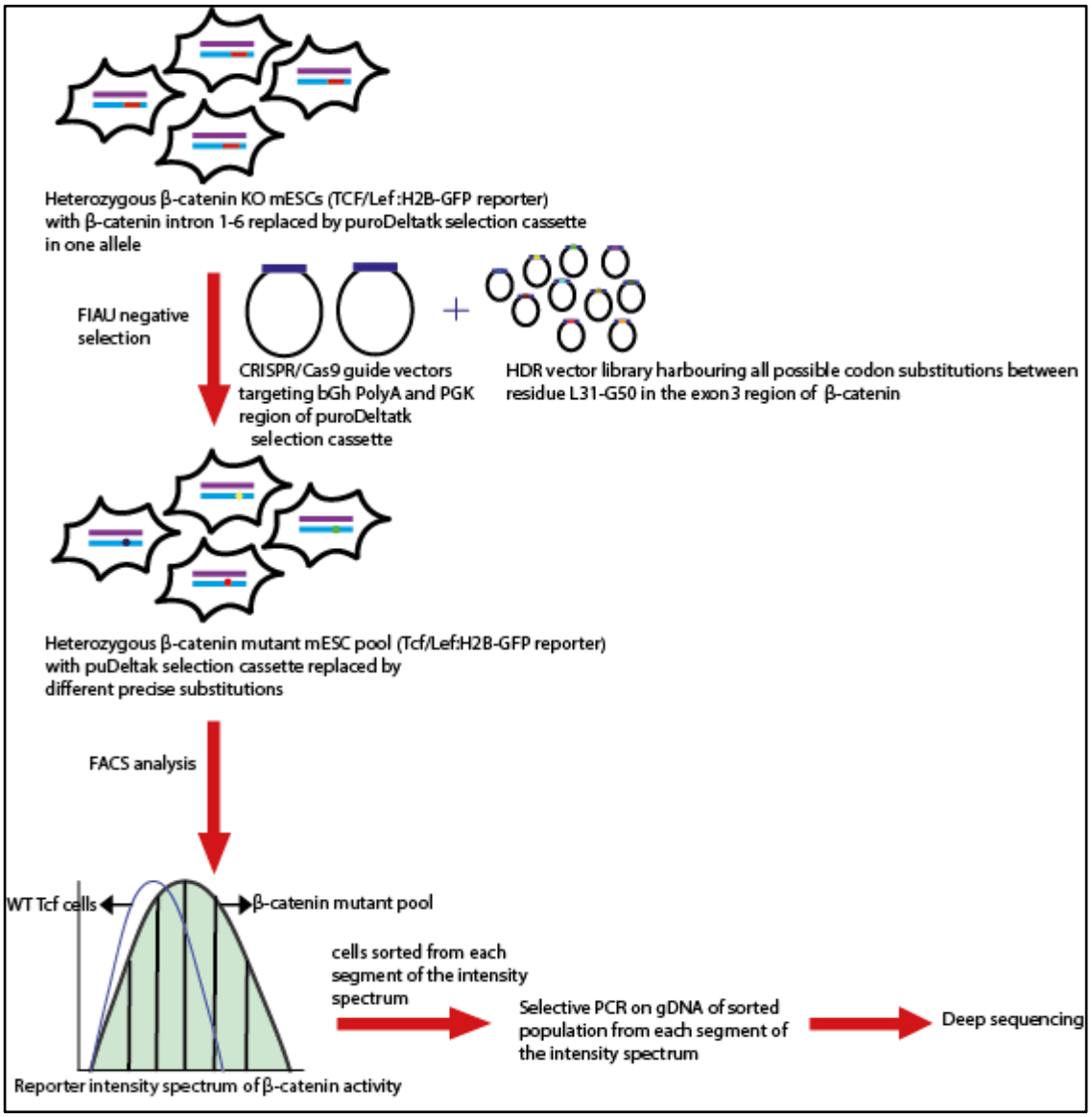


Figure 4-1: Schematic representation of the experimental design of saturation editing assay. CRISPR/Cas9 system coupled with saturation editing to be used to induce all possible mutations at the mutational hotspot of exon 3 region of β-catenin.

4.2 Results

4.2.1 Cloning of saturation HDR vector library

4.2.1.1 Optimization of various cloning strategies

As shown in the previous chapter, ensuring similar efficiency of HDR at each residue within the saturation region forced the use of plasmid as an HDR template. The need for 400 TVs required a more sophisticated cloning approach than the conventional methods. Golden gate system proved to be a very efficient method for cloning, and involved having a template vector with two type IIS restriction enzyme forming a cloning site. Therefore, the first step was to generate a β-catenin backbone vector with two type IIS restriction enzyme BbsI site around the region to be targeted along with homology arms. The generation of this vector was discussed in chapter 3. In order to find the most efficient way of generating the insert with the correct overhangs for the BbsI sites, we tested 3 strategies.

Firstly, two complementary strands (182bp) with BbsI sites were ordered separately and then phosphorylated, annealed and cloned. This approach was adapted from the sgRNA cloning which worked with 100 percent efficiency. However, this approach failed to yield any transformants, probably due to the length of the strands giving rise to internal base pairing or hairpin loop formation, and not allowing perfect complementary base pairing between the two strands. Secondly, a PCR based approach was attempted whereby ssODNs including a BbsI site were amplified using high fidelity DNA polymerase that was later digested and ligated into the backbone vector. This approach had a cloning efficiency of over 85 percent (Fig 4-2A). However, the most efficient and easiest approach that allowed single step cloning was by ordering double stranded (ds) oligos with BbsI sites that when directly digested along with the destination vector followed by ligation, resulted in 100 percent cloning efficiency (Fig 4-2B). This was the most convenient and efficient approach, therefore we followed this strategy for generating all of the targeting vectors.

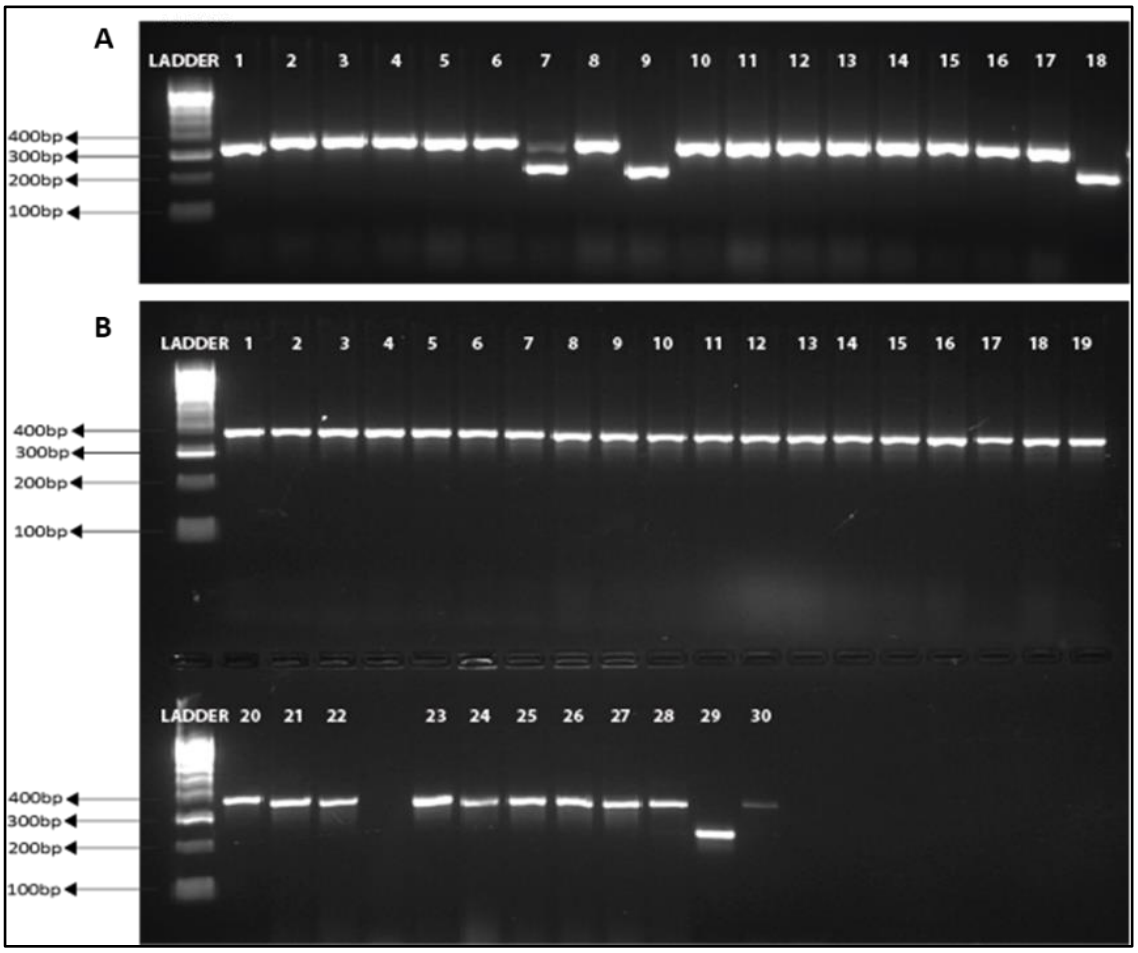


Figure 4-2: Colony PCR of ssODN PCR vs ds oligo based approach optimized for cloning of Saturation TVs. Different approaches were taken for the generation of ds oligos, and tested for cloning (efficiency) into the designation vector. A) Agarose gel electrophoresis image of colony PCR of TVs cloned by generating ds amplicon by oligo PCR based approach yielding an efficiency of over 85%. Except for two clones (lane 7 and 9) all clones (lane 1 to lane 16) show the expected band size. Lane 17 and 18 are + and –controls. B) Agarose gel electrophoresis image of the colony PCR of TVs cloned by ordering synthetically generated ds oligos yielding 100% cloning efficiency. All clones (Lane 1 to lane 28) show an expected band size. Lane 29 and 30 are – and + control.

4.2.1.2 Design of ds oligo

The ds oligo library to generate the TVs was designed so that, it included all the amino acid (20) variants between residues L31 and G50 in the β-catenin hot spot region. Each of the ds oligos also included 3 synonymous mutations downstream of residue G50 that could be used as a handle PCR to be able to amplify only the edited alleles for deep

sequencing. In addition, 2 synonymous mutations were included to disrupt the PAM sequence (of g9B β-catenin guide) and to prevent re-cutting. Although, the plan was to use guides specific for the puromycin selection cassette and keep the WT β-catenin intact, including these mutations gave us the flexibility to use the same TVs in the future and target the WT β-catenin allele to generate hemizygous mutations.

4.2.1.3 Cloning of the ds oligo for generation of TV library

The 400 ds oligo library was synthesized by Twist Biosciences and cloned into the β-catenin backbone vector as a pool using the golden gate strategy, as described previously. Colony PCR revealed a 100 percent cloning efficiency. Individual colonies were picked and analysed by sequencing to confirm the presence of the inserts in the cloned vector (Fig 4-3).

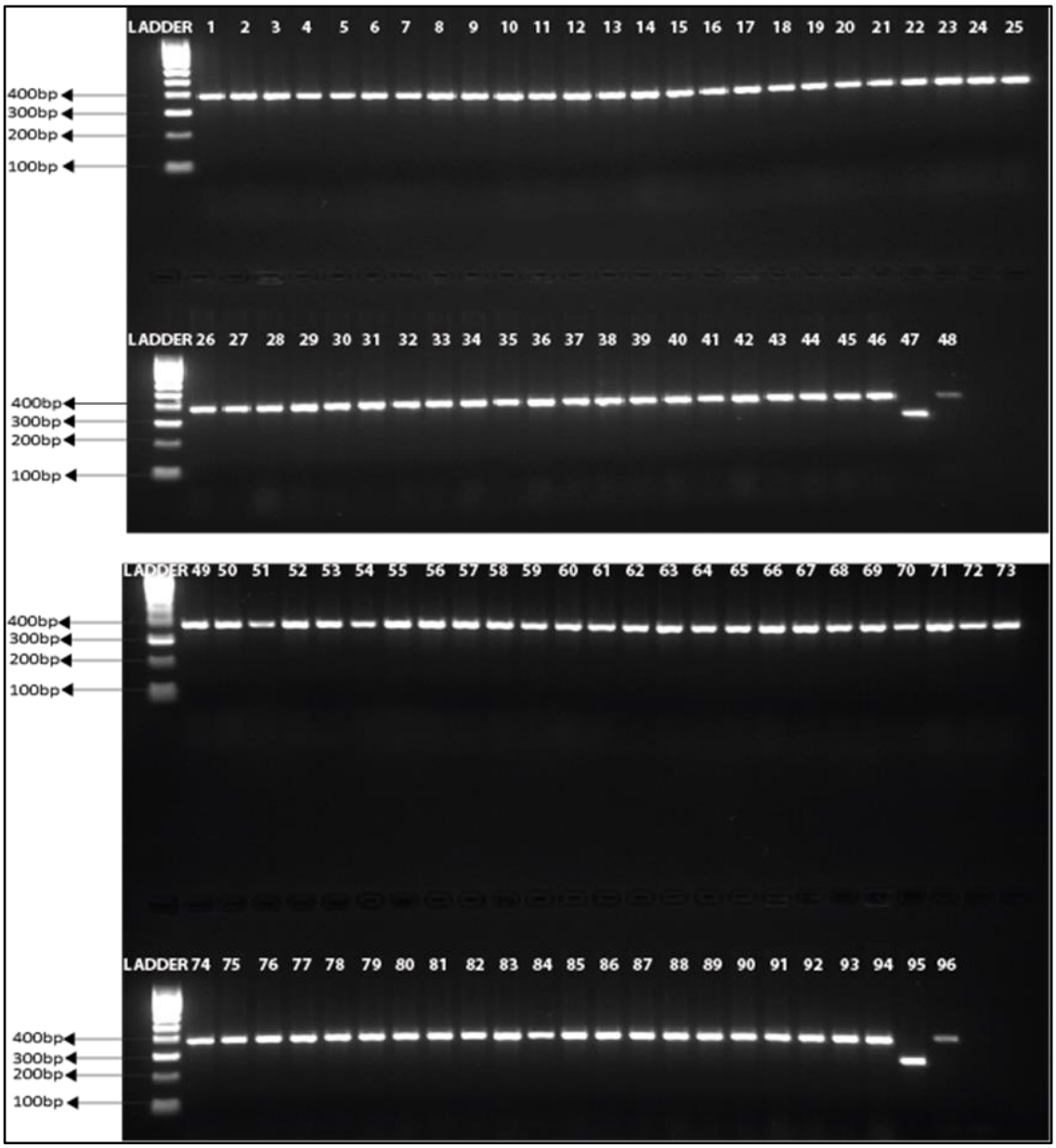


Figure 4-3: Colony PCR image of 100% efficient cloning of Saturation TVs using ds oligo based approach. Agarose gel electrophoresis image of colony PCR of TVs cloned by ordering synthetically generated ds oligos yielding 100 percent cloning efficiency. All clones (Lane 1 to 46 and lane 48 to 94) show the expected band size. Lanes 47 and 95 are – controls and lanes 48 and 96 are + controls. The positive clones were sequenced and TVs with desired mutations were selected for targeting.

4.2.2 Design and cloning of guides targeting puDeltatk selection cassette

For the replacement of the selection cassette with a β-catenin allele harbouring the desired mutation, two guides were designed for each of the 5' and 3' regions of the puDeltatk selection cassette (Fig 4-4). The guides were designed within the selection cassette, or spanning the boundary between β-catenin and the selection cassette, such that they would induce a DSB only within the knock out allele, and hence keeping the WT β-catenin allele intact. The designed guides were cloned into mCherry pX458 and the insertion was confirmed by sequencing.

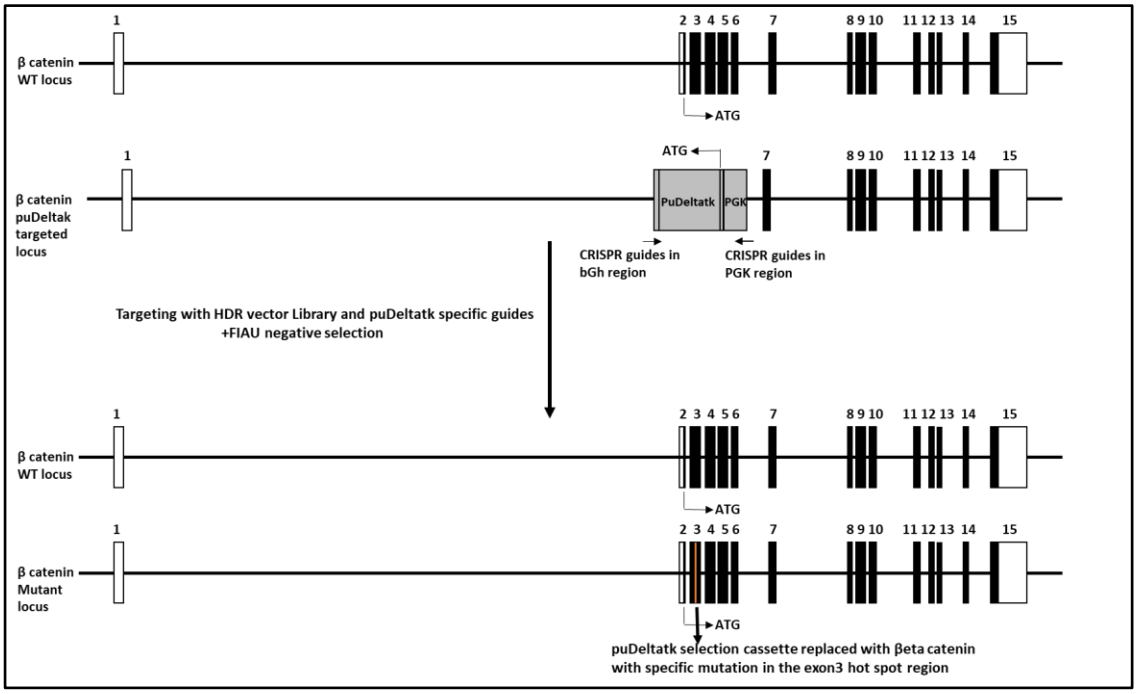


Figure 4-4: Schematic representation of generation of heterozygous β-catenin mutant cell lines. Guides were specifically designed targeting bGh polyA and PGK regions of the puDeltatk selection cassette. The β-catenin heterozygous KO cell line was targeted with these guides along with the HDR vector library followed by treatment with FIAU negative selection. This allowed the replacement of puDeltatk selection cassette with β-catenin with specific mutation in the exon 3 hot spot region, providing a successful strategy for the generation of β-catenin heterozygous mutant cell lines.

4.2.3 Optimizing PCR strategies for deep sequencing

Deep sequencing by the MiSeq sequencing platform required the amplicon size to be around 200-300bp. However, to avoid amplification of random integration of the TV, a two-step PCR approach was taken. First PCR was a long range PCR, with forward primer being in the β-catenin sequence outside the homology arm, and the reverse primer spanning the 3 nucleotide substitution downstream of residue G50 I placed to serve as a mutant allele specific amplification. This 3.2 kb amplicon was then used as a template to amplify the shorter second PCR product, which included the barcodes for the MiSeq sequencing. Various strategies were incorporated to avoid false amplification. For the purpose of optimization, a mock transfection was performed using the CRISPR guides and TV library. As negative controls, untransfected cells and cells transfected only with the TV library (excluding CRISPR guides) was used to detect any false amplification. The genomic DNA isolated from all three transfections was used to amplify 3.2 kb PCR product. To avoid the contamination with any residual TV library, which might have been still present in the cells, the PCR product was gel eluted. This product was then used as a template for the second shorter PCR. The first PCR worked as expected, giving a band only in the cells transfected with the guides and the library. Even though there were no visible bands, the gel elution was performed on the negative controls as well, using the correct band size as guidance. Although I expected a band only in the samples transfected with CRISPR guides and TV library, the shorter PCR yielded false positive amplicons in the library-only and the untransfected controls, in several attempts. I assumed this could only be a result of general DNA contamination, as the TV library was considerably larger in size than 3.2 kb where the gel elution was performed. As even very small amounts of DNA carried over either from buffers in gel tanks or other sources, would result in false amplification, I decided to run each sample in a different tank. The gel tanks, combs, and the gel trays were all soaked in 5M NaOH O/N, washed thoroughly to remove any residual DNA next day, and filled with fresh buffer. The gel eluted PCR product was also treated with DpnI restriction enzyme to remove any residual TV coming with the buffer, and in addition, all PCRs were performed in T/C hoods. As a result, the false positive amplification was visibly reduced and I was able to reduce amplification from contaminating DNA to below the detection level, ensuring specificity of PCR (Fig 4-5).

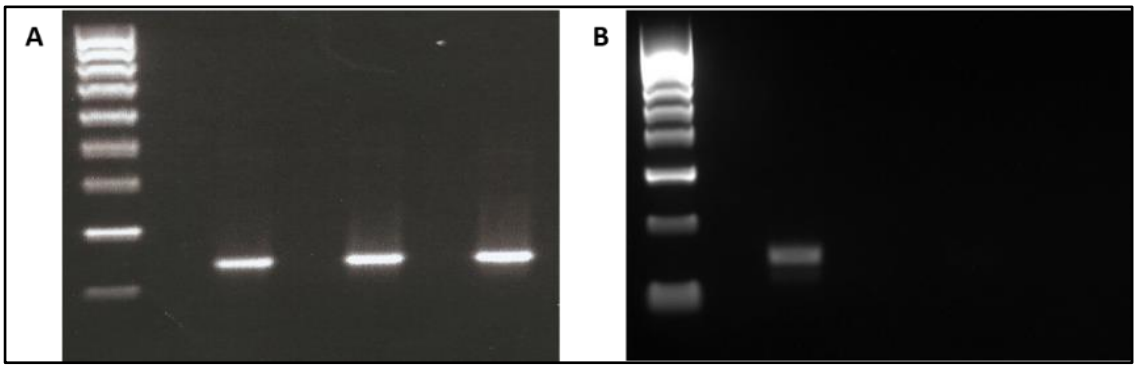


Figure 4-5: PCR optimization strategies to overcome false positive amplification. Agarose gel image of the second PCR performed on sample transfected with TV and CRISPR guide, TV only transfected and untransfected control. A) PCR products all run in the same gel- Strong PCR bands observed in all three samples B) PCR products run in separate cleaned tanks- Specific amplification observed only from DNA transfected with TV and CRISPR guide.

4.2.4 Saturation Assay

4.2.4.1 Selection of time frame for culturing TCF reporter cells in normal ES media

The TCF reporter cells could only be cultured in normal media for a short period, and prolonged passaging in normal media resulted in a change in morphology and proliferation. Therefore, I had to maintain the cells in R2i media which contains a GSK inhibitor activating the Wnt pathway. As this would interfere with phenotyping the mutant cells, I tested how long it takes for β-catenin activity to reduce after the removal of R2i media. As shown in the flow cytometry analysis in figure 4-6, there was considerable reduction in the levels of β-catenin activity even after overnight culture in normal media. I observed that culturing the TCF reporter cells in normal ES media for a short duration of 2-3 days had little effect on the morphology and they could still be maintained in an undifferentiated state. Based on these observations I decided that culturing the reporter cells for 2-3 days was optimal for both reduction of β-catenin activity levels and maintaining their morphology.

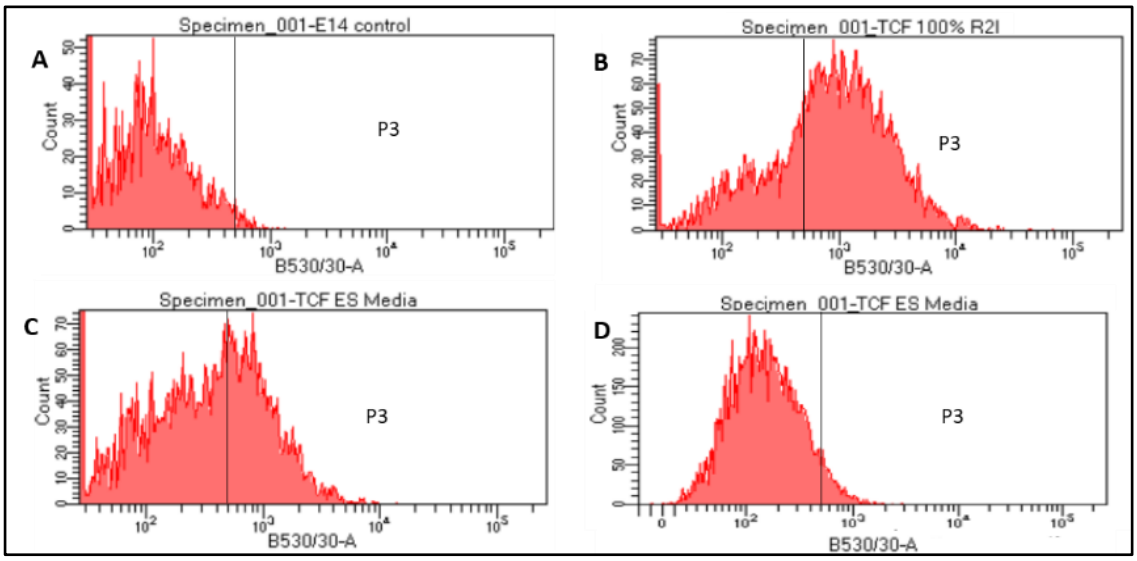


Figure 4-6: Flow cytometry analysis of β-catenin activity of TCF cells cultured in normal ES media. (A) E14 cells used as control to gate the GFP negative population. (B) TCF reporter cells cultured in R2i media. (C) TCF cells cultured overnight in normal ES media. (D) TCF cells cultured in ES media for three days. TCF cells cultured overnight in normal ES media already showed a considerable reduction β-catenin activity levels compared to the TCF cells cultured in R2i mediand culturing in normal ES media for three days resulted in a further reduction of β-catenin activity.

4.2.4.2 Saturation editing and FACS sorting

On day1, 200X10^6 β-catenin heterozygous KO TCF cells were transfected with the HDR vector library and CRISPR guides targeting the selection cassette. On the third day post transfection, the R2i media was replaced with normal ES media supplemented with FIAU for negative selection. A control plate with untransfected β-catenin KO TCF cells was also placed on FIAU negative selection, to ensure the cell death took place before I started the FACS analysis. Flow cytometric analysis on the fifth day post transfection showed a shift in the intensity spectrum of the saturation edited mutant pool in comparison with the untransfected TCF WT and β-catenin heterozygous KO TCF control cell lines (Fig 4-7). Following flow cytometric analysis, the saturation edited mutant pool was divided into 7 segments of varying intensity, and cells were sorted from each segment.

The sorting strategy of the saturation edited mutant pool was carefully designed based on the various mock experiments, and also the previous functional analysis using the S33Y and ΔS45 mutants (discussed in chapter 3 section x). Using E14 as control, I set

the P1 and P2 gates to sort the GFP negative population. Next, considering that many of the mutant population may be similar to the WT GFP signal, the P3 gate was set, which was the range of GFP occupied by majority of WT TCF population. Many mutants might be capable of increasing the β-catenin activity levels, however, as seen in our preliminary experiment with ΔS45 and S33Y, the intensity of signal may vary among the different mutants (refer chapter 3 Figure 3-28A). The S45 heterozygous mutant pool resulted in lower and medium levels of increase in β-catenin activity, taking this into consideration, I set the P4 and P5 gate to represent this range of activity. Finally, based on the observation that the S33Y heterozygous mutant population mainly selected for higher increase in activity, I set the P6 and P7 gates to sort for cells with higher increase in activity levels.

However, due to technical reasons the cells from P1 gate could not be sorted, and hence cells were sorted from the remaining six gates (P2_1 to P7_1).

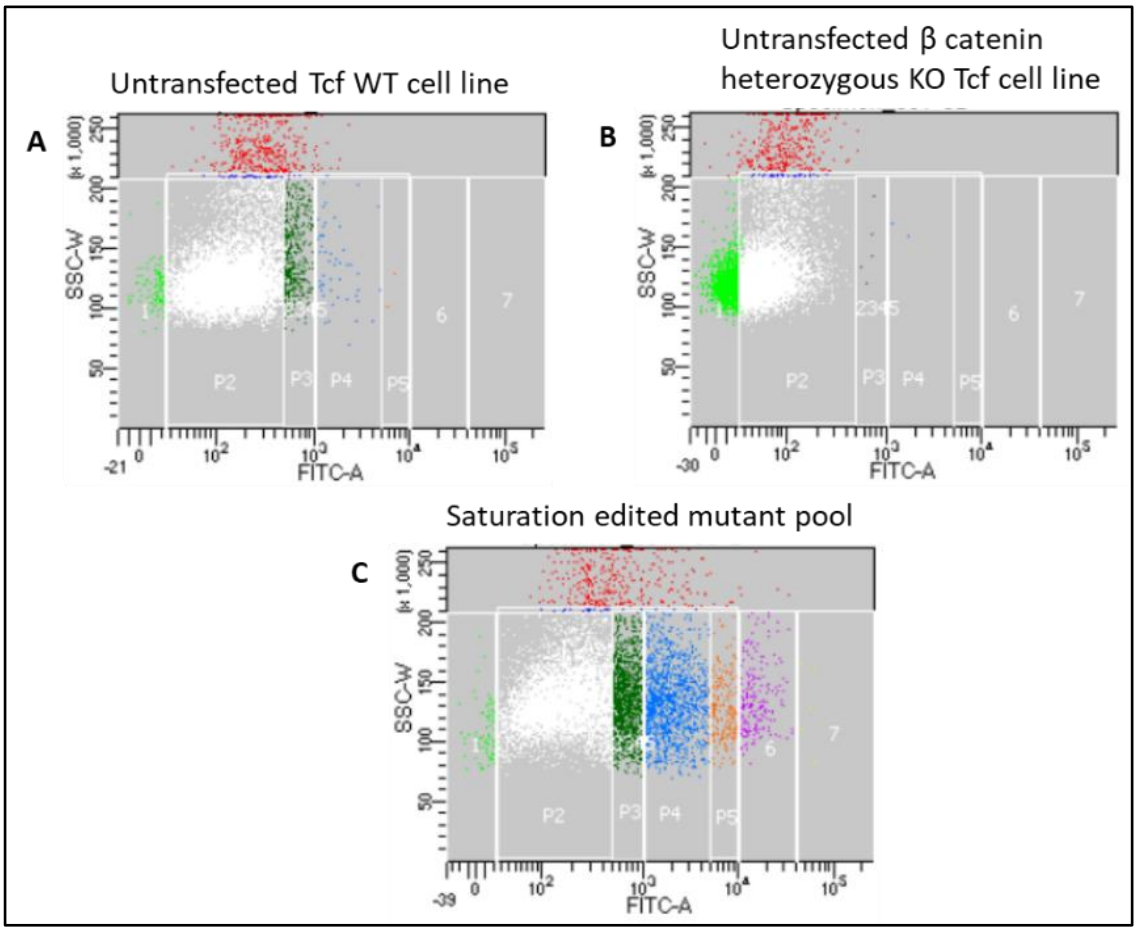


Figure 4-7: Flow cytometric analysis for sorting cells from different intensity segments of the saturation edited mutant pool. (A)Flow cytrometric analysis of untransfected TCF WT cell line. (B) Flow cytrometric analysis of untransfected β-catenin heterozygous KO TCF cell line (C) Flow cytrometric analysis of saturation edited mutant pool. Flow cytometric analysis shows a shift in GFP expression in the saturation edited mutant pool (on day 5) in comparison with the untransfected WT TCF and β-catenin heterozygous KO TCF cell lines. Cells from each of the segments P2-P7 were sorted for genotypic assessment by deep sequencing.

To be able to account for the variation in the experimental set up, the experiment was repeated a second time independently using the above described protocol and parameters, and cells were similarly sorted from the 6 gates (P2_2 to P2_7). In addition to the 6 sorted gates, a small sample of cells (pool) was collected before sorting from both experiments (pool_1 and pool_2). The cell numbers sorted from both the experiments are given in table 4-1.

Segment	Replicate1 cell numbers	Replicate2 cell numbers
P2	200,000	200,000
P3	200,000	200,000
P4	200,000	200,000
P5	80,000	1,130,000
P6	445,427	942,000
P7	8,309	11,641

Table 4-1: Number of cells sorted from different segments of replicate1 and replicate2

4.2.4.3 DNA processing and Deep-sequencing

Following cell sorting, the DNA was immediately extracted using Qiagen DNeasy kit. Taking the precautionary measures to avoid false amplification as detailed in section 4.2.3, the PCRs were performed. Briefly, the first round 3.2 Kb PCR (with handle specific R primer and F primer outside the homology arm) was gel eluted, followed by DpnI digestion. The DpnI digested first PCR was used as template to perform a second round of handle specific PCR with Illumina barcoded primers. These PCRs were performed for all the 14 samples (samples P2-P7 and the pooled sample from both replicates).

As the ds oligos to generate the targeting vectors was received and cloned as a pool, it was important to sequence this library alone itself, to see what the plasmid distribution within the library was. For this purpose the TV library used in transfections of the independent replicates (plasmid_1 and plasmid_2) was used as template and the short handle specific PCR was performed using the Illumina barcoded primers.

Following amplification, all 16 samples were quantitated using qubit, the samples were then pooled in equimolar quantity and finally the integrity of the DNA pool was assessed using a bioanalyser. The pooled library was submitted to Edinburgh genomics, and a paired end sequencing (200bp read1 and 100bp read2) was performed using the Illumina Miseq sequencing platform.

4.2.4.4 Processing and analysis of deep sequencing data

The processing of raw sequencing data and analysis was done by Martijn Kelder PhD student in Andrew Wood's lab, IGMM, University of Edinburgh. Initially, a quick quality check of the raw sequencing data (FastQ format) was performed using FastQC, which gives an overview of various parameters including per sequence quality score, sequence length distribution, total sequences etc for each of the sequenced samples. Next, as the forward and reverse reads were respectively, 200nt and 100nt in length, the forward reads were stripped down to 100nt using Python.

Reads were then trimmed for adapter sequences (<1 percent of reads) using Trimgalore and aligned to the reference sequence using BowTie2. As the amplicons were designed to specifically amplify the HDR edited cells, the handle and PAM mutations being common in all the 16 samples, a β-catenin WT sequence (GRCh38) with handle and PAM mutation was used as a reference to map the reads. The mapping of the reads from each of the 16 samples to the β-catenin reference sequence showed a very high percentage of alignment (Table 4-2), with over 98-99 percent of reads from each of the samples being correctly aligned to the reference sequence. These results not only indicated a good sequencing run, but also shows how good the quality of the library that we provided was, and highlights the strength of our design approach and robustness of the implemented strategies in determining the quality of the project outcome.

Following the alignment of the reads, the forward and reverse reads were merged into a single 100nt contig using Python. Only contigs without indels and with consensus between forward and reverse read for each position were passed on for further analysis. Finally, the amino acid substitutions were counted for each saturated codon using Python.

Sample	Number of aligned pairs	Percentage of aligned pairs
p2_1	709521	98.74%
p3_1	733891	98.73%
p4_1	729201	98.75%
p5_1	613649	98.89%
p6_1	640743	98.90%
p7_1	585110	99.17%
Plasmid_1	658587	98.69%
pool_2	736914	98.64%
p2_2	851189	98.28%
p3_2	679890	98.74%
p4_2	728792	98.58%
p5_2	695346	98.96%
p6_2	604142	98.94%
p7_2	582289	99.11%
Plasmid_2	596072	98.71%
pool_1	789250	98.71%

Table 4-2: The number and percentage of aligned pairs of deep sequencing data for each of the 18 samples.

4.2.4.5 Correlation between replicates

As this experiment was repeated twice to be able to account for any variation in the experimental set up, we tested how these two replicates correlated for each segment. A very high correlation was observed between the replicates, with most of them having R values in the range of 0.7-0.99 (Fig 4-8). The very low measurable variation between the replicates confirmed that these results were highly reproducible. The slightly lower R value (0.653) for the replicates from the P7 gate was probably due to the lower cell numbers, however, the correlation was still high enough to draw conclusive evidence of the biological effect of the observed mutants in this segment.

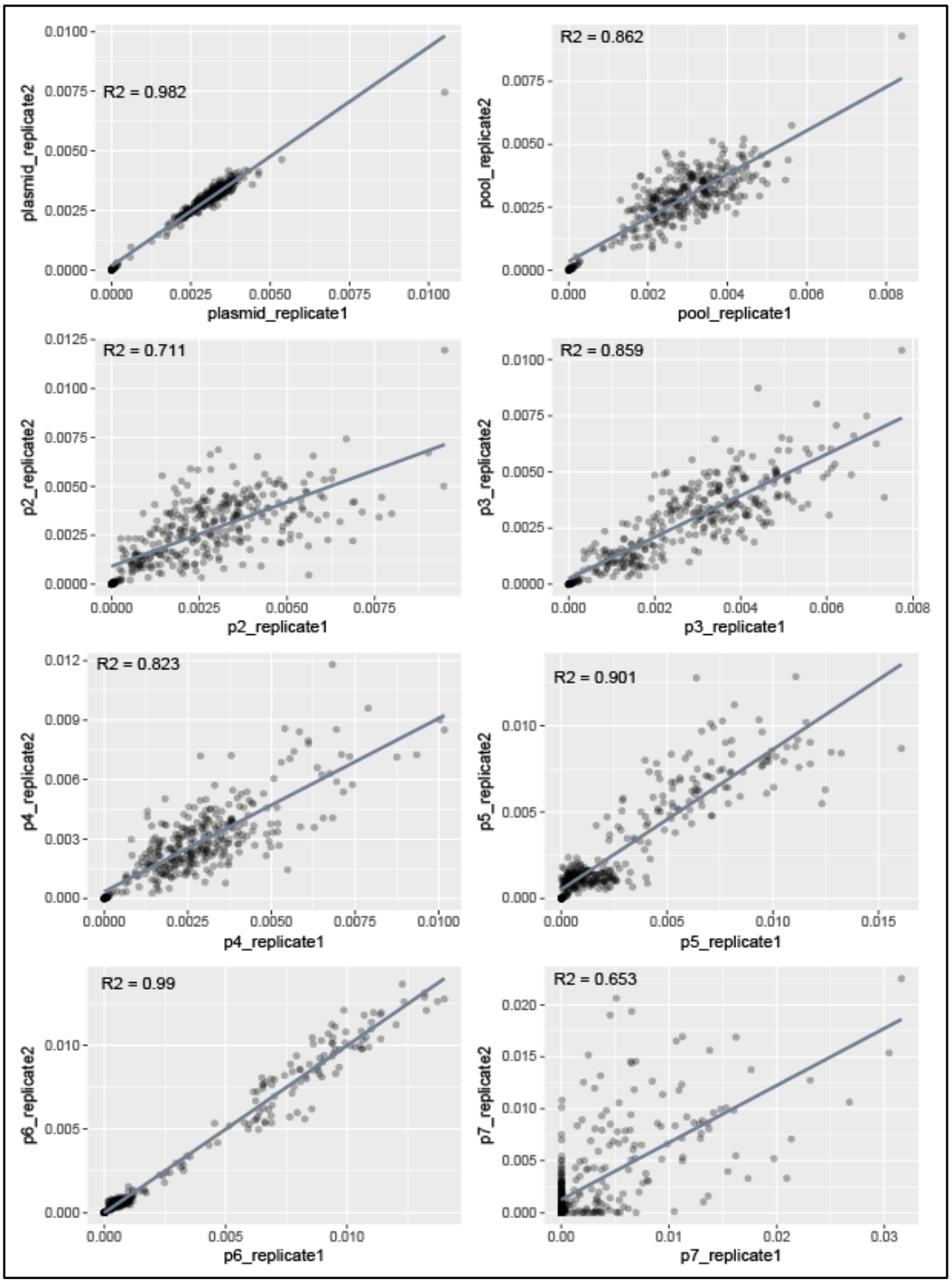
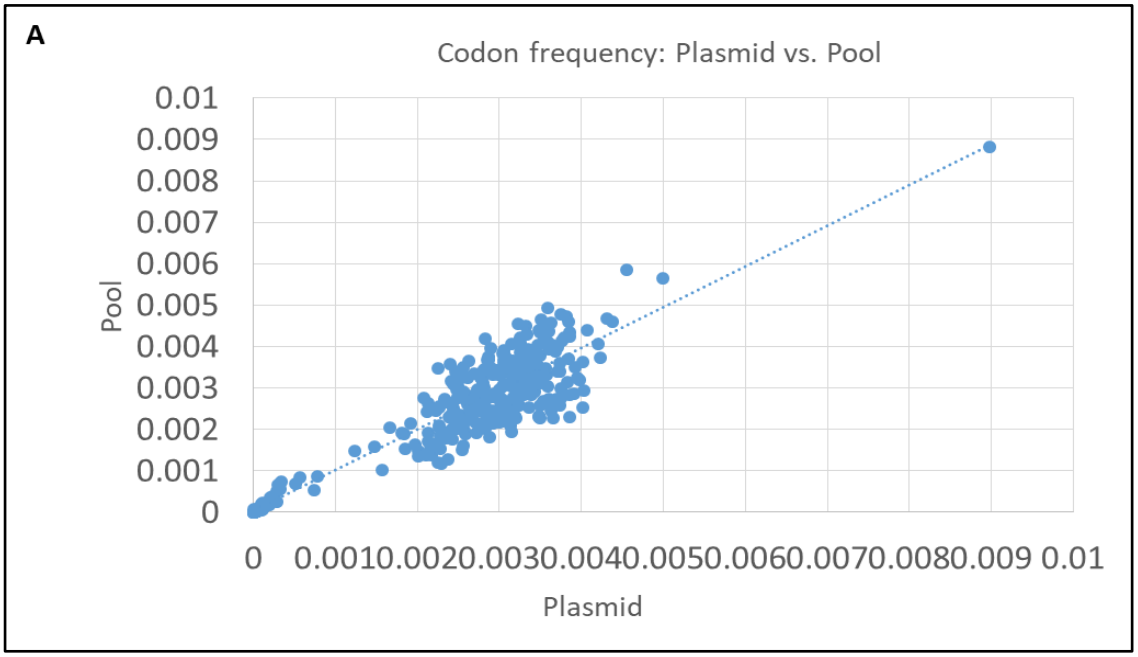


Figure 4-8: Correlation between replicates. A correlation analysis was performed for comparing the reproducibility between the samples P2 to P7 from replicates 1 and 2.

A very high correlation was obtained between the pool and plasmid library (R=0.93), confirming no bias in the rate of incorporation of the different variants (Fig 4-9). In addition, the histograms of the amino acid variants across L31 to G50 from the pooled sample shows a similar profile as that of the plasmid library, validating the efficient and equally weighted incorporation of the substituents, reflecting an un-biased HDR editing rate across the target site. As seen in Figure 4-9B and C, there was good representation of each residue in the plasmid library except for K49 and G50, similar to that observed in the pool sample. This could be an artifact the way the DNA synthesis was done by Twist and because of this low incorporation, these 2 residues were later removed from the analysis. The substitutions within each residue also seemed to be well distributed and comparable among both the plasmid and pool samples.



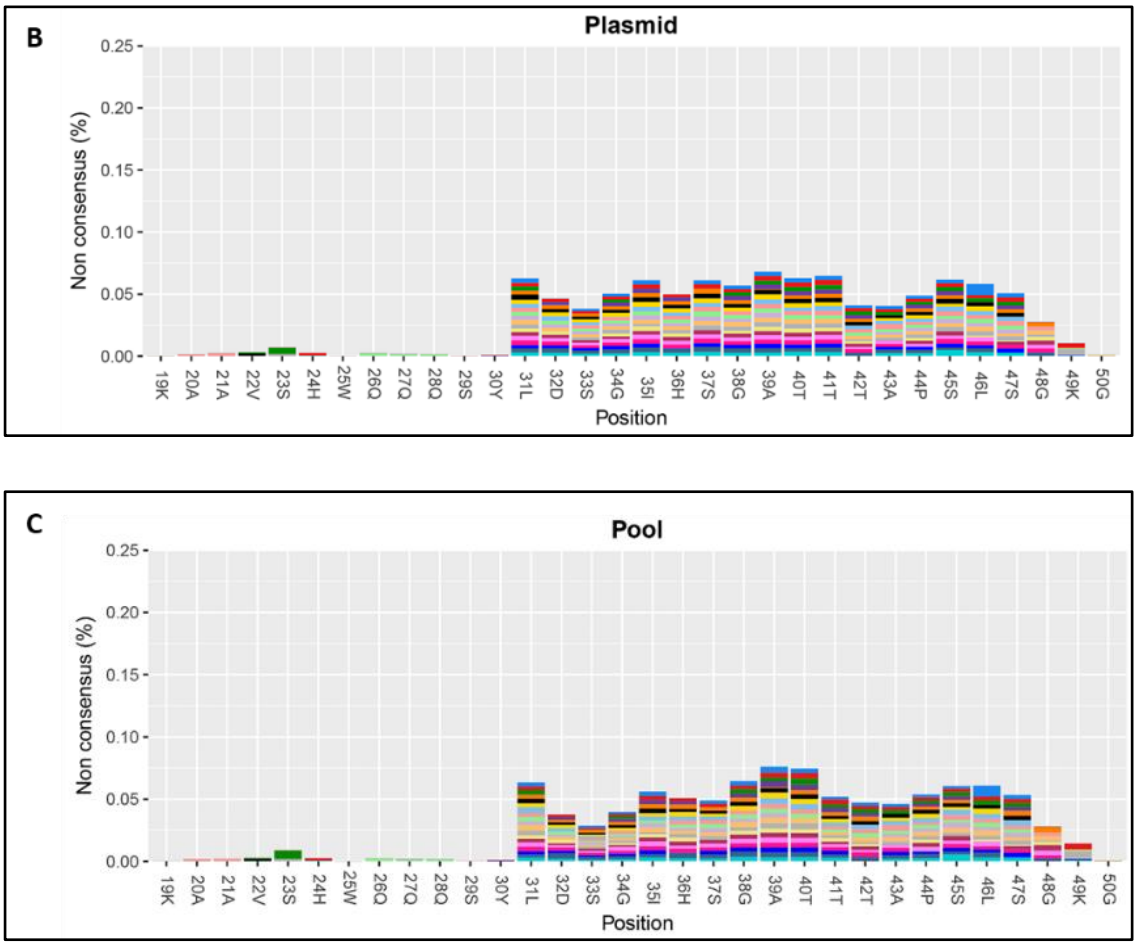


Figure 4-9: Comparison between pool and plasmid sample. (A) Correlation plot of plasmid vs pool and (B) Histogram of proportion of amino acid variants across the target site in the plasmid sample (C) Histogram of the proportion of the incorporated amino acid variants across the target site in the pool sample

4.2.4.6 Combined overview of P2-P7 from both replicates normalized to pool

Although each residue and each substitution were well represented in both the library and the pool, there was still some variation. In order to eliminate the effect of this variation from our analysis, I decided to normalize the values from each gate to the pool. This way, the differences we found between each mutation could not be due to being under/overrepresented in the plasmid or any small difference in HDR efficiency across the region.

As the replicates for each segment showed good correlation, the normalized values of the two independent replicates were combined for further analysis.

Comparison of the line graphs between the P2 (segment with the lowest activity) and P7 (segment with the highest activity) populations reveals the variation in the activity levels by the different residues (Fig 4-10). The most frequently mutated residues from the COSMIC database analysis including the phosphoylatable S and T residues 33,37,41,45 and residues D32 and G34, showed clear difference in the activity levels across the segments. While the rest of the amino acid variants across the target site were largely present in the lower intensity segments P2-P4 which also represents the activity range of the WT TCF cells.

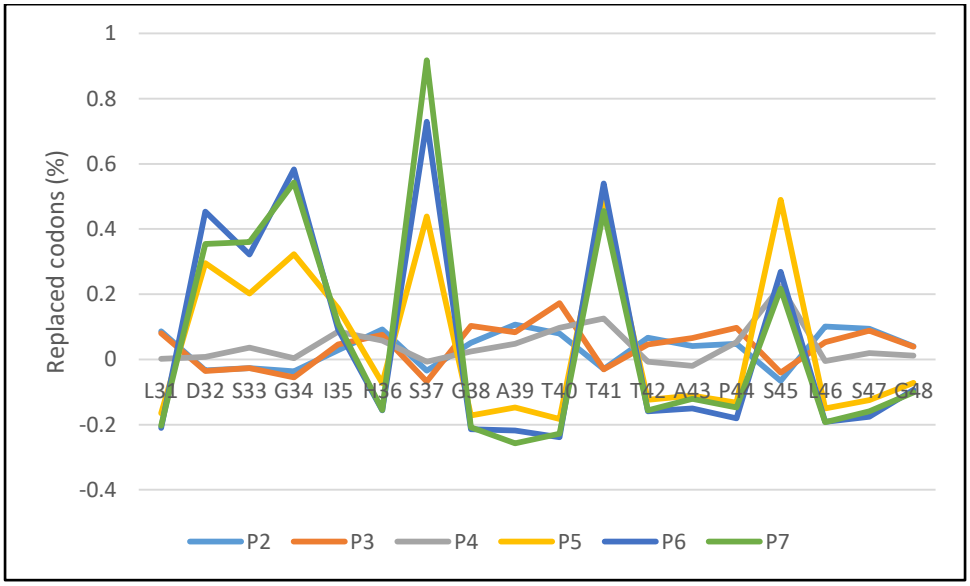


Figure 4-10: Line graph of the overall β-catenin activity of residues L31-G48 across the different segments of the sorted population.

As expected the top six residues were underrepresented in the low expressing segments (P2 and P3), and were prominently present in the high expressing segments (P6 and P7) (Fig 4-11). Furthermore, the proportion of the substituted amino acid variants differed across the segments for each of these 6 residues, which is the first sign of each residue having a different β-catenin activity (Fig 4-12). While in the P4 segment, the number of total S45 mutants were significantly higher than S37, the opposite trend was observed in the P7 segment. This shows that S37 mutation results in higher β-catenin activity than

S45 residue. A trend similar to that of S45 was observed for the T41 variants, which could also be seen increasing in the P4 segment. The highest proportion of S45 variants were present in the P5 segments and decreases in P6 and P7, with only a handful of S45 variants being selected for in the P7 segment. The proportion of T41 variants increase through P4 and P5, with the highest levels in P6 and reducing in P7. A linear increase in the proportion of S37 variants can be observed, with the highest proportion in the P7 segment and with majority of the variants capable of increased activity. The S33 variants follow a similar trend as S37, with a linear increase in proportion, however the overall proportion of S33 variants was considerably lower than that of S37. Further, the activity of D32 and G34 residues increase in the P5 and P6 segments and dips in the P7 segment. In addition to D32 and G34, a similar increase in the proportion of amino acid variants was observed for residue I35, especially in the P5 segment, and slightly reducing in the P6 and P7 segments. This enrichment in specific amino acid variants of residues across the hot spot region in different segments of the sorted population indicates the capability of different mutants to confer different activity levels.

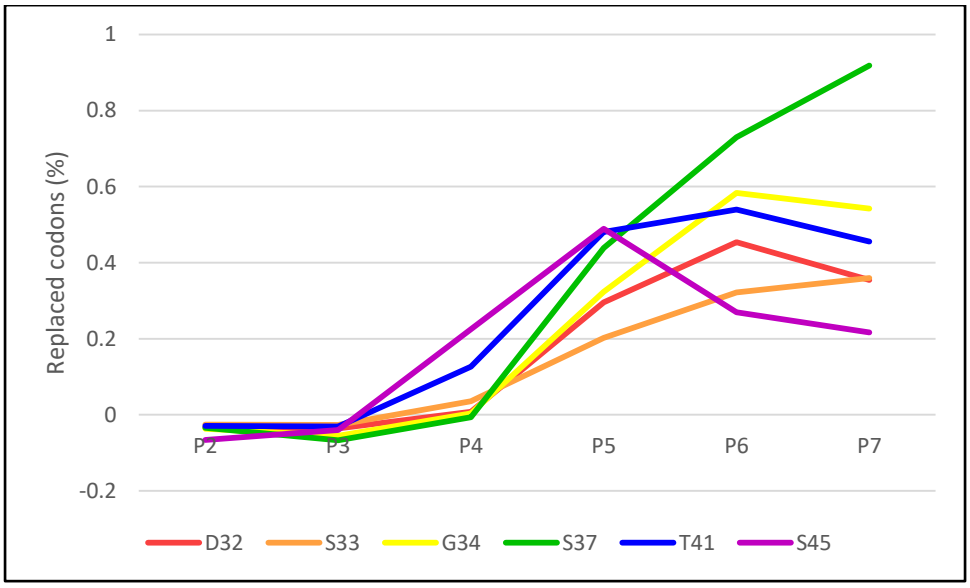
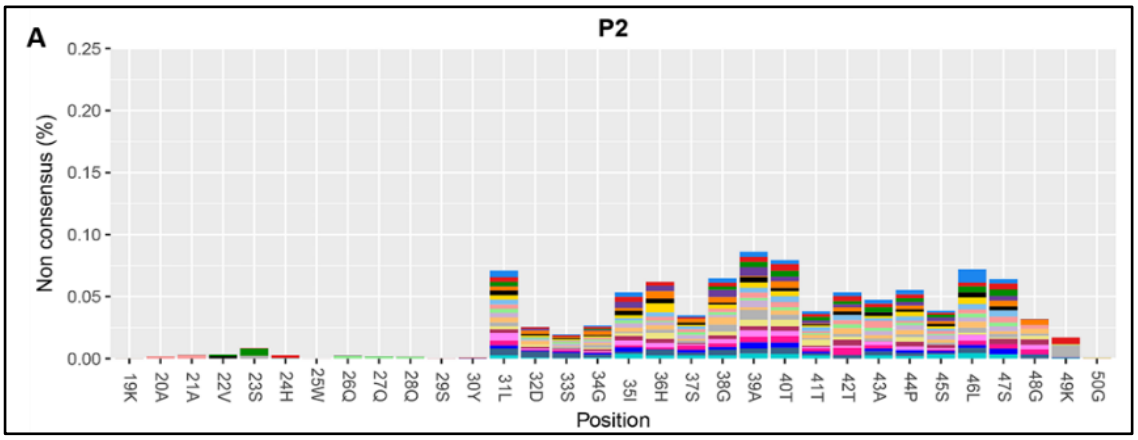
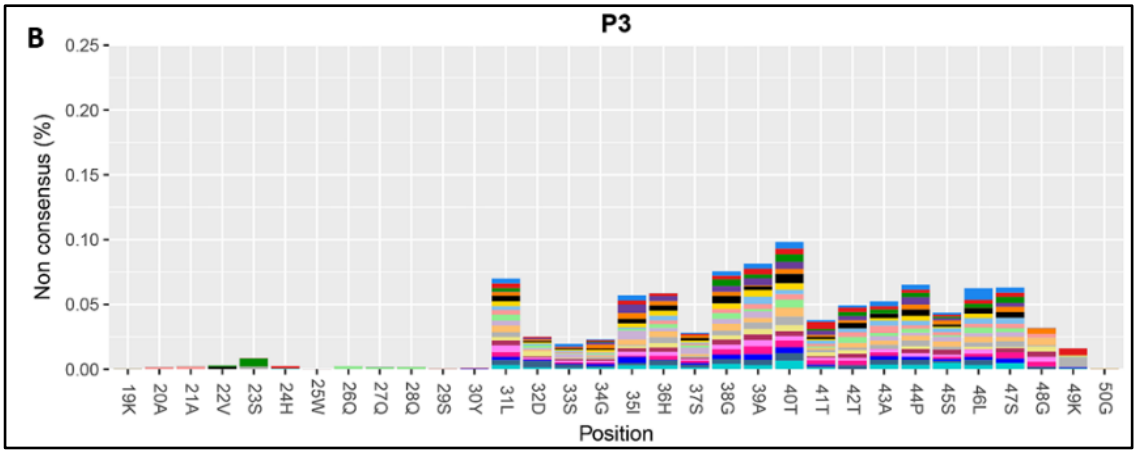
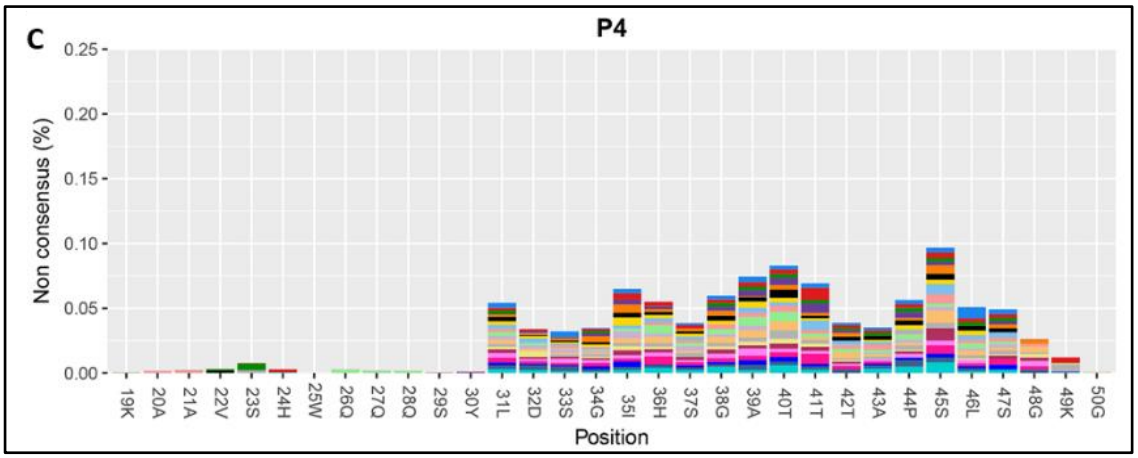


Figure 4-11: Line graph of the overall β-catenin activity of residues D32 S33 G34 S37 T41 and S45 across the different segments of the sorted population.


A
P2
Non consensus (%)
0.25
0.20
0.15
0.10
0.05
0.00
19K 20A 21A 22V 23S 24H 25W 26Q 27Q 28Q 29S 30Y 31L 32D 33S 34G 35I 36H 37S 38G 39A 40T 41T 42T 43A 44P 45S 46L 47S 48G 49K 50G
Position


B
P3
Non consensus (%)
0.25
0.20
0.15
0.10
0.05
0.00
19K 20A 21A 22V 23S 24H 25W 26Q 27Q 28Q 29S 30Y 31L 32D 33S 34G 35I 36H 37S 38G 39A 40T 41T 42T 43A 44P 45S 46L 47S 48G 49K 50G
Position


C
P4
Non consensus (%)
0.25
0.20
0.15
0.10
0.05
0.00
19K 20A 21A 22V 23S 24H 25W 26Q 27Q 28Q 29S 30Y 31L 32D 33S 34G 35I 36H 37S 38G 39A 40T 41T 42T 43A 44P 45S 46L 47S 48G 49K 50G
Position

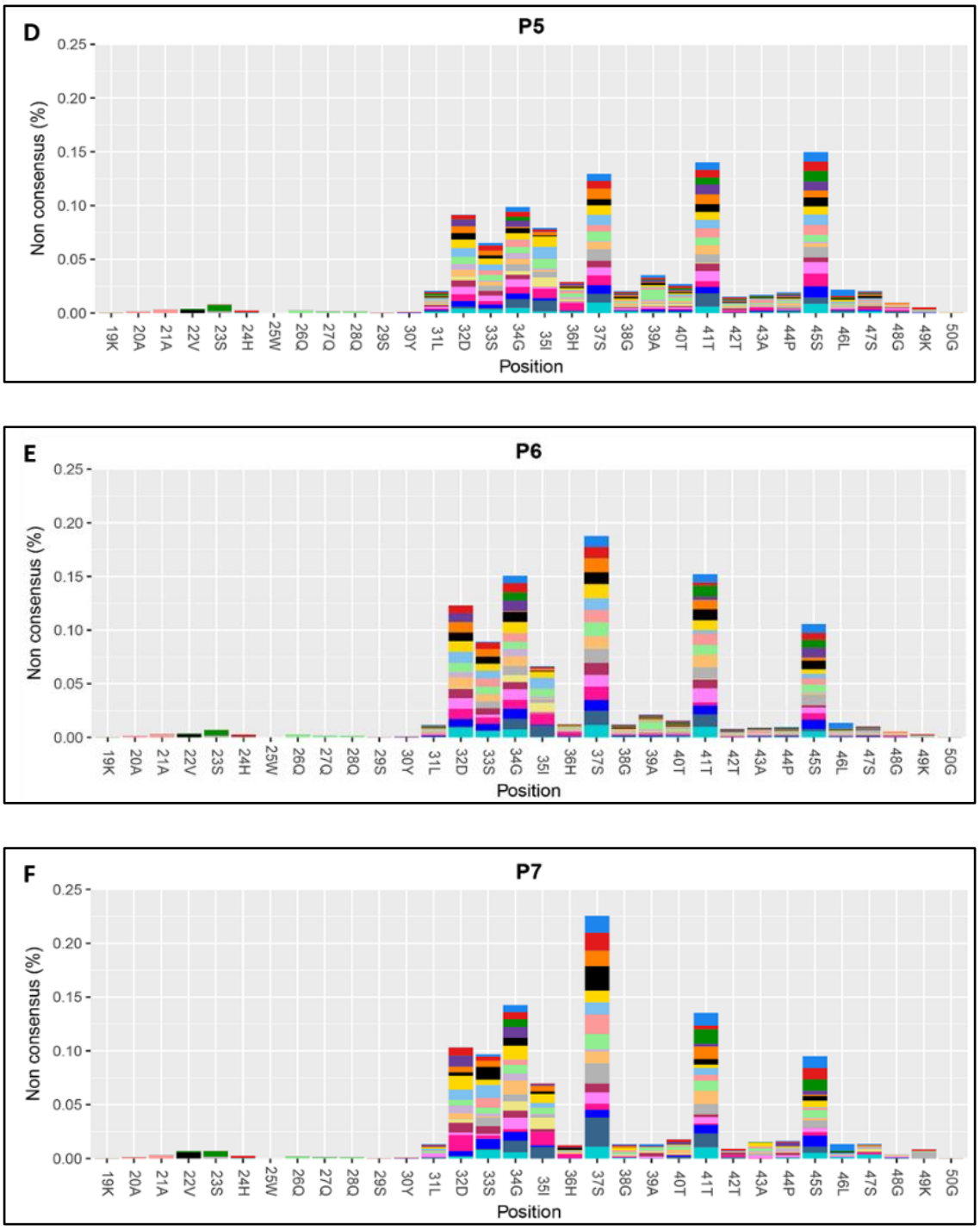


Figure 4-12: Analysis of the different segments of sorted population. (A-F) Histograms of the segments P2-P7 representing the proportion of the amino acid variants for each of the residues across the hotspot region. Each colour block represents a different amino acid variation.

4.2.4.7 Assigned value heat map

To statistically predict whether the activity levels of each amino acid variant (for each of the 18 residues from L31-G48) changes across the six segments of the intensity spectrum, regression analysis was performed. This analysis was done by Helen Brown (senior statistician, The Roslin Institute).

The p values, slope and SE of slope were calculated. Both p value and slope give a measure of the change in activity for each amino acid variant across the six segments. The normalized values of the two independent replicates were separately used to calculate the p values, to test if the observed change is significant. The slope (m) is a measure of the 'rate of change' in activity for each of the amino acid variants across the six segments.

The individual m values of the amino acid variants were assigned as their overall activity across the six segments (hereafter referred to as the mutational effect). A heat map of the mutational effect of the different amino acid variants across L31-G48 residue was plotted (Fig 4-13). The assigned value heat map gives a clear indication of the differential mutational effect among the amino acid variants. The mutational effect of the amino acid variants for of majority of the residues across the target site remain low or unaltered, with enrichment observed especially among the amino acid variants at the top 6 residues. Similar to our previous observations, among the top 6 residues, the highest mutational effect were conferred by the S37 variants and the S45 variants are among those conferring a lower mutational effect. In addition to the top six residues, few of the amino acid variants at the residue I35 also show enrichment in the observed mutational effect. The mutational effect of synonymous substitutions i.e. D32D, S33S, G34G etc, was almost always zero across all the residues, which in itself validated the rate of change of activity scores obtained by regression analysis.

This regression analysis was important, not only in assigning a single score for the β-catenin activity levels for each amino acid variants across the six segment, but also provided a statistically significant validation of the observed differential activity among the different amino acid variants.

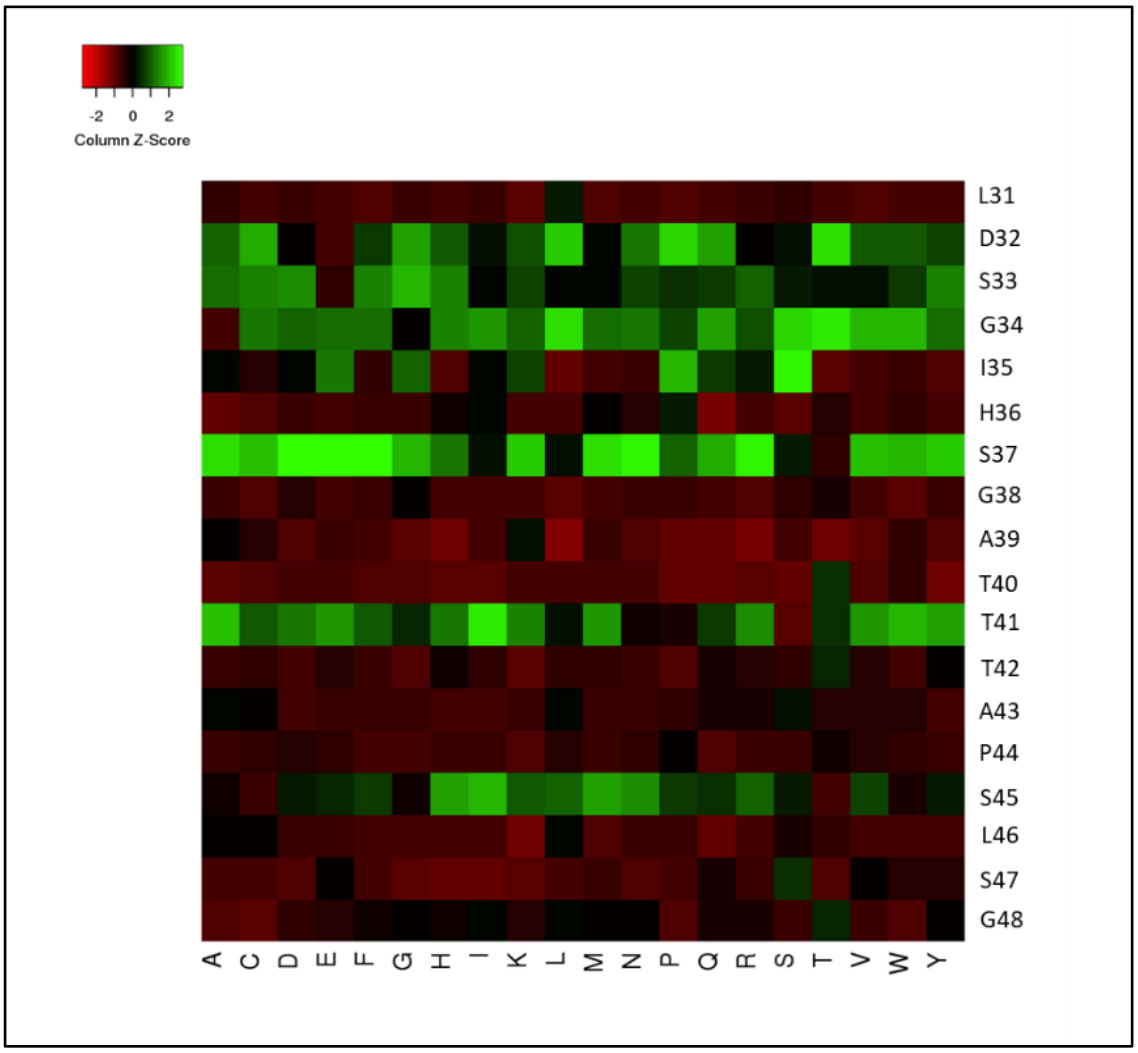


Figure 4-13: Regression analysis of the β-catenin activity across the target region. A heatmap of the mutational effect was plotted to analyse the differential activity among the mutants.

4.2.4.8 Analysis of β-catenin mutational effect for the mutations observed across different cancer types

In Chapter 2, I had tabulated the mutational spectrum for each tumour type using the COSMIC database. In order to establish how the mutations found in a particular cancer type aligned with the mutational scores from the regression analysis, the mutation frequency vs mutational effect graph was plotted for each cancer. These graphs were made by Deepti Vipin in Anagha Joshi's group in The Roslin Institutute.

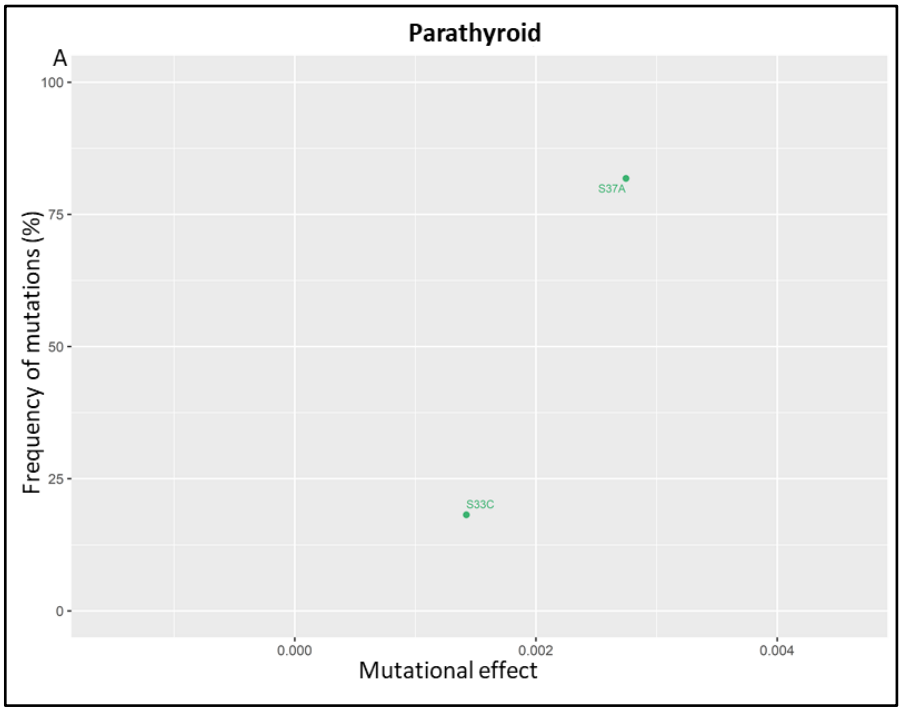
The mutations observed in the tumours of the parathyroid, prostrate, testis, urinary tract and oesophagus, specifically select for medium to higher range of mutational effect (Fig 4-14 A-E). In none of these tumour types are there any mutations with lower effect, which suggests a preferential requirement an increased β-catenin activity among these tumour types.

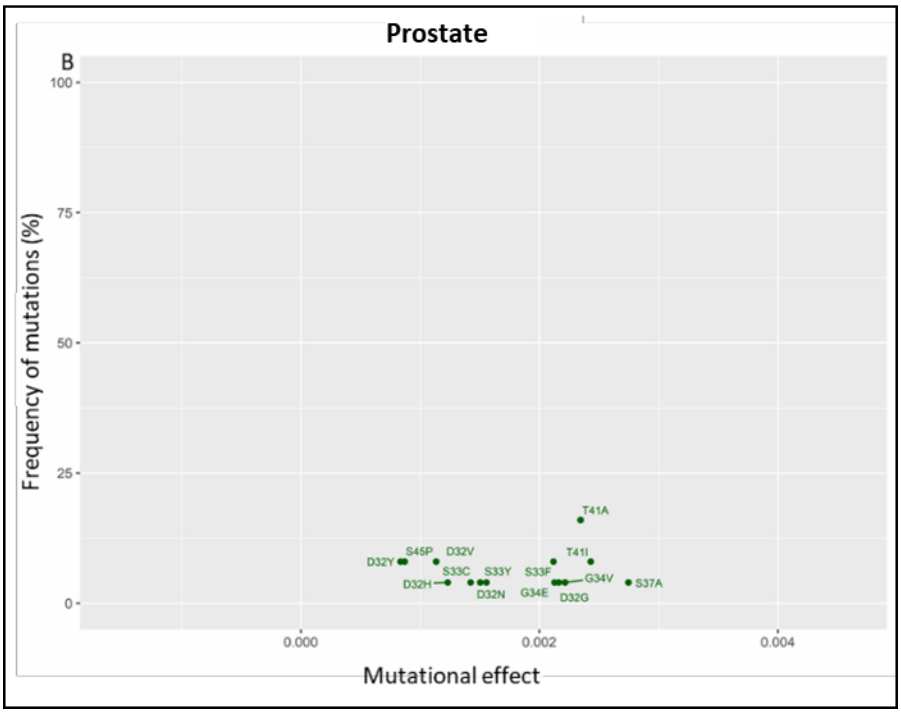
Contrary to the selection of mutations with an increased mutational effect observed in the above tumour types, the mutations selected for in the salivary gland yield a very low mutational effect (Fig 4-14 F). The β-catenin mutation in the salivary gland are confined to two residues, I35 and T41.I35T is the most frequently occurring mutation, followed by a lower incidence of T41P mutations, and both these mutations contribute to a very low mutational effect.

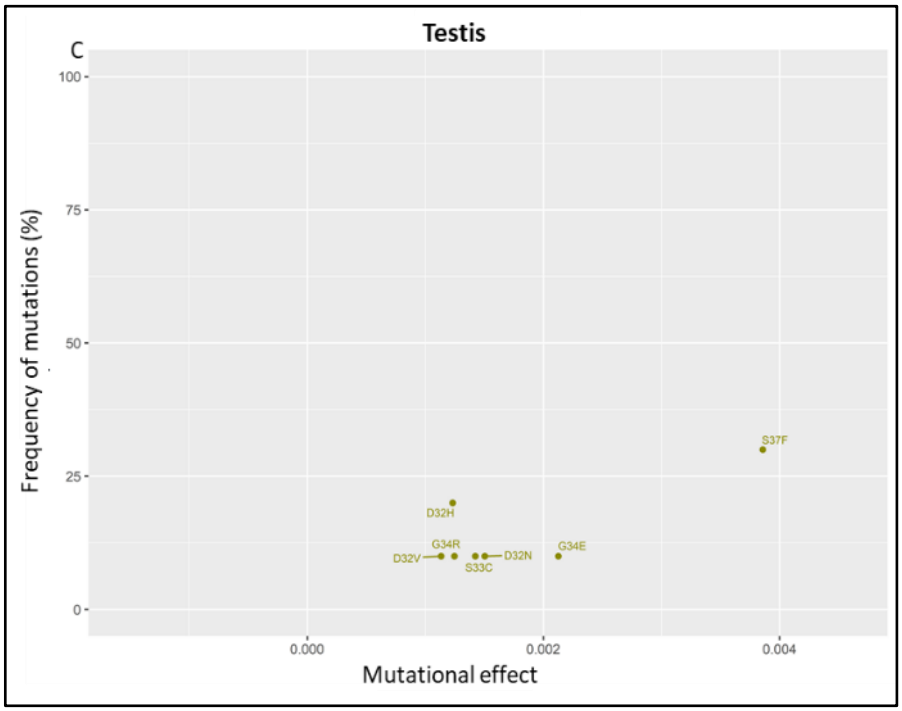
Next, the tumour types could be categorized into those having mutations of both lower and medium to higher mutational effect. Such a pattern was observed in the tumours of bone, biliary tract, CNS, lungs, ovaries, pancreas, pituitary, small intestine and breast (Fig 4-14 G-0). However, very few mutations that contribute to a lower effect could be observed in these tumour types and were present at a very low frequency when compared to the mutations that yield a medium to higher activity level. A similar pattern of mutations of both lower and medium to higher effect was seen even in the tumours stomach, however, these tumours had a slightly larger number of mutations that contribute to the lower mutational effect (Fig 4-14 P). The tumours of the thyroid and hematopoietic and lymphoid tumours although they had mutations of both lower and medium to higher mutational effect, in these tumours the mutations that contribute to the lower mutational effect outnumber those that yield a medium to higher effect (Fig 4-14 Q-R).

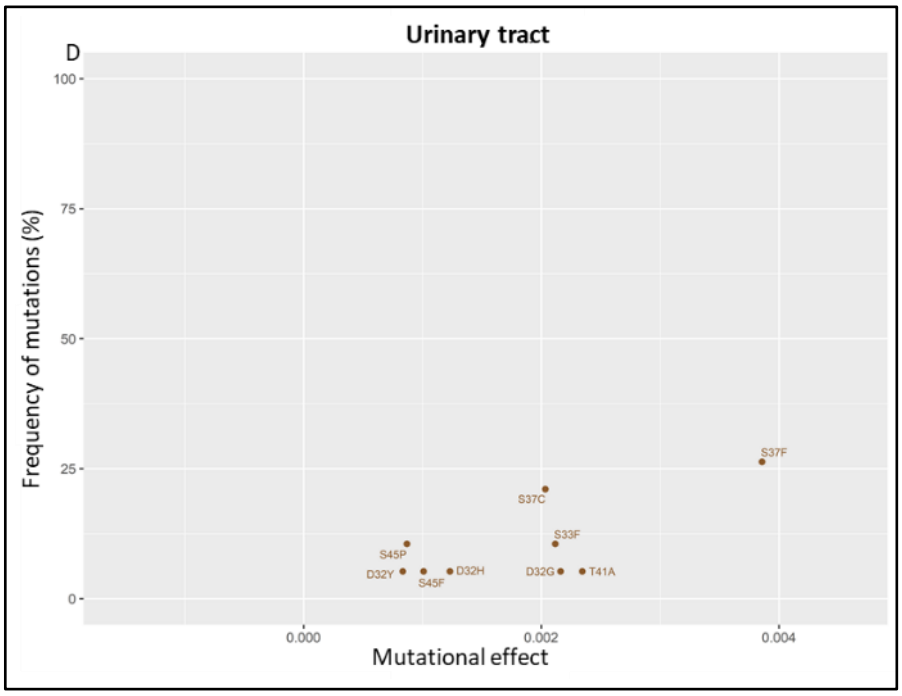
The tumours of the endometrium, large intestine, liver, skin and soft tissue, had very a wide spectrum of β-catenin mutations with varied mutational effect giving a scattered pattern (Fig 4-14 S-W). The large intestine and soft tissue, although they had a scattered pattern with multiple mutations of different mutational effect, the two mutations S45F and T41A giving a medium to higher effect occurred at an increased frequency in both these tumour types, especially prominent in the tumours of the soft tissue. The mutations in the kidney and adrenal gland tumours had a similar varied mutational effect, however, the

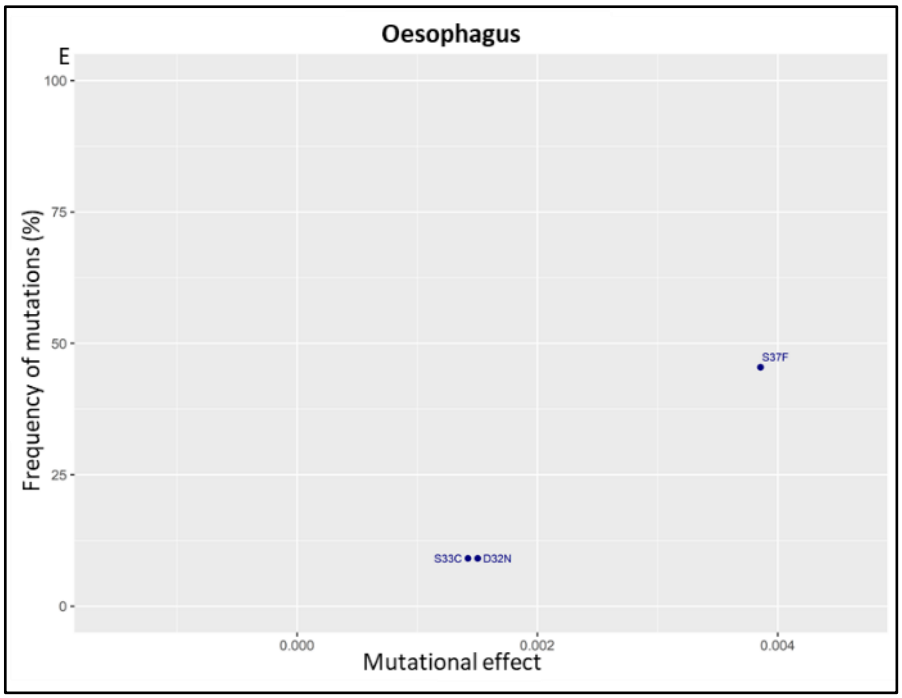
number of different mutations observed were lower in this tumour type (Fig 4-14 X-Y). Also, the kidney and adrenal gland had a higher frequency of S45P and S45F mutations. These two mutations present at an increased frequency, especially among the adrenal gland tumours, conferred a medium mutational effect. The distinct pattern of mutational effect conferred by the observed spectrum of mutations across the different types of tumours, indicates the presence of an activity based selection mechanism to be one of the major factor contributing to tumour type specific mutational bias.

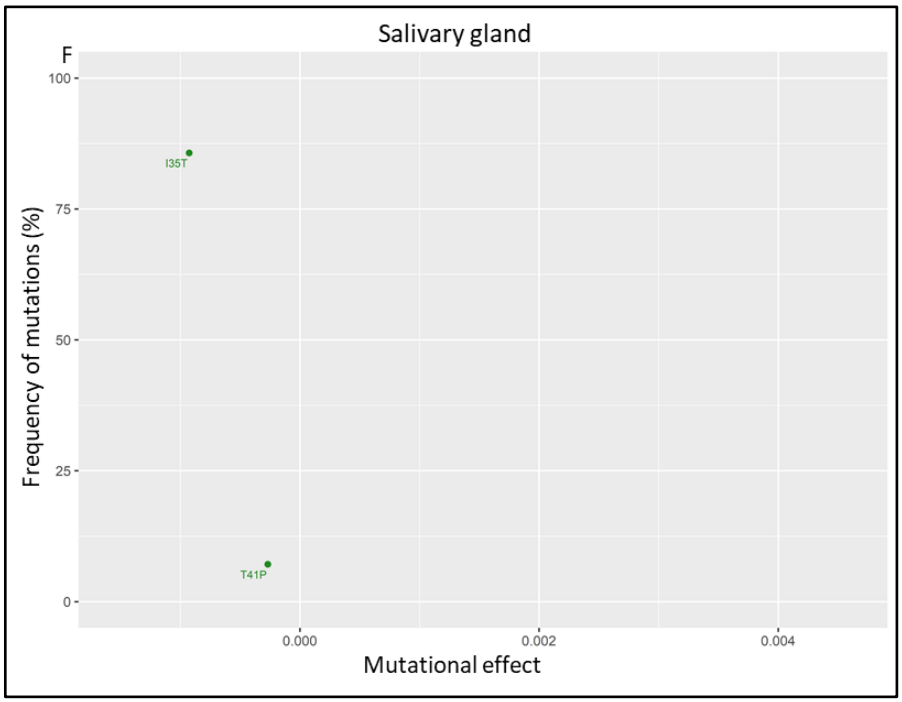


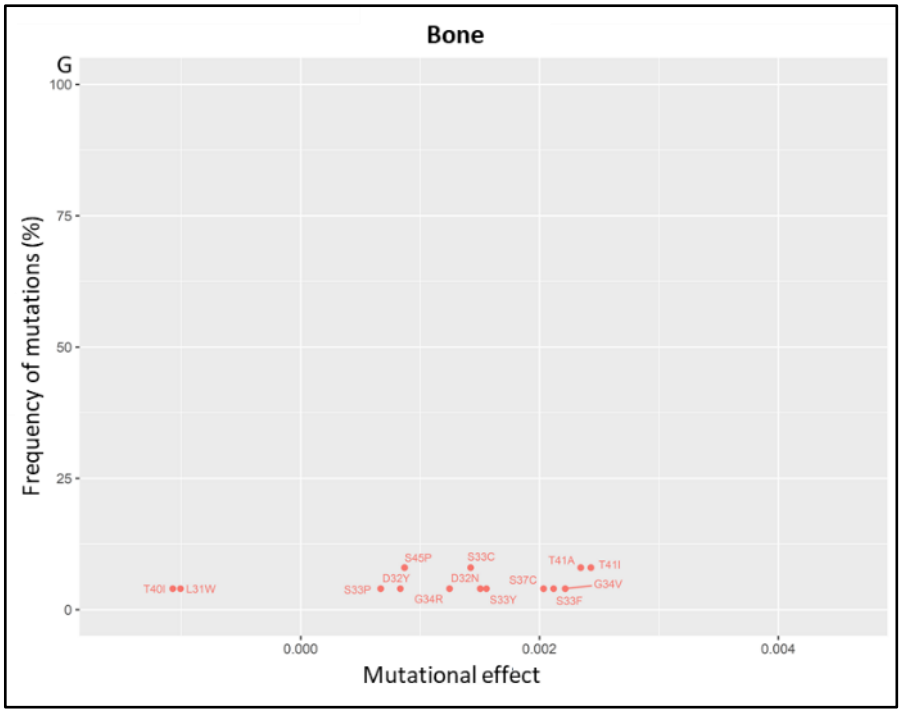

Prostate
B
Frequency of mutations (%)
100
75
50
25
0
T41A
S45P
D32V
D32Y
T41I
S33C
S33Y
S33F
G34V
S37A
D32H
D32N
G34E
D32G
0.000
0.002
0.004
Mutational effect

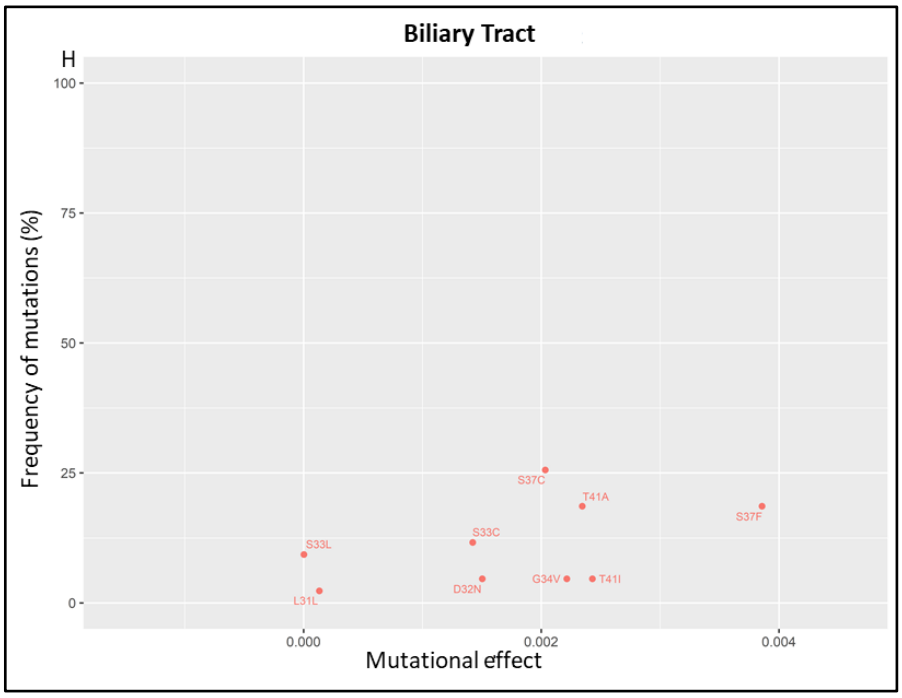

Testis
C
Frequency of mutations (%)
100
75
50
25
0
S37F
D32H
G34R
G34E
D32V
D32N
S33C
0.000
0.002
0.004
Mutational effect

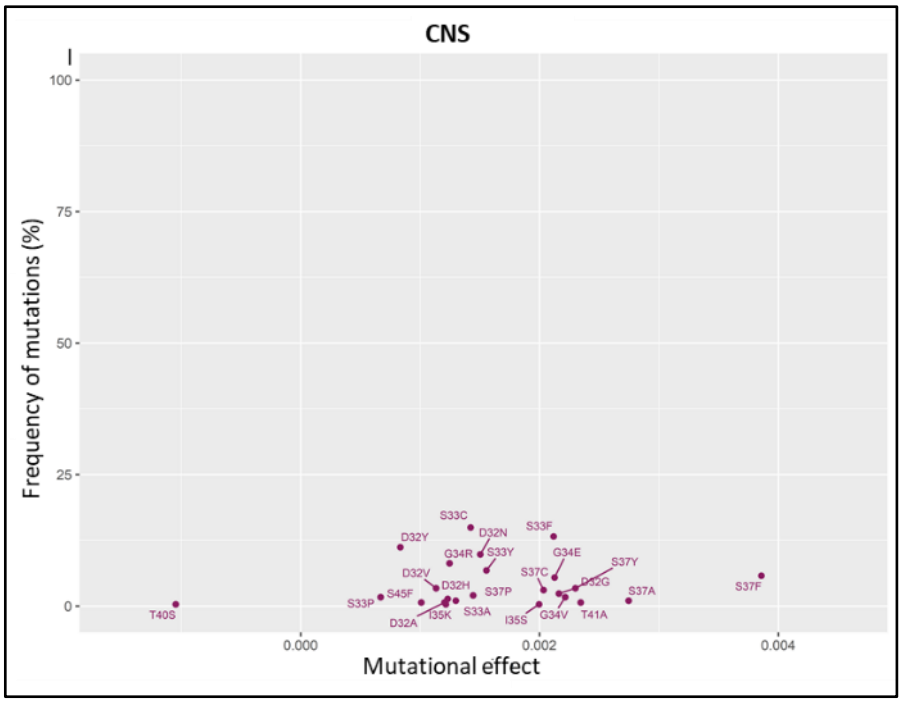

Urinary tract
D
Frequency of mutations (%)
100
75
50
25
0
S37F
S37C
S33F
S45P
D32Y
S45F
D32H
D32G
T41A
0.000
0.002
0.004
Mutational effect

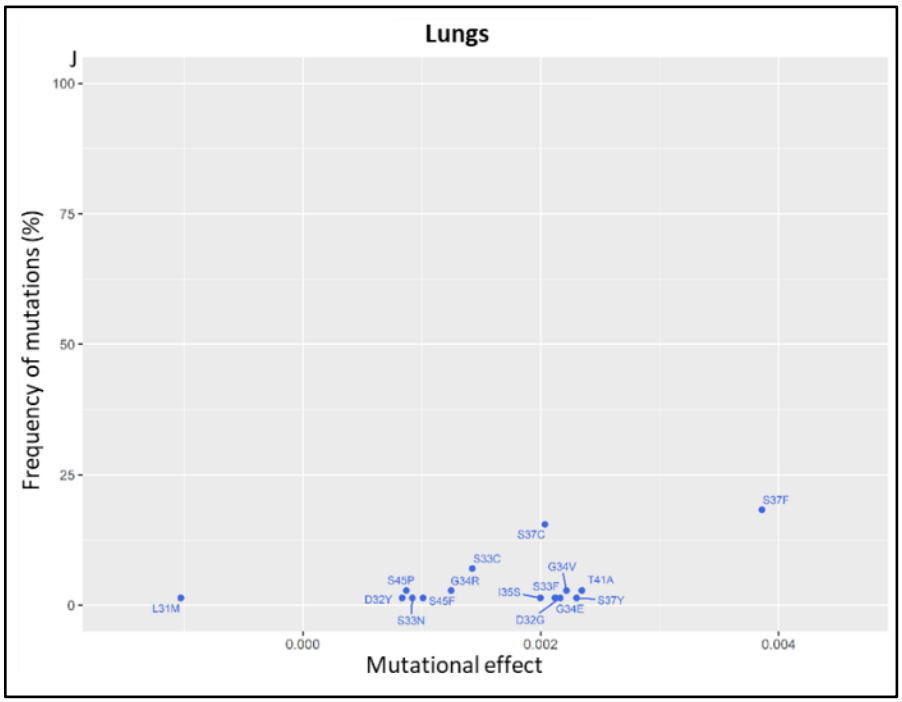

Oesophagus
E
Frequency of mutations (%)
100
75
50
25
0
S37F
S33C
D32N
0.000
0.002
0.004
Mutational effect

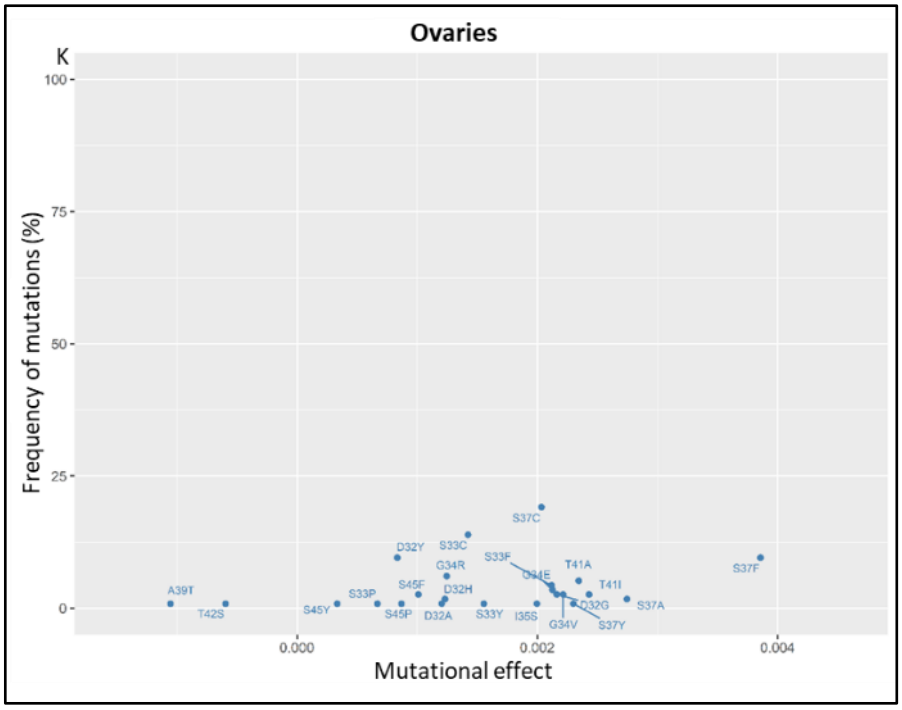

Salivary gland
F
Frequency of mutations (%)
100
75
50
25
0
I35T
T41P
0.000
0.002
0.004
Mutational effect

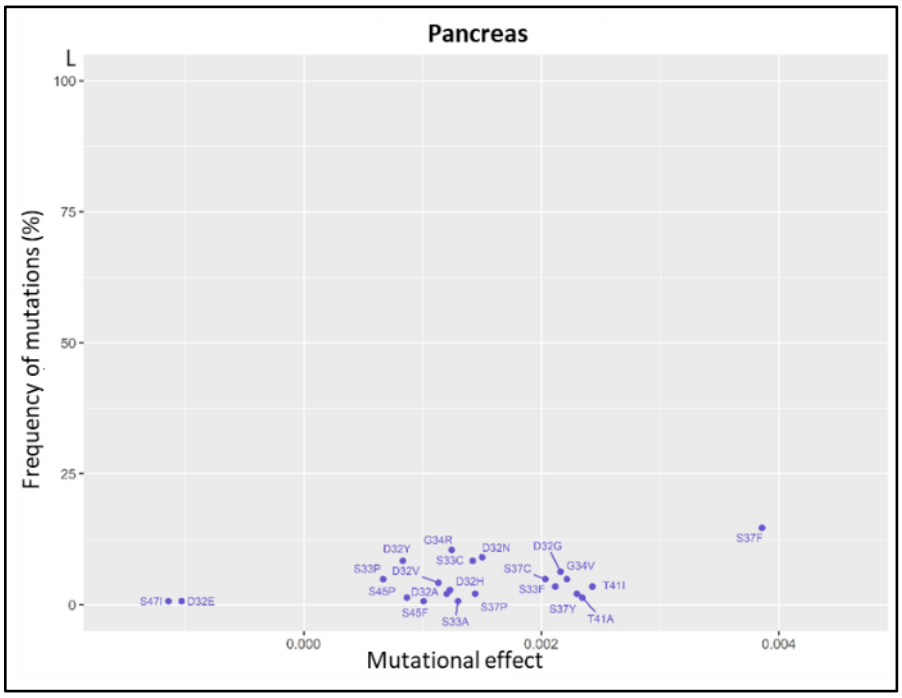

Bone
G
Frequency of mutations (%)
100
75
50
25
0
T40I
L31W
S33P
D32Y
S45P
G34R
D32N
S33C
S33Y
S37C
T41A
S33F
T41I
G34V
0.000
0.002
0.004
Mutational effect

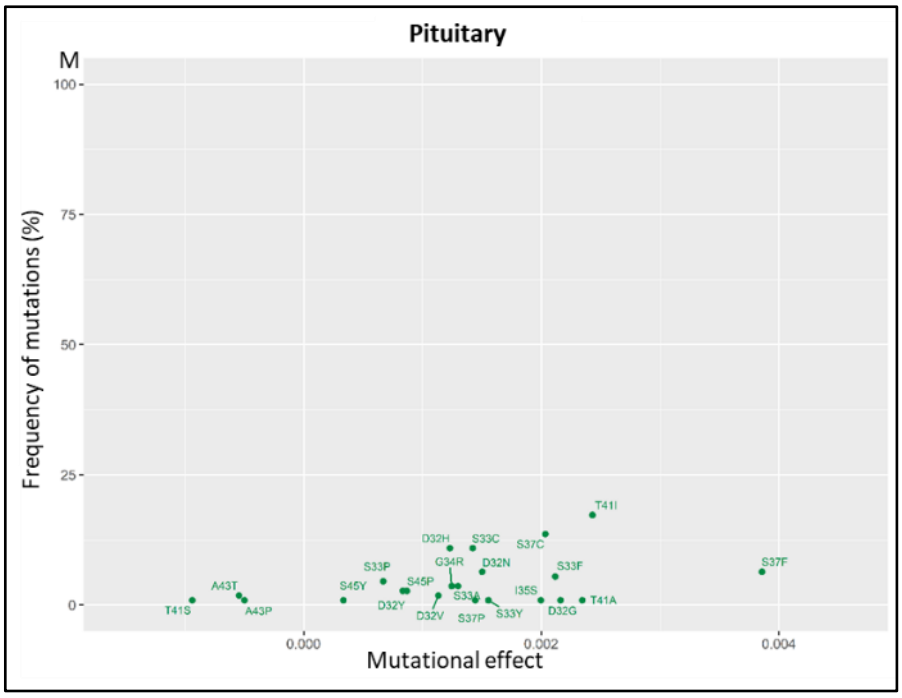

Biliary Tract
H
Frequency of mutations (%)
100
75
50
25
0
S37C
T41A
S37F
S33C
S33L
G34V
T41I
D32N
L31L
0.000
0.002
0.004
Mutational effect

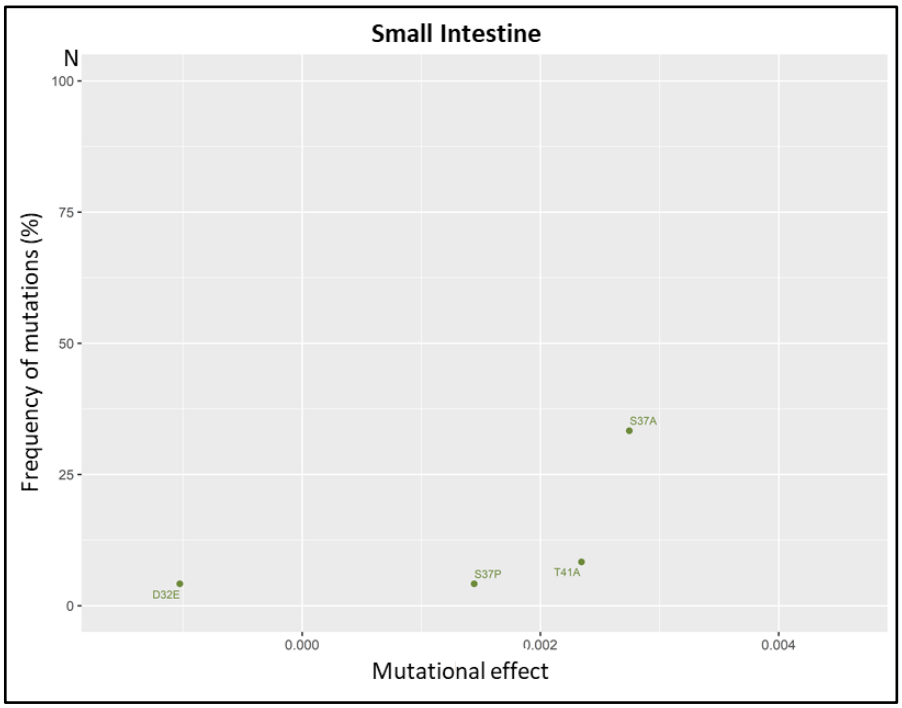

CNS
I
Frequency of mutations (%)
100
75
50
25
0
S33C
D32N
S33F
D32Y
G34R
S33Y
G34E
S37Y
D32V
S37C
D32H
D32G
S37F
S37P
S37A
S45F
S33P
T40S
I35K
S33A
I35S
G34V
T41A
D32A
0.000
0.002
0.004
Mutational effect

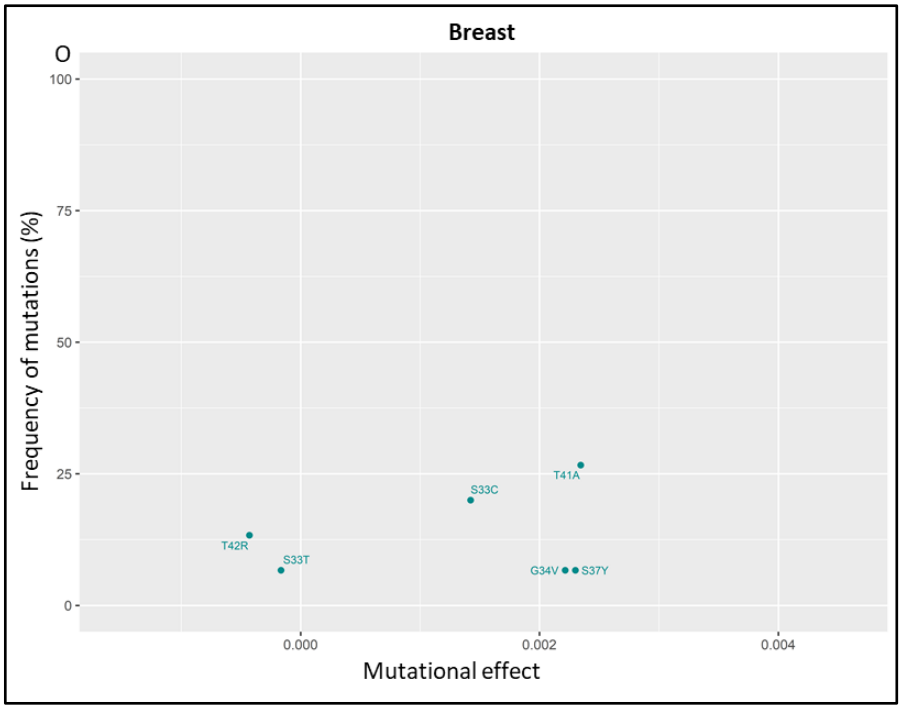

Lungs
J
Frequency of mutations (%)
100
75
50
25
0
L31M
S45P
D32Y
S33N
S45F
S33C
G34R
S37C
G34V
S33F
I35S
D32G
G34E
T41A
S37Y
S37F
0.000
0.002
0.004
Mutational effect

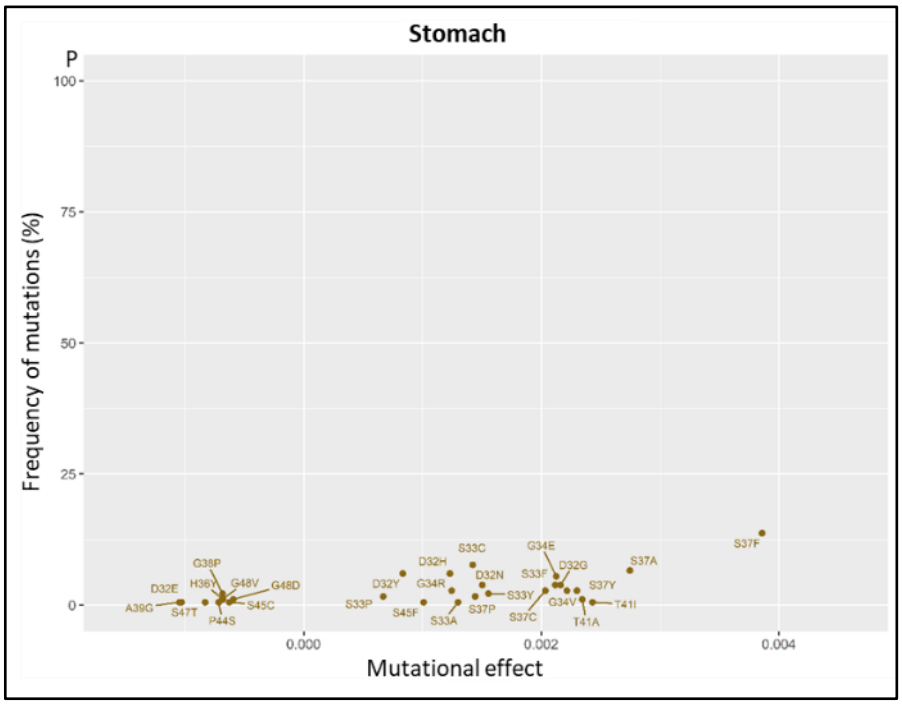

Ovaries
K
Frequency of mutations (%)
100
75
50
25
0
A39T
T42S
S45Y
S33P
S45F
S45P
D32Y
S33C
G34R
D32H
D32A
S37C
S33Y
S33F
G34E
I35S
T41A
D32G
G34V
T41I
S37A
S37Y
S37F
0.000
0.002
0.004
Mutational effect

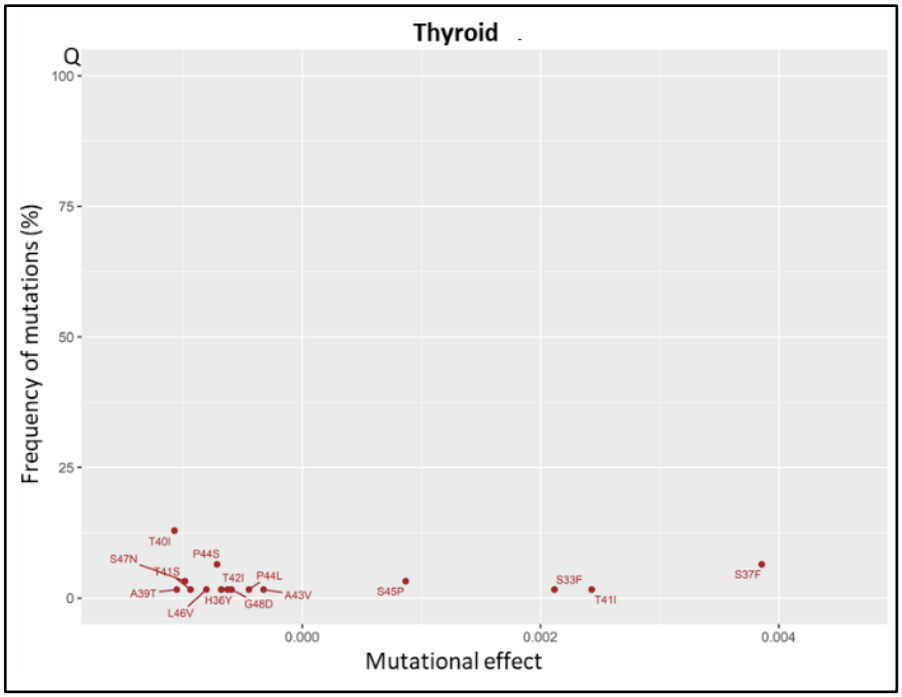

Pancreas
L
Frequency of mutations (%)
100
75
50
25
0
0.000
0.002
0.004
Mutational effect
S47I
D32E
D32Y
G34R
D32N
D32G
S33C
S33P
D32V
S37C
G34V
S45P
D32A
D32H
S33F
T41I
S45F
S33A
S37P
S37Y
T41A
S37F

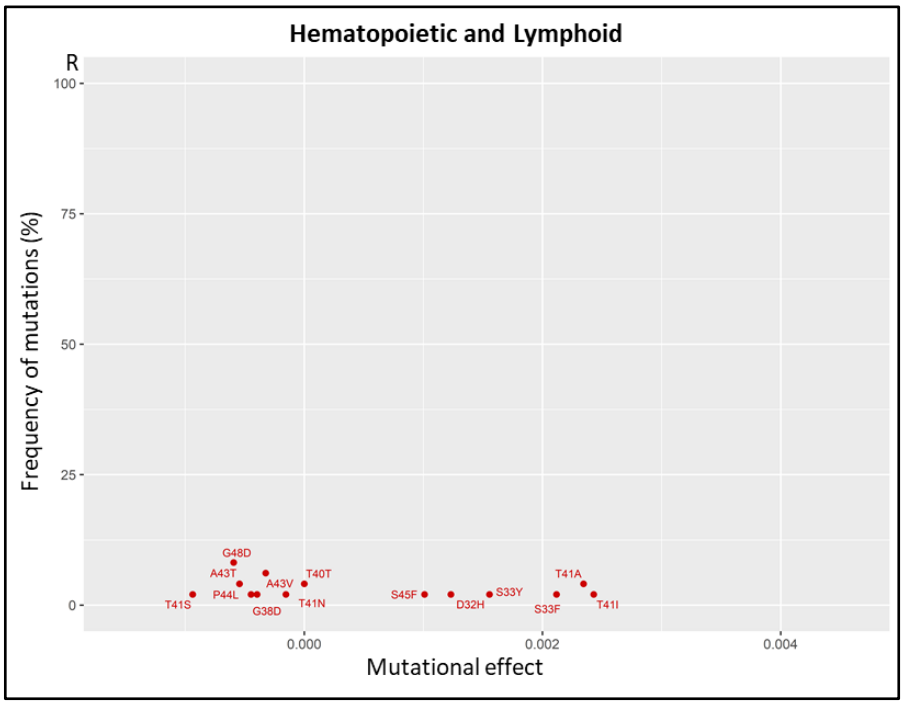

Pituitary
M
Frequency of mutations (%)
100
75
50
25
0
0.000
0.002
0.004
Mutational effect
T41I
D32H
S33C
S37C
S33P
G34R
D32N
S33F
S37F
A43T
S45Y
S45P
I35S
T41A
T41S
A43P
D32Y
S33A
D32V
S37P
S33Y
D32G

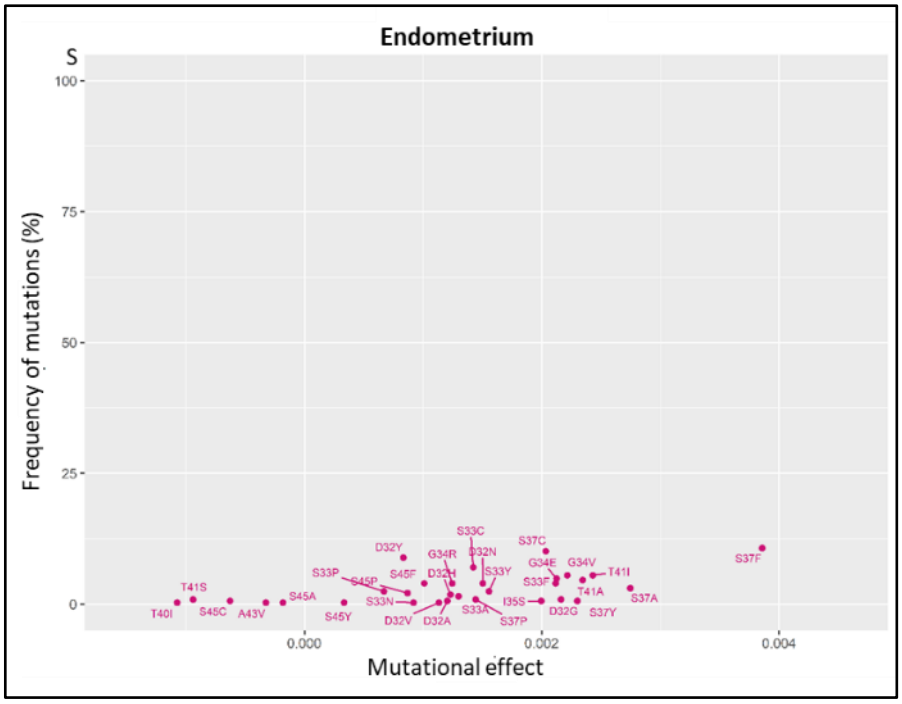

Small Intestine
N
100
75
50
25
0
Frequency of mutations (%)
S37A
S37P
T41A
D32E
0.000
0.002
0.004
Mutational effect

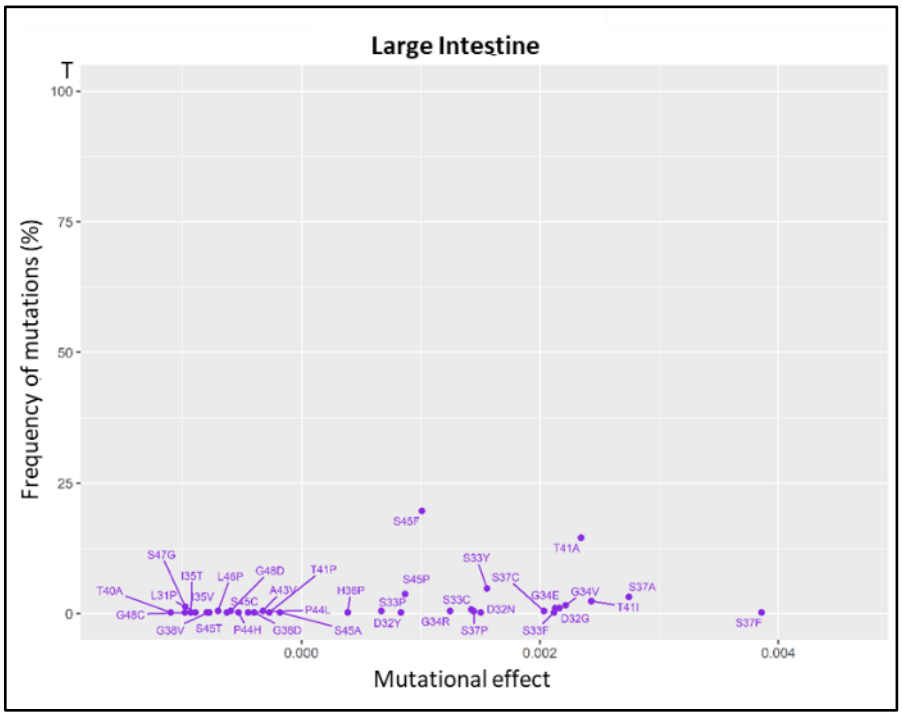

Breast
O
100
75
50
25
0
Frequency of mutations (%)
T41A
S33C
T42R
S33T
G34V
S37Y
0.000
0.002
0.004
Mutational effect

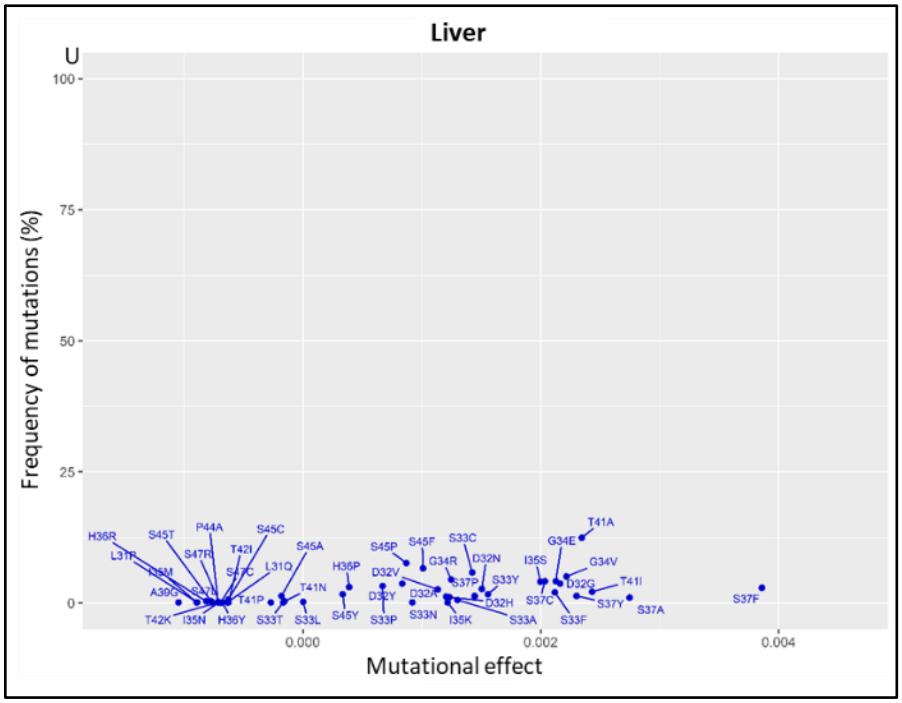

Stomach
P
Frequency of mutations (%)
100
75
50
25
0
G38P
H36Y
G48V
G48D
D32E
A39G
S47T
S45C
P44S
S33C
G34E
D32H
D32G
S37A
D32N
S33F
D32Y
G34R
S37Y
S33Y
S33P
S45F
G34V
T41I
S37P
S33A
S37C
T41A
S37F
0.000
0.002
0.004
Mutational effect

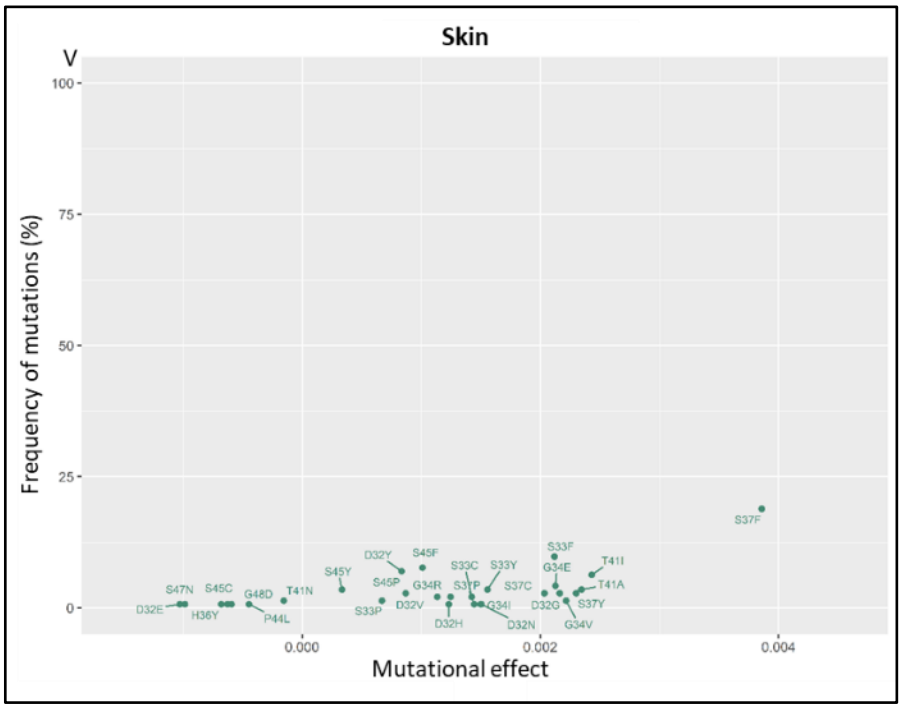

Thyroid
Q
Frequency of mutations (%)
100
75
50
25
0
T40I
S47N
P44S
T41S
T42I
P44L
A39T
H36Y
G48D
A43V
L46V
S45P
S33F
T41I
S37F
0.000
0.002
0.004
Mutational effect

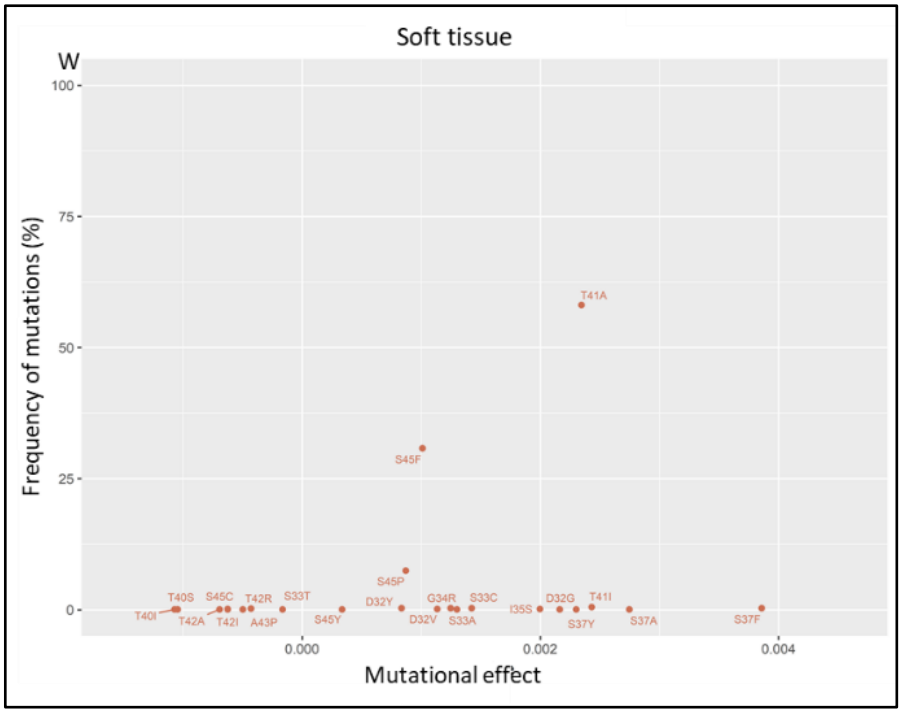

Hematopoietic and Lymphoid
R
Frequency of mutations (%)
100
75
50
25
0
G48D
A43T
T40T
A43V
P44L
T41S
G38D
T41N
S45F
S33Y
D32H
T41A
S33F
T41I
0.000
0.002
0.004
Mutational effect


Endometrium
S
Frequency of mutations (%)
100
75
50
25
0
S33C
D32Y
S37C
G34R
D32N
G34E
G34V
T41I
S33P
S45F
D32H
S33Y
S33F
T41S
S45P
T41A
S37F
S37A
S33N
I35S
S45A
S37A
T40I
S45C
A43V
S45Y
D32V
D32A
S33A
S37P
D32G
S37Y
0.000
0.002
0.004
Mutational effect


Large Intestine
T
100
75
50
25
0
Frequency of mutations (%)
S45F
T41A
S47G
S33Y
I35T
L46P
G48D
T41P
S37C
S45P
T40A
L31P
I35V
S45C
A43V
H36P
S33P
S33C
G34E
G34V
S37A
G48C
P44L
D32N
T41I
D32Y
G34R
D32G
S37F
G38V
S45T
P44H
G38D
S45A
S37P
S33F
0.000
0.002
0.004
Mutational effect


Liver
U
100
75
50
25
0
Frequency of mutations (%)
T41A
H36R
S45T
P44A
S45C
S45P
S45F
S33C
G34E
L31P
S47R
T42I
S45A
G34R
D32N
I35S
G34V
I35M
S47C
L31Q
H36P
D32V
S37P
S33Y
D32G
T41I
A36G
S47L
T41P
T41N
D32Y
D32A
S37C
S37Y
S37A
T42K
I35N
H36Y
S33T
S33L
S45Y
S33P
S33N
I35K
D32H
S33A
S33F
S37F
0.000
0.002
0.004
Mutational effect


Skin
V
Frequency of mutations (%)
100
75
50
25
0
S37F
S33F
D32Y
S45F
G34E
T41I
S45Y
S33C
S33Y
S45P
G34R
S37C
T41A
S47N
S45C
G48D
T41N
S37P
S37Y
D32E
D32V
G34I
D32G
H36Y
S33P
D32H
D32N
G34V
P44L
0.000
0.002
0.004
Mutational effect


Soft tissue
W
Frequency of mutations (%)
100
75
50
25
0
T41A
S45F
S45P
T40S
S45C
T42R
S33T
D32Y
G34R
S33C
D32G
T41I
T40I
T42A
T42I
A43P
S45Y
D32V
S33A
I35S
S37Y
S37A
S37F
0.000
0.002
0.004
Mutational effect

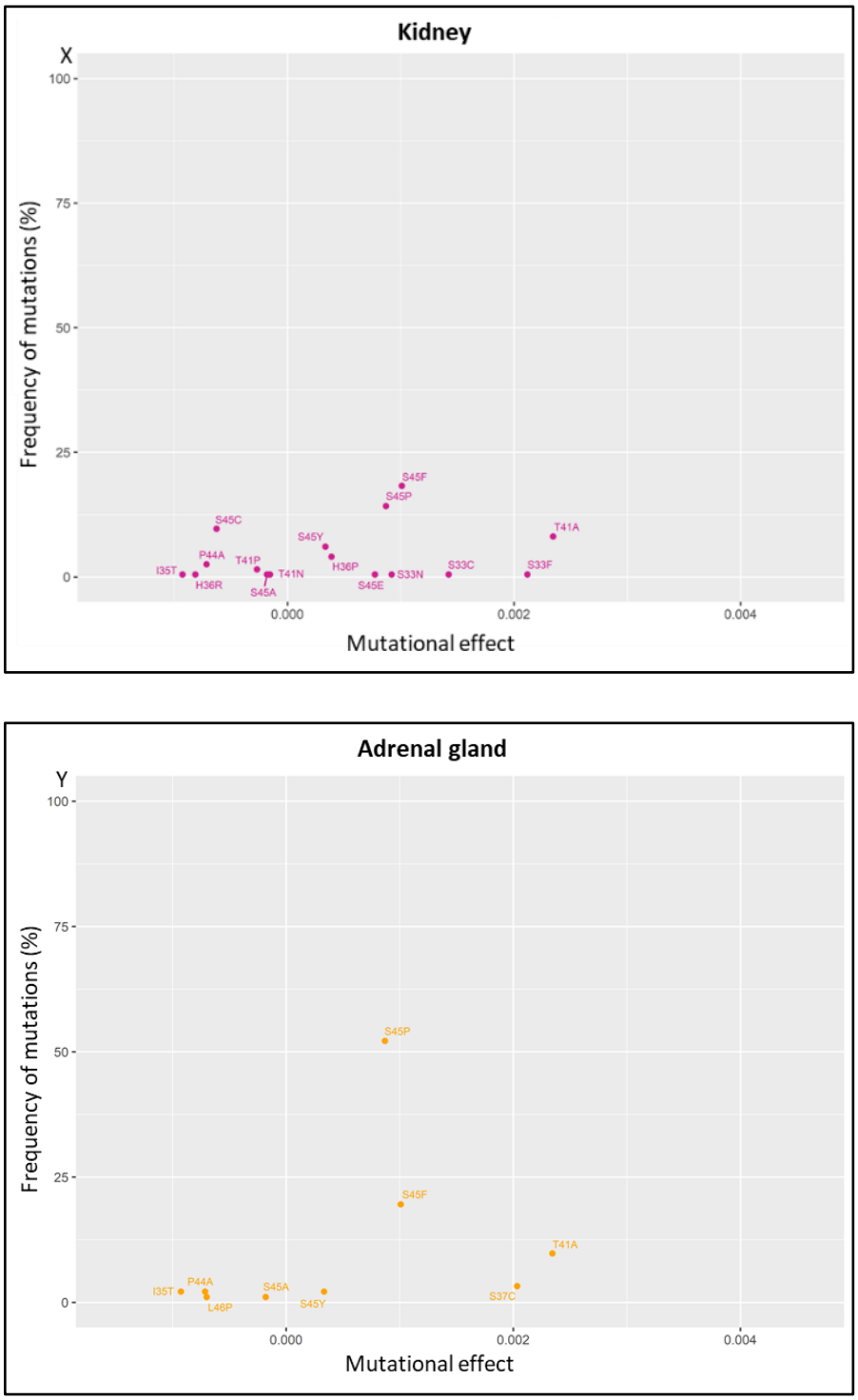


Figure 4-14: Analysis of β-catenin mutational effect for the mutations observed across different cancer types. Graphs A-Y represents the mutational effect (m value) vs frequency of mutations observed for each of the different tumour types compiled from COSMIC database.

4.2.4.9 Analysis of the Background mutational rate

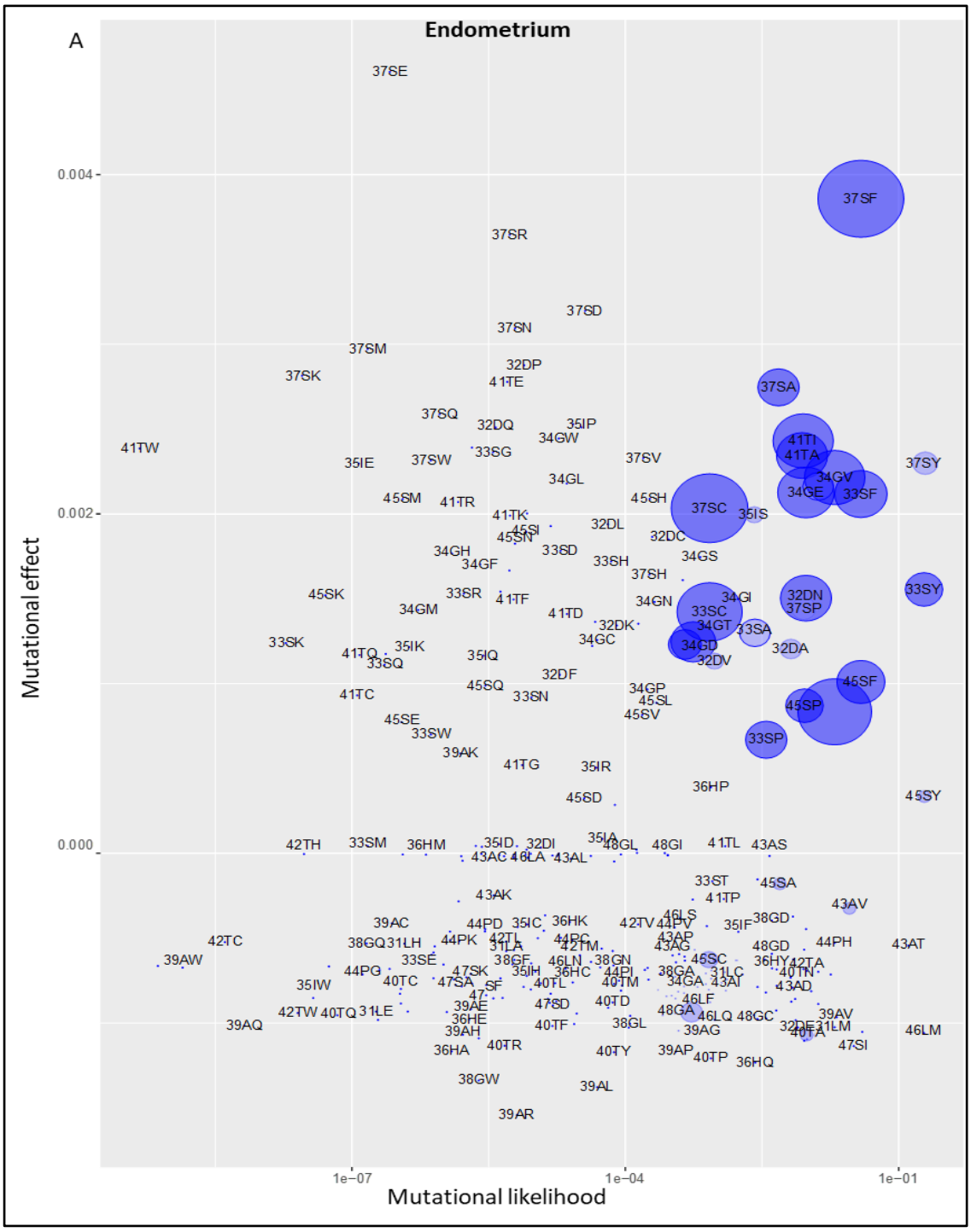
In addition to selection of mutations based on their functional significance, the differences in the sensitivity to endogenous and exogenous agents, and mutational processes, have been known to contribute to a distinct tumour type specific mutational signature. The preferential selection of G to T transversion in the p53 tumour suppressor gene is commonly observed in lung cancers associated with smokers (Pfeifer *et al.*, 2002). The AFB1-N7-Gua, the primary adduct formed by the reaction of the aflatoxin B1 with guanine, predominantly results in a specific G-T transversion in liver tumours, in patients with increased exposure to this fungal metabolite (Aguilar, Hussain and Cerutti, 1993). Similarly, the C to T transitions of the dipyrimidine residues and CC-TT double base change in the p53 gene are a distinct feature in Ultraviolet B (UVB) induced tumours of the skin (Brash *et al.*, 1991). In addition to the environmental mutagens, several other endogenous processes are also capable of leaving a distinct fingerprint in the cancer genome. The cytosine deaminase activity of the APOBEC enzymes is the source of the mutations observed in the helical domain of PIK3CA gene across various cancer types (Henderson *et al.*, 2014). The mutations in the mismatch repair genes are also associated with specific mutational signature. Included among these, are mutations in MUTYH gene that codes for the DNA glycosylase involved in BER, which are known to result in increased proportion of 8-oxoguanine induced G-T transversion, in various oncogenes and tumour suppressors in CRCs (Viel *et al.*, 2017). With evidence of various mutational processes in contributing to the tissue specific mutational bias, in the recent years, more and more studies are beginning to focus on the analysis of background mutational profile, to be able to categorize tumour types based on the mutational signatures.

It was thus important for us to analyse the background mutational rate, to understand how much of the difference observed in the β-catenin mutational pattern is due to the underlying differences in the mutational processes that different cancers with β-catenin mutations are exposed to. The analysis of the background mutational rate was performed by Ailith Ewing in Colin Semple's group in IGMM. The whole genome (or exome) sequencing data for samples with β-catenin mutations were collected from the TCGA database. However, except for liver and endometrial tumours, the number of tumour samples with whole genome sequencing data for other tumor types was sparse, hence the analysis was done only on these two tumour types.

The likelihood of a given amino acid substitution within the β-catenin hotspot region was calculated for liver and endometrial tumours using whole exome sequencing data from the TCGA project (described in detail in the materials and methods section). The mutational likelihood was then plotted against the previously calculated regression score (mutational effect), also taking into account the relative incidence of the mutational variants (depicted by the circle size).

The majority of β-catenin mutations selected for in both tumour types are typically those having a higher likelihood, and contributing to medium to higher mutational effect (Fig 4-15A and B). Although mutations with high likelihood and lower mutational effect can be seen in both the tumours, their incidence is much lower than those that yield an increase in the mutational effect. Comparison of both the tumour types indicate that specific mutations selectively happen in each of these tumour types despite having a high likelihood of occurrence. For example, the 36HP mutation is observed at a relatively high frequency in the liver, when compared to the endometrium, where no incidence of H36P mutation has been recorded (from sequencing data obtained from TCGA database), although the likelihood of occurrence of 36HP is high in both these tumour types. Likewise, specific mutations having a similar likelihood of occurrence in both these tumors do not always happen at a similar frequency in the two tumours. For example the 41TA mutation have similar likelihood of occurrence, but are observed at a higher frequency in the tumours of the liver compared to the endometrium, indicating a tissue specific selection bias based on functional significance.

Overall, these results indicate that both background mutational profile and the β-catenin activity are essential forces that together contribute to the specificity of the observed mutational pattern in the liver and endometrial tumours. Although the likelihood of the occurrence of a given mutation plays an important role in determining the presence of these mutations, selectivity based on tumour specific functional significance seemed to be a much more dominant force in determining the distinct mutational spectrum observed for the different tumour types.


A
Endometrium
Mutational effect
Mutational likelihood
0.004
0.002
0.000
1e−07
1e−04
1e−01

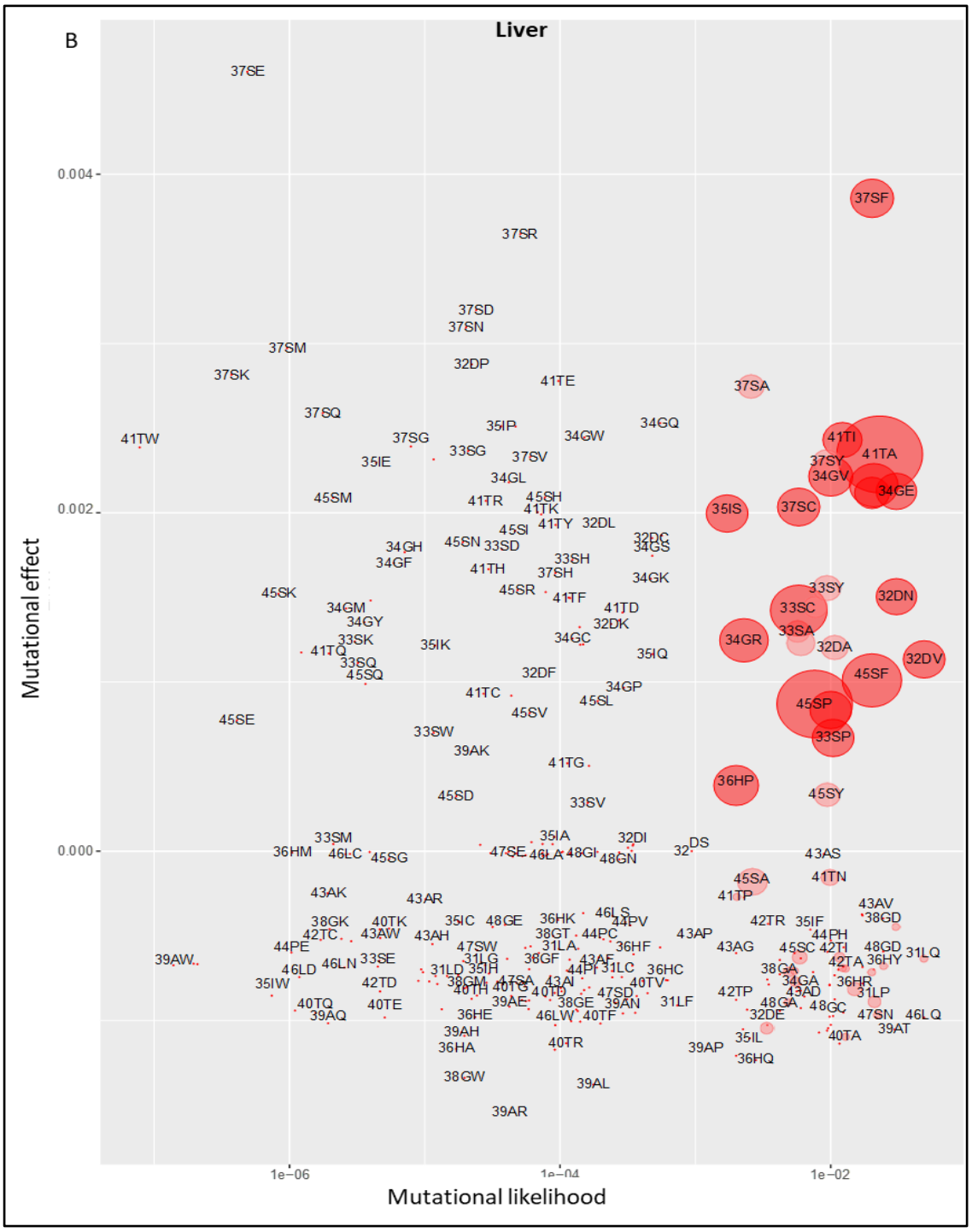


Figure 4-15: Analysis of the background mutational rate. Graphs representing the likelihood of amino acid substitution vs the corresponding regression score (effect) and also taking into consideration the relative frequency of occurrence for tumour types (A) endometrium and (B) liver.

4.3 Discussion

The complexity of the cancer genome landscape is defined by a variety of mutations, often leaving a distinct imprint specific to the particular tissue/organ in which they are observed. The spectrum of mutations in a particular cancer type may be a result of various selection processes. Two of the well documented selection principles include A) selectivity based on the specific activity conferred by the variant, particularly of those increasing the fitness and in the process providing a specific tumorigenic advantage to the clonal population – in accordance with the Darwinian evolutionary principles or B) the susceptibility to exogenous and endogenous agents/processes producing a specific background mutational signature. These two principles either on their own or in combination have been known to be the major contributors of the mutational bias observed across different cancer types.

Having observed a strong preferential selection of specific residues and amino acid substitutions in the mutational spectrum of *CTNNB1* across various tumour types, firstly I sought to explore if different β-catenin activity levels can be the Darwinian force to select for specific mutations by performing a saturation mutagenesis screen. The highest frequency of mutations was observed between the residues L31-G50 in the exon 3 region of β-catenin, and hence I choose to perform a saturation mutagenesis screen of this entire stretch of 20 residue region of the β-catenin mutational hotspot.

The TCF/Lef:H2B-GFP reporter mESC was the cell line of choice as it allowed quantitation of β-catenin activity at single cell resolution, which meant that we could easily FACS sort the cells into different segments of the activity spectrum, and this would provide an efficient system to assess the allele specific activity of the β-catenin variants. These cells were derived from blastocysts cultured in R2i and were sent to us after being cultured in the same media for a few passages. Since the presence of GSK3β inhibitor in the R2i media would interfere in phenotyping the mutant cells, I tried to ween them off from R2i to normal ES media. Even after culturing on feeders and weening them of R2i by gradual reduction of 2i, these cells still failed to adopt to normal ES media conditions evident by increased differentiation and cell death when cultured for over a week. However, I observed that the cells could keep their morphology intact and maintained an

undifferentiated state when cultured normal media for 2-3 days, and this short time scale was also sufficient for reduction of β-catenin activity.

It was further necessary to carry out this screen in heterozygous condition by targeting only one allele and keeping the other allele as WT. This was important, not only to maintain physiological relevant scenario, but as the screen is based on NGS of pooled cells, and with CRISPR being biallelic, it is not possible to differentiate clean heterozygous mutants from compound heterozygous mutants with indel on the other allele. Furthermore, the presence of indel in one of the alleles may alter the phenotypic response of the specific missense allelic variant. For this purpose, as discussed in chapter 3, I tried to initially generate a PAM mutant heterozygous cell line but even after several attempts I was unable to generate this cell line. In addition, our initial strategy of using ssODN as HDR repair template also remained unsuccessful. However the targeting approach based on positive negative selection strategy puDeltatk selection along with the HDR vector library that was designed to overcome these drawbacks proved to be very successful.

The analysis of each of the six segments of the sorted population provided the initial evidence of the allele specific variation in β-catenin activity. As expected, the serine and threonine residues that govern the regulation of the phosphorylation dependent stabilization of β-catenin were among those having a maximum impact on the TCF mediated β-catenin activity. The overall increase in β-catenin activity by the amino acid variants of the priming S45 residue was lower in comparison to the other residues in the sequential cascade. These results already contradict the current dogma of β-catenin activity based on the sequential phosphorylation. Previously, it has been shown that a colorectal cancer cell line with endogenous S45 mutation (S45F) had phosphorylated T41/S37/S33 pool in the absence of S45 phosphorylation showing, S45 priming is not essential for GSK3 mediated phosphorylation of residues T41,S37 and S33 (Wang, Vogelstein and Kinzler, 2003). These observations may suggest S45 priming dependent sequential phosphorylation model of β-catenin regulation is not sufficient to explain this complex pathway.

In addition to the S45 residue, the overall β-catenin activity also varied among the T41, S37 and S33 mutants. The S37 mutants showed the highest overall increase in β-catenin

activity. The overall β-catenin activity conferred by the S33 variants also follow a similar trend as S37 mutants. The phosphorylation of both S37 and S33 residues are known to be important for the recognition by E3 ubiquitin ligase which might be the reason for the observed effect (Hart *et al.*, 1999). However the proportion of the S33 variants in the high activity segments was much lower than that observed for S37 mutants which indicate that the two residues might be differently regulated. In addition to the phosphorylatable residues, the D32 and G34 variants are also capable of increasing the β-catenin activity. The D32 and G34 residues are also known to be a part of the degron motif and the observed effect of mutations at these residues underscores their significance in regulation of β-catenin activity.

Furthermore, for each of these six residues, there exists variation in the proportion of amino acid variants across the different segments, indicating that not all amino acid variants are capable of increasing β-catenin activity to a similar extent. The phosphorylatable S33, S37, S45 and T41 residues when mutated to S33T, S37T, S45T and T41S, respectively, fail to increase the β-catenin activity suggesting that the serine and threonine residues are interchangeable at these sites. The aspartic (D) and glutamic acid (E) due to their structural resemblance to phosphoserine and phosphothreonine are often used as their phosphometics. However, except for S33E, the substitution of phosphorylatable S and T residues with D or E did not mimic the WT β-catenin activity levels in our screen. In addition, variable increase in activity levels were observed among the amino acid variants for a specific residue. For example, among the T41 variants the T41P had the highest frequency in the P4 segment, whereas majority of the residues were seen to have higher frequencies either in the P5 or P6 segments, and the T41I, T41A (also the most frequently observed T41 variants in our COSMIC analysis) are among the only few variants with highest frequency in the P7 segment. This variability in TCF dependent β-catenin activity among the amino acid substituents was observed for all the four phosphorylatable residues and also for residues D32 and G34. The observed differential activity response by the amino acid variants, however, cannot be explained based on the current knowledge of β-catenin regulation and will need to be further investigated by performing biochemical and structural analysis of these variants on their own and also in the combination with their interacting partners.

The structural analysis of the β-catenin degron motif predicted that any hydrophobic residue can be tolerated at position I35 and any residue can be tolerated at residue H36 (Wu *et al.*, 2003). A similar observation was made from our screen with majority of the hydrophobic residues seem to be well tolerated at the residue I35 in comparison to the hydrophilic residues (with the exception of asparagine and glycine where opposite trend was seen). Similarly, at the residue H36 any nonsynonymous mutation was tolerated except proline. H36 substitution to proline was the only non-synonymous substitution giving an increased β-catenin activity and incidentally among the H36 mutations observed in various cancer types, H36P substitution was most commonly selected for especially among the tumours of the liver. However among all the H36 variants how proline alone is capable of increasing the β-catenin activity remains to be understood.

The graphs of the regression vs the frequency of mutations for each of the individual tumour types obtained from COSMIC database showed a distinct pattern of activity levels for each of the tumour types. The mutations observed in majority of the tumours were not confined to a single activity level but instead selected for different activity ranging from low, mid to high levels. The genotype-phenotype study of liver tumours by Rebouissou et al have provided evidence of a similar selection of mutations based on activity levels (Rebouissou et al 2016). In this study, the HCC were seen to particularly select for mutations with medium or low activity levels, compared to HCAs wherein a selection of mutations with a lower or medium activity was preferred. The mutations of lower activity when present in HCCs, almost always had an additional mechanism (either LOH, duplication of allele or an additional β-catenin activating mutation) leading to higher activity levels. In addition, different β-catenin mutations were seen to associate with mutations in specific genes. In HCAs, the mutation at residue T41 was observed to be frequently associated with mutations in the *IL6ST* gene. However, the β-catenin S45 mutations were mutually exclusive to these JAK/STAT activating *IL6ST* mutations. A similar association of specific genetic alterations is observed in Wilm's tumour, where the inactivating *WT1* mutations are found to co-occur with mutations at S45 residue of β-catenin (Maiti *et al.*, 2000). These studies indicate that many factors including the requirement of specific activity levels by tumour sub types, additional genetic alterations producing an increased activity level, preferential selection of certain β-catenin activity level together in association with specific genetic alterations producing a synergistic effect

may contribute to the selection of specific activity levels, and provide a possible explanation for the selection of a distinct pattern of low, mid and higher mutational effect observed in our study for the majority of the tumour types.

In addition to analyzing the selection of mutation based on activity levels, it was also important to understand the contribution of the background mutational profile in the observed preferential selection of the different β-catenin mutations in the hotspot region. Increased exposure to environmental mutagens such as tobacco smoke, UV rays, fungal metabolite aflatoxin and other endogenous processes including the cytosine deaminase activity of the APOBEC enzymes, transcriptional strand bias DNA mismatch repair genes have all been known to account for the tumour type specific mutational signature. Besides saturation mutagenesis screen, we also performed an analysis of the background mutational rate of β-catenin. The investigation of the likelihood of amino acid substitution in the β-catenin hotspot region was however limited to the liver and endometrial tumour types, due to the unavailability of good sample size of whole exome sequencing data in the TCGA database for the other tumour types. The analysis of mutational likelihood, together with the corresponding regression score of the various amino acid substitutions in the β-catenin hotspot region, and also taking into consideration the relative frequency of occurrence for liver and endometrial tumours suggests that, although the background mutational rate does contribute to the presence of the mutation, the selectivity based on tissue specific functional significance might play a more dominant role in contributing to the observed mutational bias for the two analysed tumour types. These results further signifies the genotype-phenotype correlation of the β-catenin mutations to be a major contributor of the preferential selection of mutations observed across the different tumour types.

Chapter 5 Multiplex targeting and β-catenin functional assay

5.1 Introduction

The analysis of COSMIC database revealed mutations at various residues in the exon 3 region of *CTNNB1* gene, and also further selection of different amino acid variants among different tumours. The saturation editing of exon 3 region of β-catenin provided insights into the existing genotype-phenotype variations among the various mutations, and as a complementary approach, we decided to generate clonal cell lines of the known mutations observed in cancers that would allow in-depth functional analysis. However, as it is not possible to study all of the observed mutations in great detail, we selected to study the top six mutations and the subsequent amino acid substitutions observed at a high frequency (that were also statistically significant) across various cancer types. In this second approach, heterozygous β-catenin mutant cell lines were generated to understand the genotype-phenotype correlation among these observed mutations.

The variability in HDR frequency, and the generation of heterozygous mutants using ssODN as repair template, as evident by the differences in editing rates in the initial round of multiplex targeting, required the need for a new strategy. The puDeltatk counter selection strategy in combination with vector based HDR template proved to be a successful combination for saturation editing assay, therefore, the same strategy was adopted for the generation of heterozygous β-catenin mutant cell lines. For this purpose, as detailed in chapter 3, a heterozygous β-catenin KO E14 cell line with puDeltatk counter selection cassette and a β-catenin golden gate destination vector was generated. In this chapter, a detailed view of the cloning of 26 targeting vectors followed by multiplex targeting in E14 β-catenin KO cell line will be provided. Using this approach, cell lines harbouring the top 6 statistically significant residues with all the significant amino acid substitution (4-5 amino acid substitution for each residue) were generated. The β-catenin activity in these mutant cell lines was measured by transfection of luciferase reporter constructs (TOPflash/FOPflash luciferase reporter). In addition, the differential gene expression in these mutant cell lines were further analysed by Taqman assay to understand the genotype-phenotype correlations.

5.2 Results

5.2.1 Cloning of multiplex Targeting vectors

The targeting vectors (TVs) for multiplex targeting were generated using the same β-catenin backbone vector that was previously used for cloning of the TVs for saturation editing assay. Double stranded (ds) oligos with BbsI site and the desired mutation were ordered as separate pools for each of the six residues. Six sets of cloning was performed, and the transformants were preliminary screened by colony PCR. Since the β-catenin backbone vector had the region of interest deleted, the successfully cloned vectors would reintroduce the insert, and could be easily identified based on size difference of the amplicon. The PCR positive clones were sequenced by Sanger sequencing. One correct clone for each mutation was selected and maxiprep plasmid isolation was individually done for of each of the 26 multiplex targeting vectors.

5.2.2 Generation of heterozygous β-catenin mutant clones by multiplex targeting

Endogenous heterozygous mutants were generated for the various β-catenin mutations observed at a high frequency and statistically significant proportions across cancer types. Six sets of targeting was performed, wherein CRISPRs targeting the puDeltatk selection cassette and the HDR templates with all the selected amino acid variants for a particular residue were transfected into the heterozygous β-catenin KO E14 clone. For example, for residue T41, HDR vector templates having T41A, T41I, T41P, T41S and T41N substitutions were transfected along with the CRISPR/Cas9 guide vectors in a single experiment (Fig 5-1). Since the heterozygous β-catenin KO clone was being cultured in R2i media, the transfections were carried out in R2i media. Next day post transfection, the cells were trypsinised and plated in 10cm dishes in normal ES media and 8hours later, the media was supplemented with FIAU negative selection analogue. Following ten days of culture, when visible colonies appeared, 200 clones were picked for each targeting, and once confluent each plate was split into two plates. One plate was frozen down, and from the second plate DNA was isolated and genotyping PCR was performed and sequenced.

The sequencing of the clones revealed over 80 percent efficiency in targeting. The strategy of using positive negative selection in combination with CRISPR Cas9 and vector based HDR template once again proved to be a successful combination for generating the various mutant clones with high efficiency.

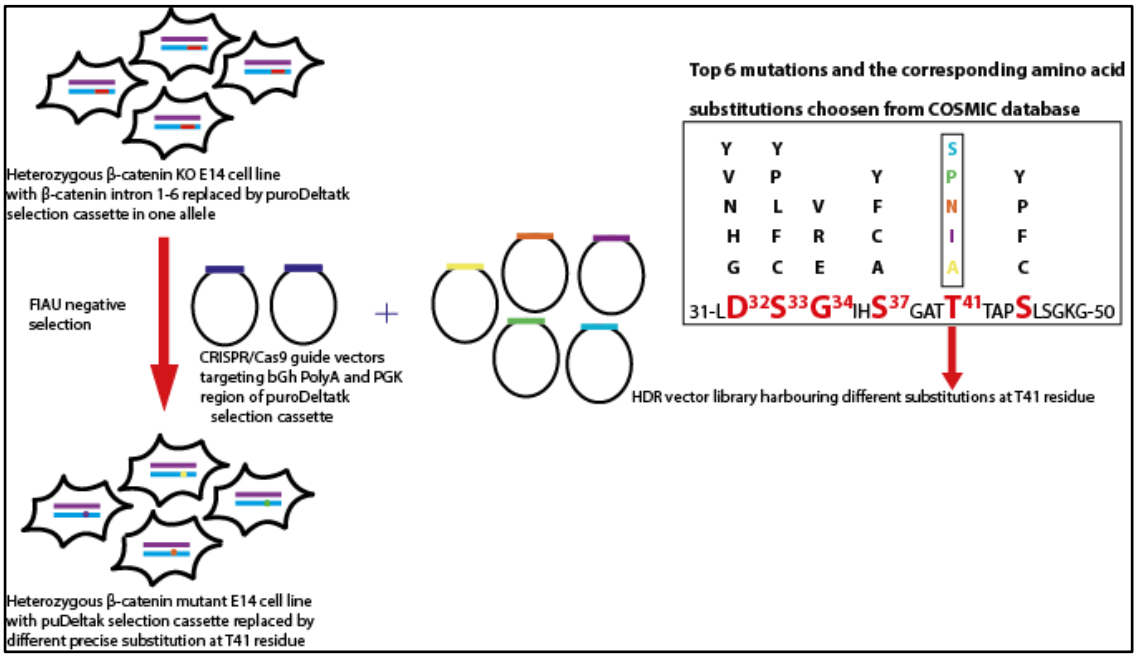
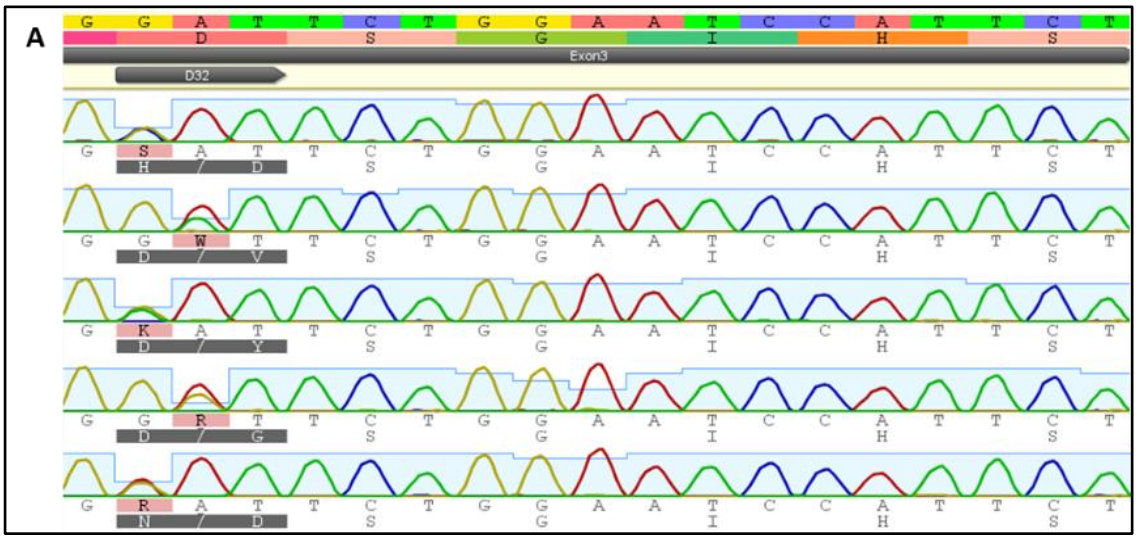
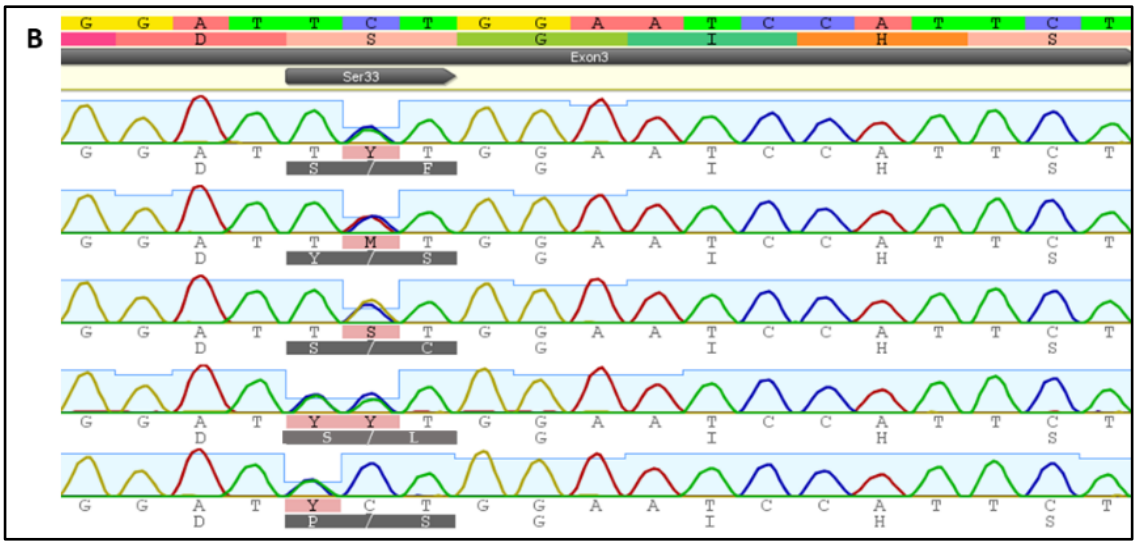
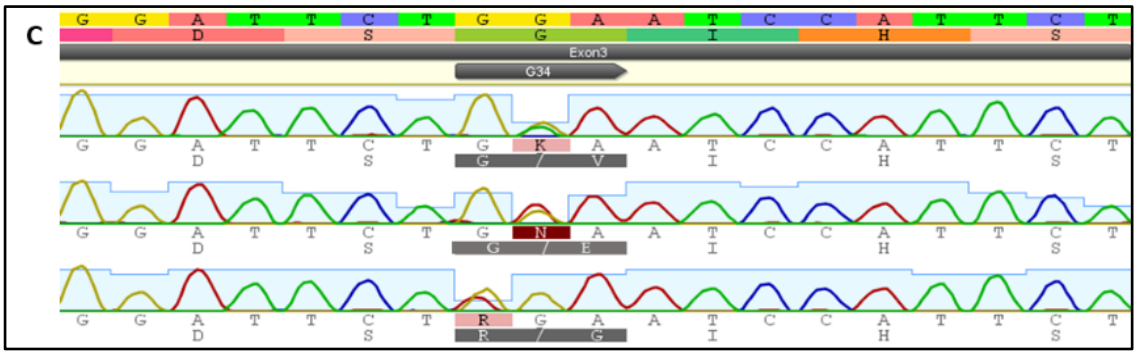


Figure 5-1: Schematic representation of the experimental design for generation of mutant cell lines by multiplex targeting. The CRISPR/Cas9 system combined with multiplex targeting to be used to induce mutations at the chosen top six residues along with the corresponding substitutions in the endogenous β-catenin gene.

In order to ensure that the effects we see is not due to the clonal variation but the effect of the mutation, I decided to analyse the clones in triplicates (3X26 mutants = 78 cell lines). Three clones for each of the mutations was started up in 24 well culture plates, expanded and frozen down, and DNA was isolated and re-sequenced for confirmation of mutation (Fig 5-2). After confirmation of mutants, cells were started up and plated for luciferase assay, and also RNA was isolated (for cDNA synthesis) to perform Taqman assay.


A
Exon3
D32


B
Exon3
Ser33


C
Exon3
G34

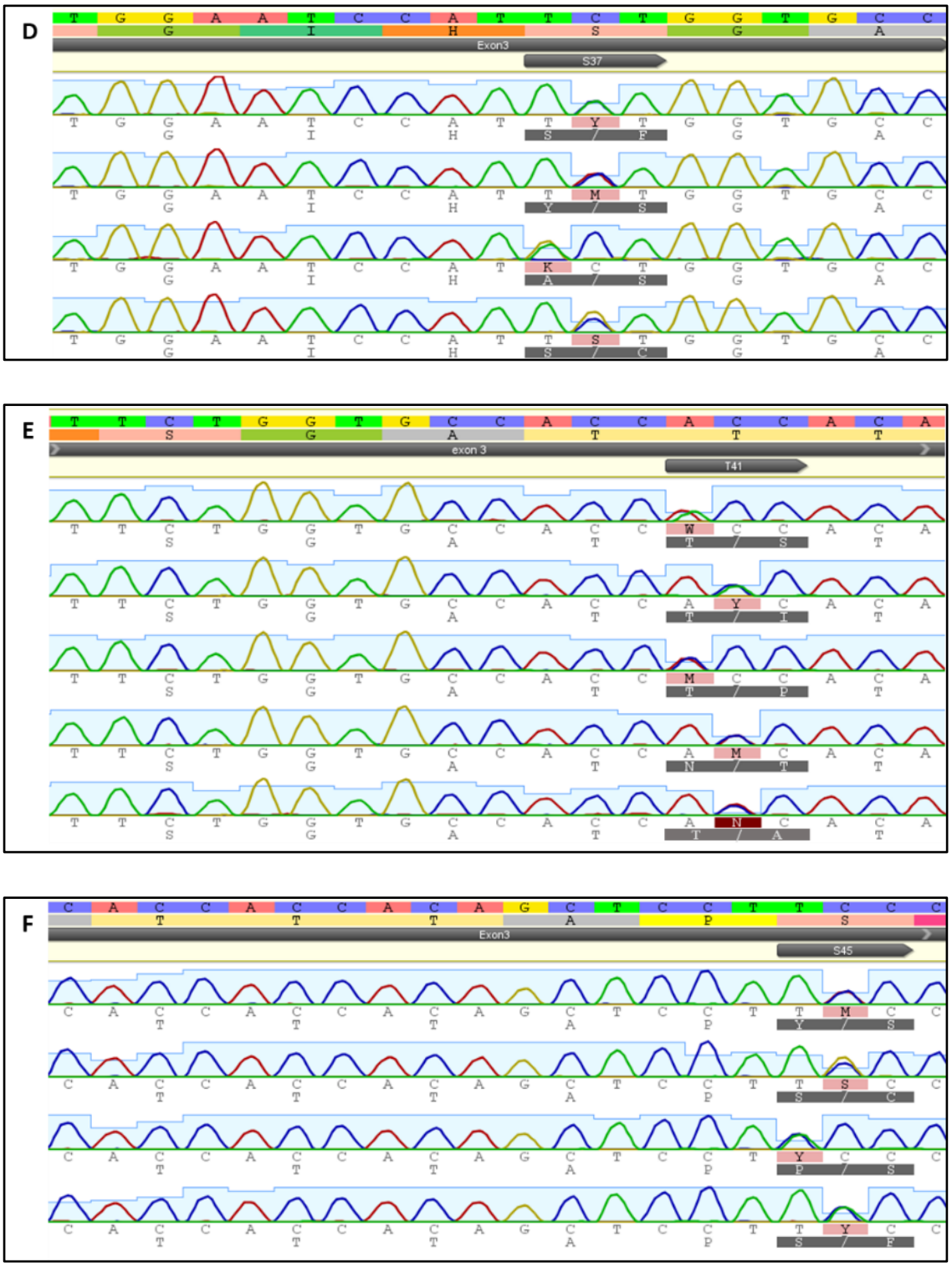


Figure 5-2: Sequence validation of heterozygous β-catenin mutations. (A-F) Representative image of the sequence analysis of the amino acid variants for each of the six residues.

5.2.3 Luciferase assay of E14 multiplex clones

Reporter gene assays based on luciferase system are commonly used to analyse gene expression and the strength of transcriptional regulatory elements such as promoters and enhancers. Such luciferase based systems have also been adapted to measure the Wnt/β-catenin activity. Currently, the most sensitive assay is the Super8XTOPflash/Super8XFOPflash which was generated in the Moon lab. These vectors are the modified versions of the pTOPflash and pFOPflash reporter constructs which were originally developed in Hans Clevers lab. The Super8XTOPflash vector construct consists of 7 sets of TCF/Lef DNA binding elements, upstream of a minimal TA viral promoter, which drive the expression of firefly luciferase enzyme and provides a sensitive and efficient approach for quantification of β-catenin activity. As a control, the Super8XFOPflash vector was generated in which these TCF/Lef binding sites were mutated. I decided to use this well establish system to compare β-catenin activity in different mutants.

Transient transfections were performed, where the Super8xTOPflash and FOPflash vectors were co-transfected with renilla luciferase construct. The signal generated from Renilla luciferase system was used as an internal vector control to normalize the signal for the cell numbers in each well. The transfections were performed in triplicates and the signal was measured 36 hours post transfection. Simultaneously, an additional set of transfections were performed for each mutant cell line to be used for treatment with DKK1.The Wnt pathway antagonist DKK1 acts by binding to the LRP6 co-receptor keeping the pathway turned off. Since the mutants generated were heterozygous, treatment with DKK1 would prevent any ligand binding and potentially abrogate the activity of WT allele, and provide evidence if the mutant allele is capable of constitutionally activating the pathway. The concentration of DKK1 for optimal suppression of WT endogenous β-catenin activity in E14 was previously optimized in the lab.

Six sets of luciferase assay were performed, with each set involving all the variants of a particular residue. In each plate, WT E14 cell line was used as a control, to make the comparison of each plate possible. As expected, both the cells transfected with the Super8x FOP flash and the untransfected controls gave very low luciferase signal. For comparison of the TOP signal from different mutant residues, the fold change was

calculated by taking the combined average (TOP/Renilla) of the independent clones, each normalized to WT E14 values from that particular experimental data set, and then plotted on a single graph (Fig 5-3). The β-catenin activity of the mutant clones differed when compared to the levels of WT E14, with certain clones of S37 and T41 mutants reaching over 800 fold increase in β-catenin activity. With the exception of T41S mutants, whose activity was similar to that of the WT E14 control, every other mutant had higher β-catenin activity levels. Furthermore, the range of β-catenin activity differed between the residues and also for the different amino acid variants for a given residue, clearly indicating the presence of a genotype-phenotype correlation among β-catenin mutants. The S37 residue showed the highest increase in activity levels along with few variants of T41 residue. This was followed by D32, G34 and S33; and S45 variants, which exhibited much lower increase in activity compared to the S37 mutants. Statistical analysis was done by performing one way ANOVA. This was followed by multiple comparison by Tukeys post ad hoc test, wherein a pairwise comparison of the 26 variants with one another and with E14 control was performed. Among the 351 possible combination, statistically significant differences ($p<0.05$) was observed between over half (180 pairs) of the compared pairs (Appendix 2A). The differences in β-catenin activity between the various amino acid variants for a given residue was significant, especially among T41 mutants, with T41I and T41A showing much higher activity than T41P and T41N and T41S.

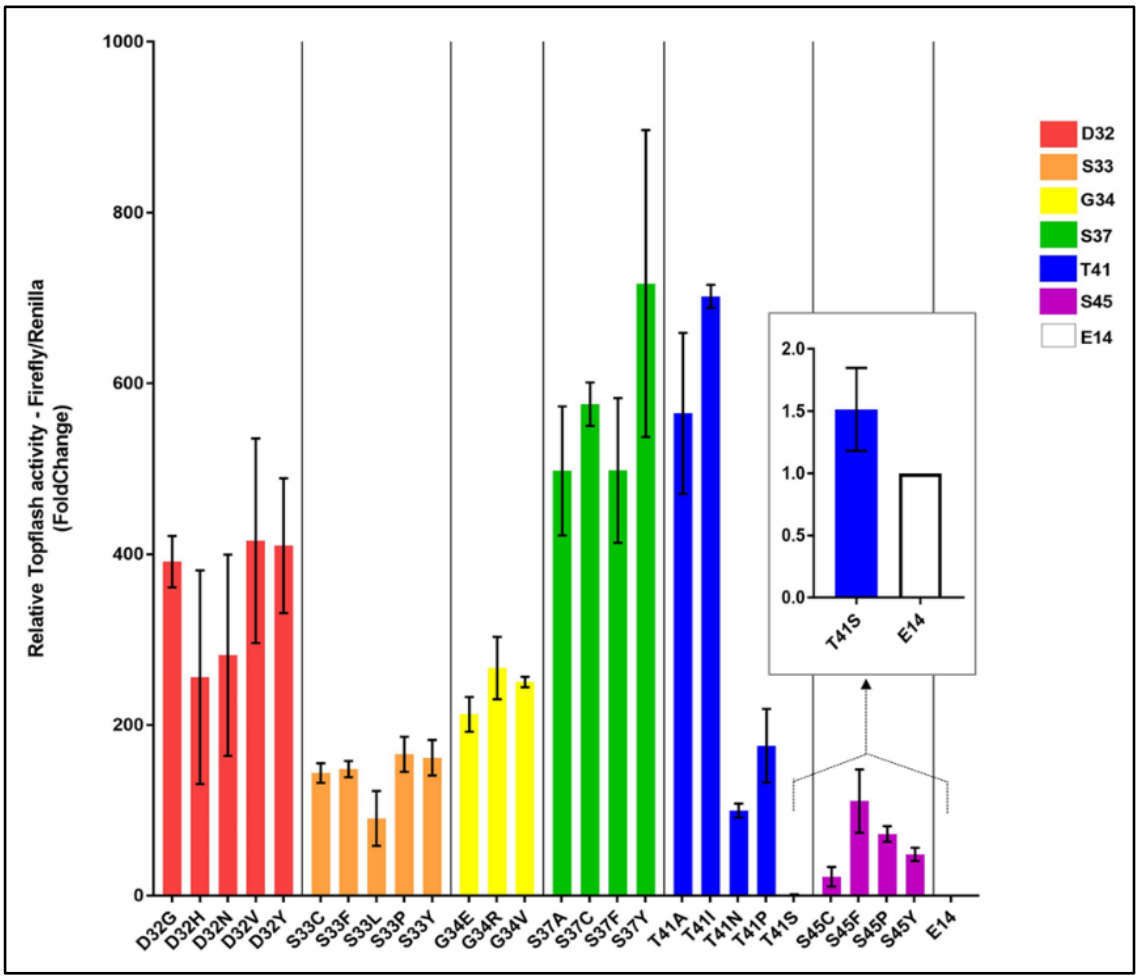


Figure 5-3: Analysis of β-catenin activity by Luciferase assay. Luciferase assay was performed for analysis of β-catenin activity among the various amino acid variants across the top six residues generated by multiplex targeting. Six sets of luciferase assay was performed and the TOP values were normalized to the E14 WT control. The values of the triplicate clones were combined and the average of the normalized TOP values were plotted for each of the mutant. Statistical analysis was done by performing one way ANOVA followed by multiple comparison by Tukeys post ad hoc test (Refer Appendix 2A). Renilla was used as internal vector control.

Next, the analysis of luciferase results from DKK1 treatment resulted in considerable reduction in the endogenous β-catenin activity in the WT E14 untreated control cell line in comparison to DKK1 treated samples (Fig 5-4).

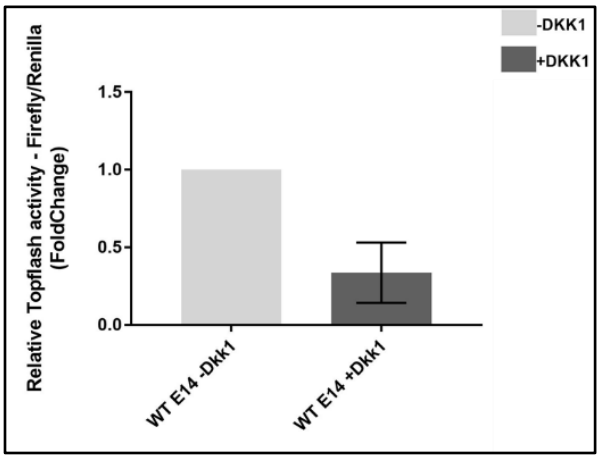
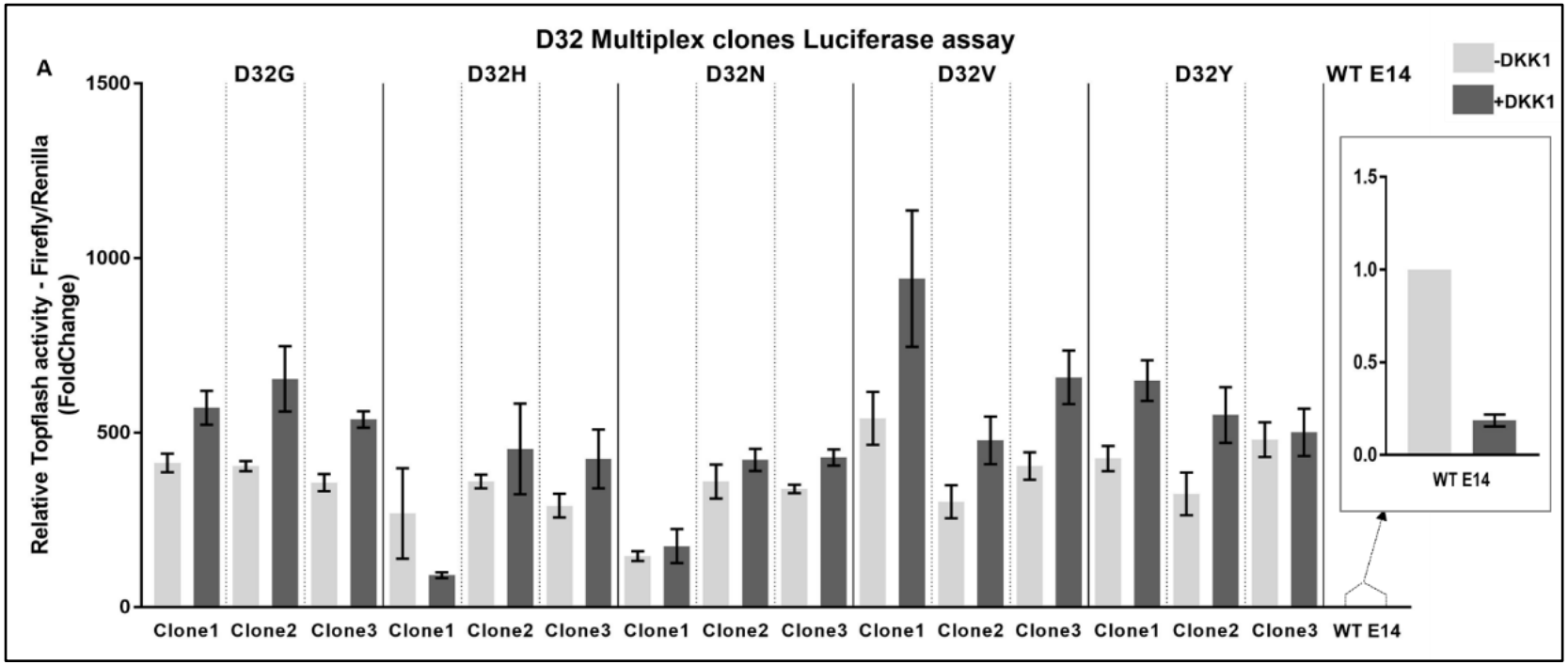
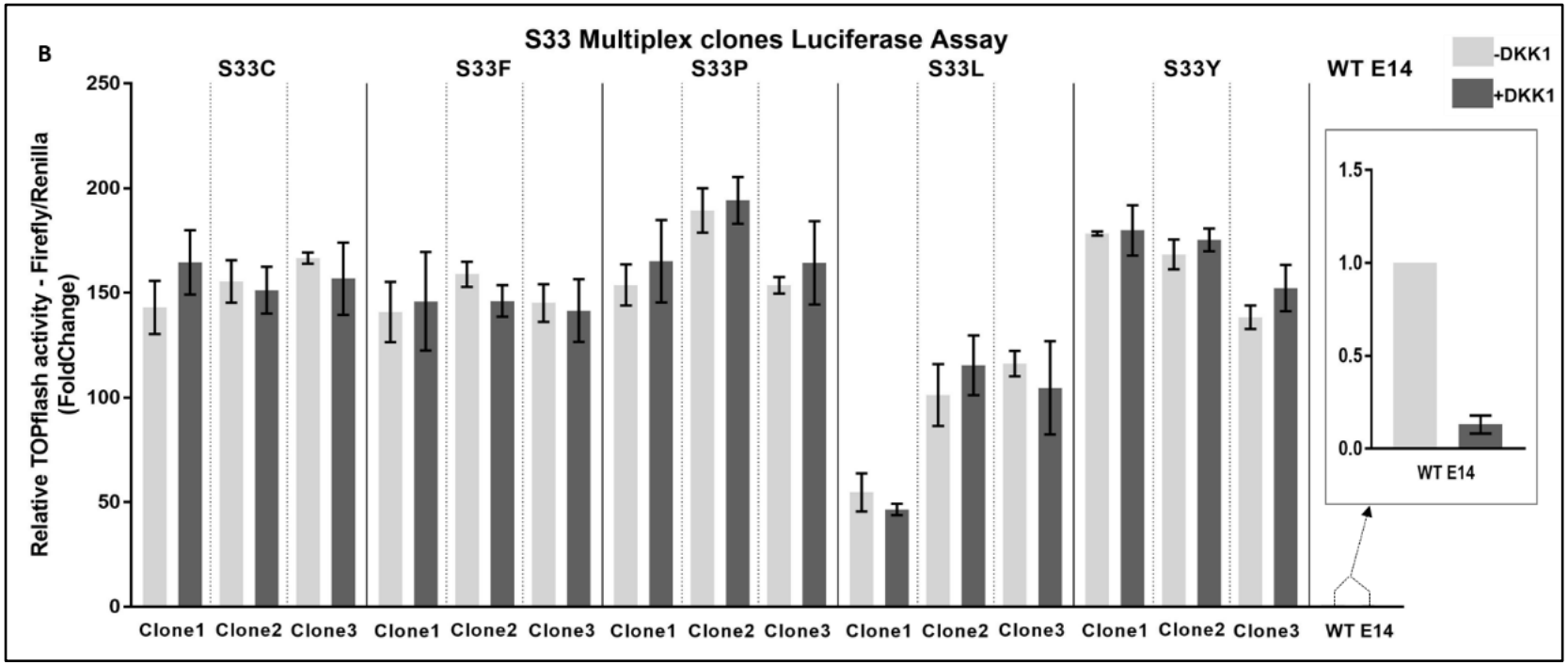
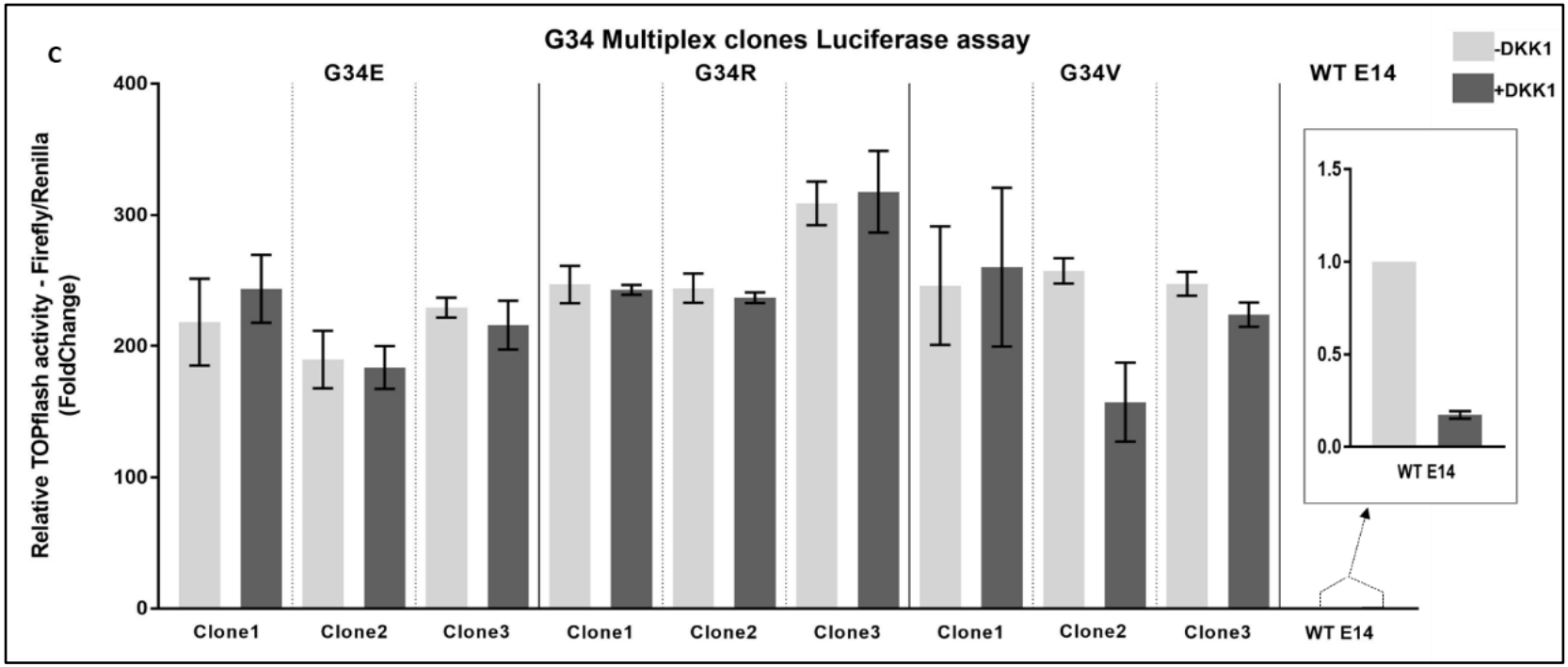


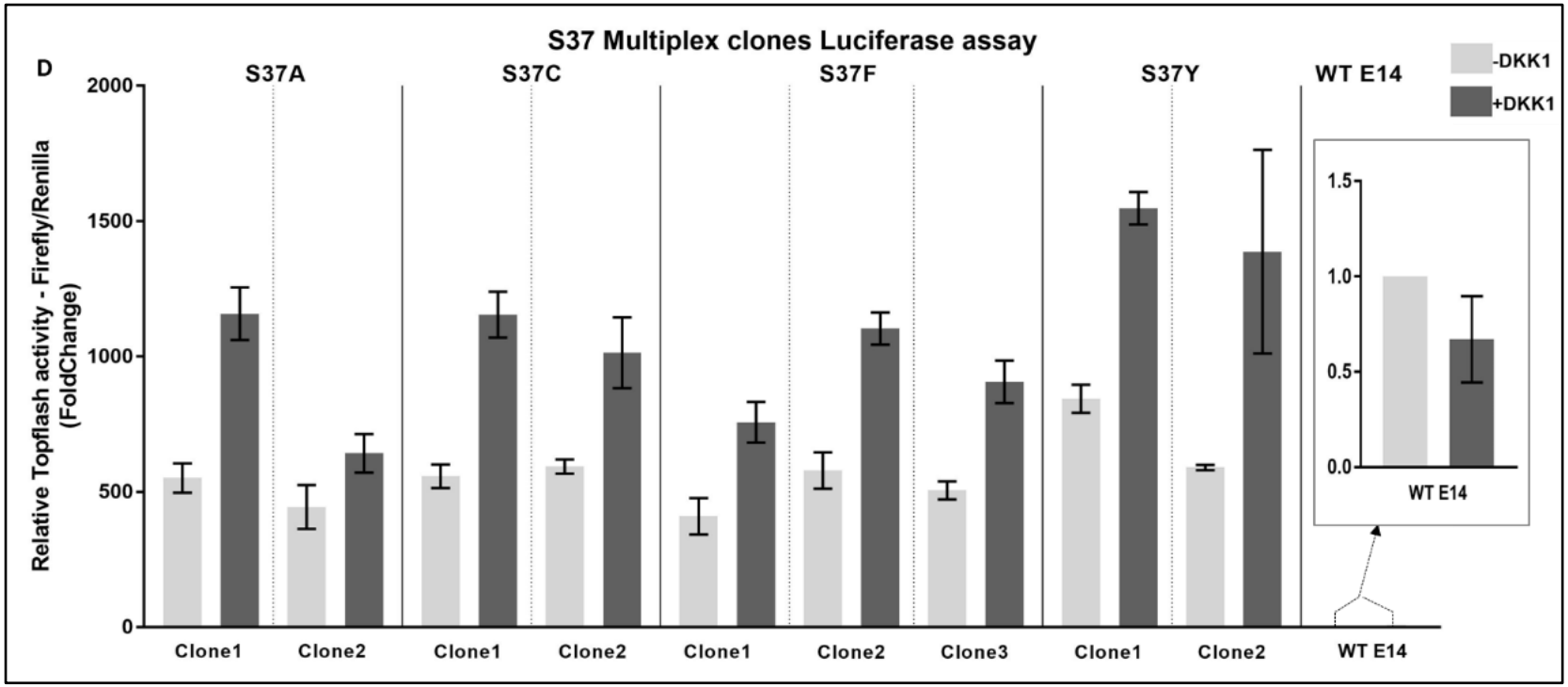
Figure 5-4: Luciferase analysis of WT E14. The β-catenin activity of the WT E14 cell were measured following treatment with Wnt antagonist DKK1. The luciferase measurements are from the 6 experiments performed for analysis of β-catenin activity of the amino acid variants for each of the 6 residues, in each case WT E14 was used as control. Error bars represent SD.

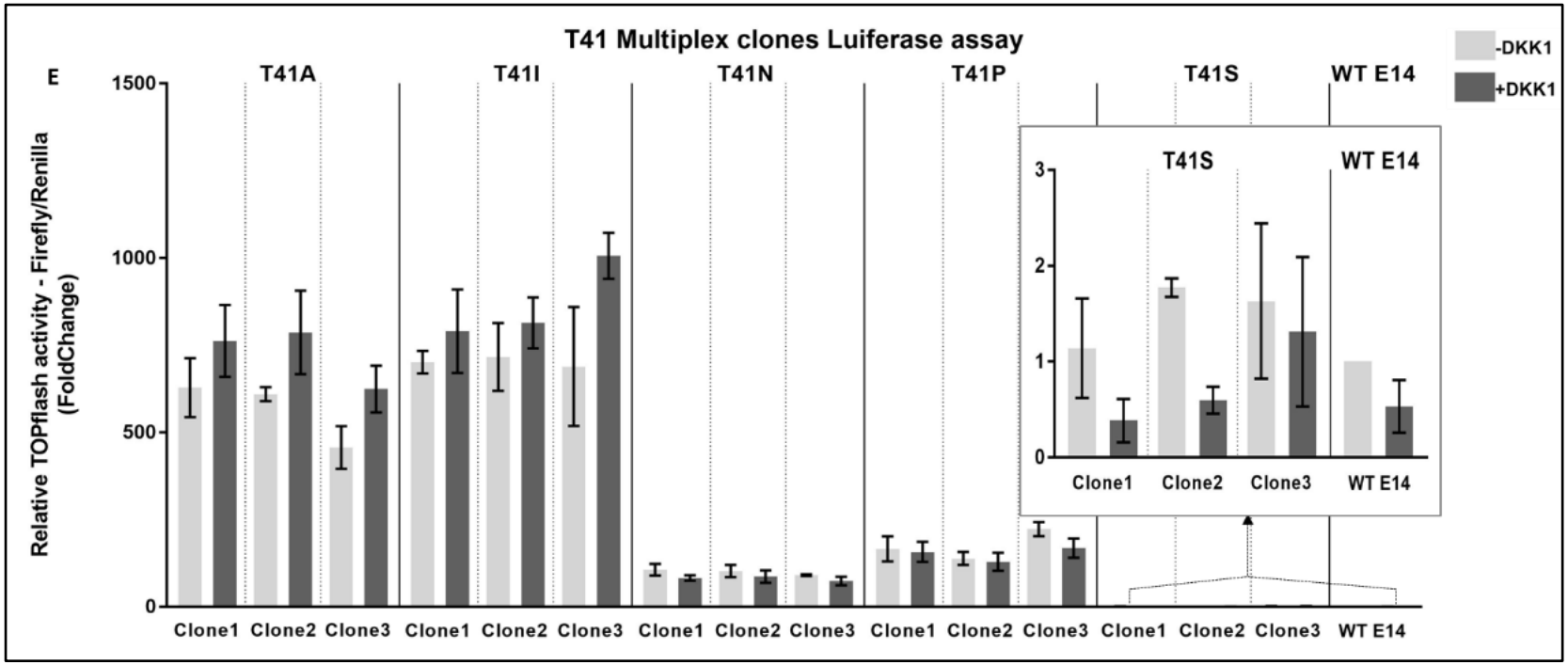
However, the effect of DKK1 varied across the mutant clones, and the independent clones for a particular mutation also showed differences in their response to DKK1. Hence, the activity of the independent clones for a particular mutation were not combined and the results are shown separately and in independent graphs. The fold change was calculated by normalizing the values (TOP/Renilla) of each clone to the WT E14 control. As observed in the graphs, clonal variation in response to DKK1 treatment was evident in G34, S33 and S45 mutants, with few clones showing slightly decreased activity, and few other clones showing increased activity even among the independent clones for a particular mutation (Fig 5-5B, C and F). However D32 (except for one of the clone), S37 and T41 I and A mutants showed consistently higher levels of β-catenin activity in response to DKK1 treatment (DKK1+) in comparison to their respective untreated controls (DKK-) (Fig 5-5A,D and E). The increase in the β-catenin activity, was more prominent especially among the S37 clones, where more than double the β-catenin activity was observed in majority of the variants, in response to DKK1.


D32 Multiplex clones Luciferase assay
A
Relative Topflash activity - Firefly/Renilla (FoldChange)
1500
1000
500
0
D32G
D32H
D32N
D32V
D32Y
WT E14
-DKK1
+DKK1
Clone1 Clone2 Clone3 Clone1 Clone2 Clone3 Clone1 Clone2 Clone3 Clone1 Clone2 Clone3 Clone1 Clone2 Clone3 WT E14
1.5
1.0
0.5
0.0
WT E14


B
S33 Multiplex clones Luciferase Assay
S33C
S33F
S33P
S33L
S33Y
WT E14
-DKK1
+DKK1
Relative TOPflash activity - Firefly/Renilla (FoldChange)
250
200
150
100
50
0
1.5
1.0
0.5
0.0
WT E14
Clone1 Clone2 Clone3 Clone1 Clone2 Clone3 Clone1 Clone2 Clone3 Clone1 Clone2 Clone3 Clone1 Clone2 Clone3 WT E14


C
G34 Multiplex clones Luciferase assay
-DKK1
+DKK1
G34E
G34R
G34V
WT E14
Relative TOPflash activity - Firefly/Renilla (FoldChange)
400
300
200
100
0
1.5
1.0
0.5
0.0
WT E14
Clone1
Clone2
Clone3
Clone1
Clone2
Clone3
Clone1
Clone2
Clone3
WT E14


D
S37 Multiplex clones Luciferase assay
S37A
S37C
S37F
S37Y
WT E14
-DKK1
+DKK1
Relative Topflash activity - Firefly/Renilla (FoldChange)
2000
1500
1000
500
0
Clone1
Clone2
Clone1
Clone2
Clone1
Clone2
Clone3
Clone1
Clone2
WT E14
1.5
1.0
0.5
0.0
WT E14


E
T41 Multiplex clones Luiferase assay
-DKK1
+DKK1
Relative TOPflash activity - Firefly/Renilla (FoldChange)
1500
1000
500
0
T41A
T41I
T41N
T41P
T41S
WT E14
Clone1
Clone2
Clone3
WT E14
3
2
1
0

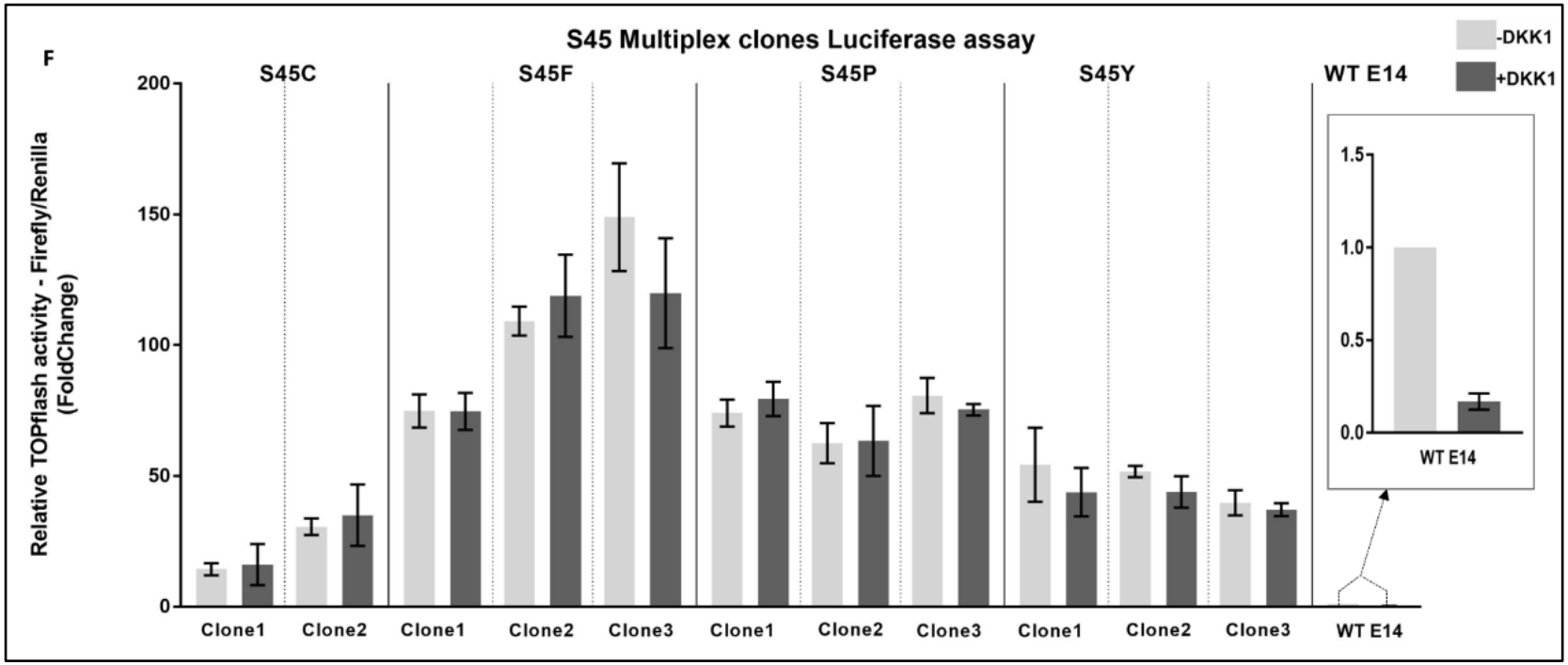


Figure 5-5: Luciferase analysis of DKK1 treated cell lines. The β-catenin activity of the various mutant clones was measured following treatment with Wnt antagonist DKK1. The relative Luciferase results of the DKK1 (+DKK1) treated samples were compared with the untreated control (-DKK1). In each case WT E14 was used as control. The error bars represent SD.

5.2.4 Comparison of luciferase with regression saturation data

In order to see if the luciferase data supported the results from the regression analysis of the saturation editing, I plotted regression values of these top 6 residues we used in multiplex targeting. As shown in figure 5-6, the activity levels measured by luciferase based system resembled the regression analysis in chapter 4, providing a convincing validation of the observed results.

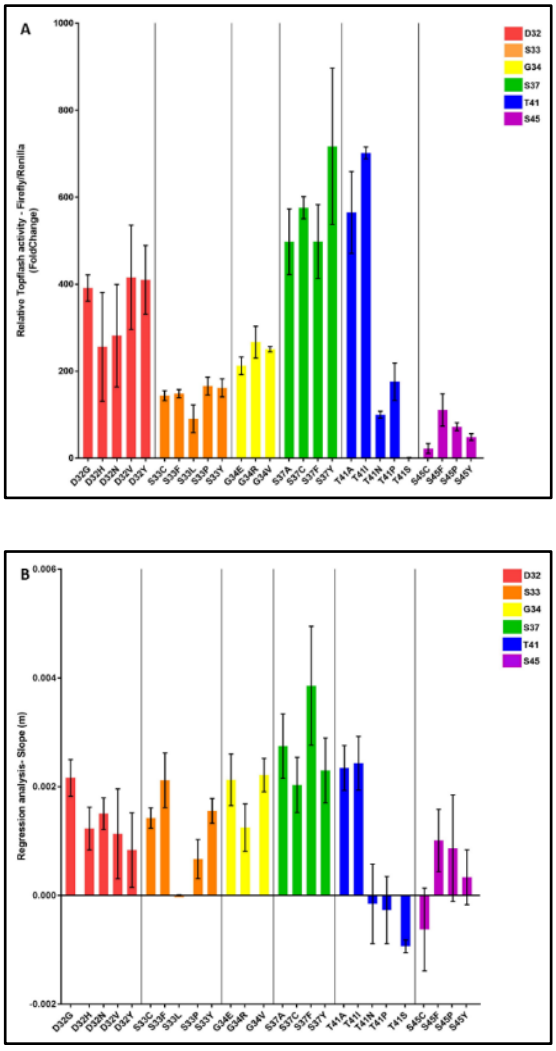


Figure 5-6: Comparison of luciferase with saturation data for multiplex clones. (A) β-catenin activity of the top six residues (E14 clones) generated by multiplex targeting, analysed using luciferase assay. (B) β-catenin activity of the top six residues (TCF clones) generated by saturation assay, analysed by the TCF reporter activity.

5.2.5 Taqman assay of E14 multiplex clones

The onset of gastrulation marks the specification of germ layers and involves various cellular transforming events characterized by spatio-temporal regulation in gene expression. The Wnt pathway is one of the important signaling cascades that plays an essential role in controlling a number of characteristic transition events of gastrulation including the β-catenin dependent direct/indirect regulation of multiple genes involved in the differentiation of the ectoderm, endoderm and mesoderm lineages. In addition, β-catenin is known to be essential for maintaining the pluripotency of ES cells. In this view, the various β-catenin mutant cell lines generated by multiplex targeting were analysed for the expression of few of the differentiation markers and genes involved in pluripotency network by performing Taqman assay.

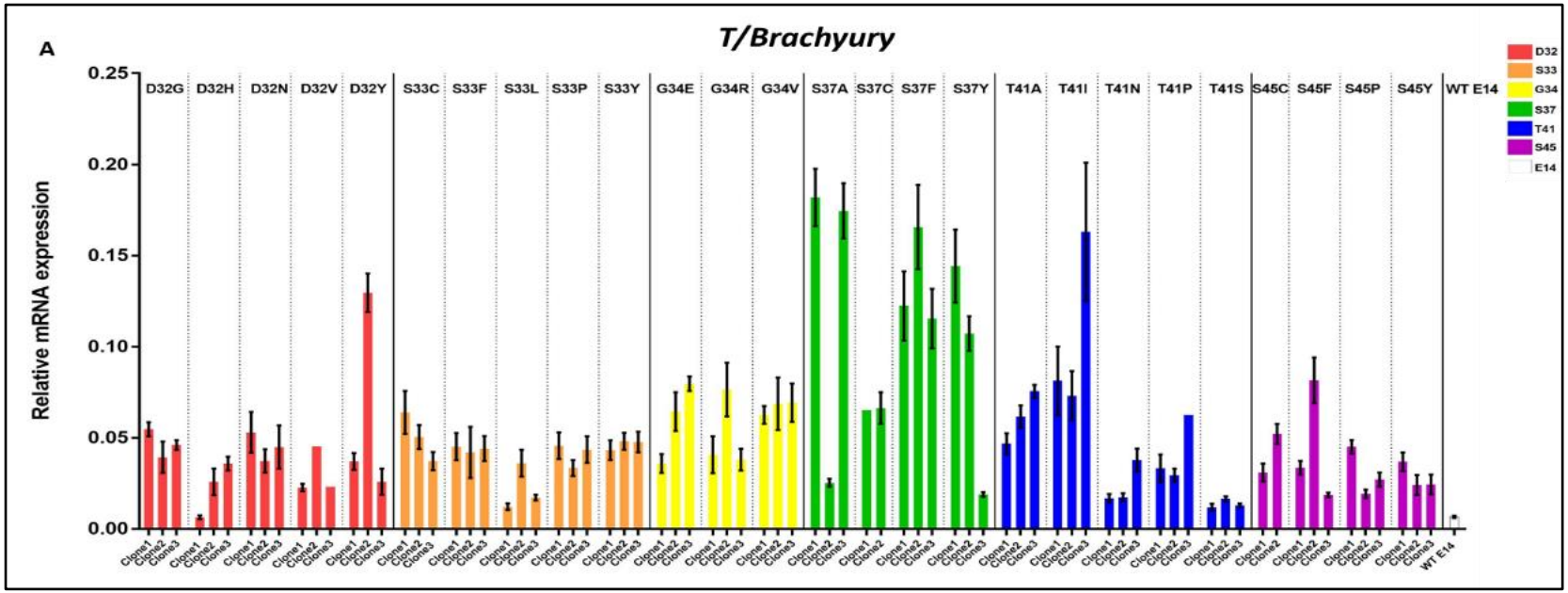
T/ Brachyury is a pan early stage mesodermal marker associated with the induction of EMT, and exhibits localised expression at the primitive streak and later in the notochord near the posterior/caudal end of the embryo. *Brachyury* expression is also known to be associated with induction of EMT. Similar to its role observed in embryonic development, *Brachyury* expression in mESCs has been reported to correlate with its in-vivo functions, with increased expression observed in differentiated mesenchymal like cells at the leading edge of EMT (Turner *et al.*, 2014). Taqman analysis of T expression showed heterogenous expression among the β-catenin mutants (Fig 5-7A). Although there exists clonal variation, overall comparison of expression indicated the S37 mutants having higher expression in comparison to the other residues and the control, possibly indicating a more mesenchymal like phenotype. *T/Brachyury* was identified as a target of Wnt/β-catenin signalling (Arnold *et al.*, 2000), and closely correlates with the β-catenin activity analysed by luciferase assay.

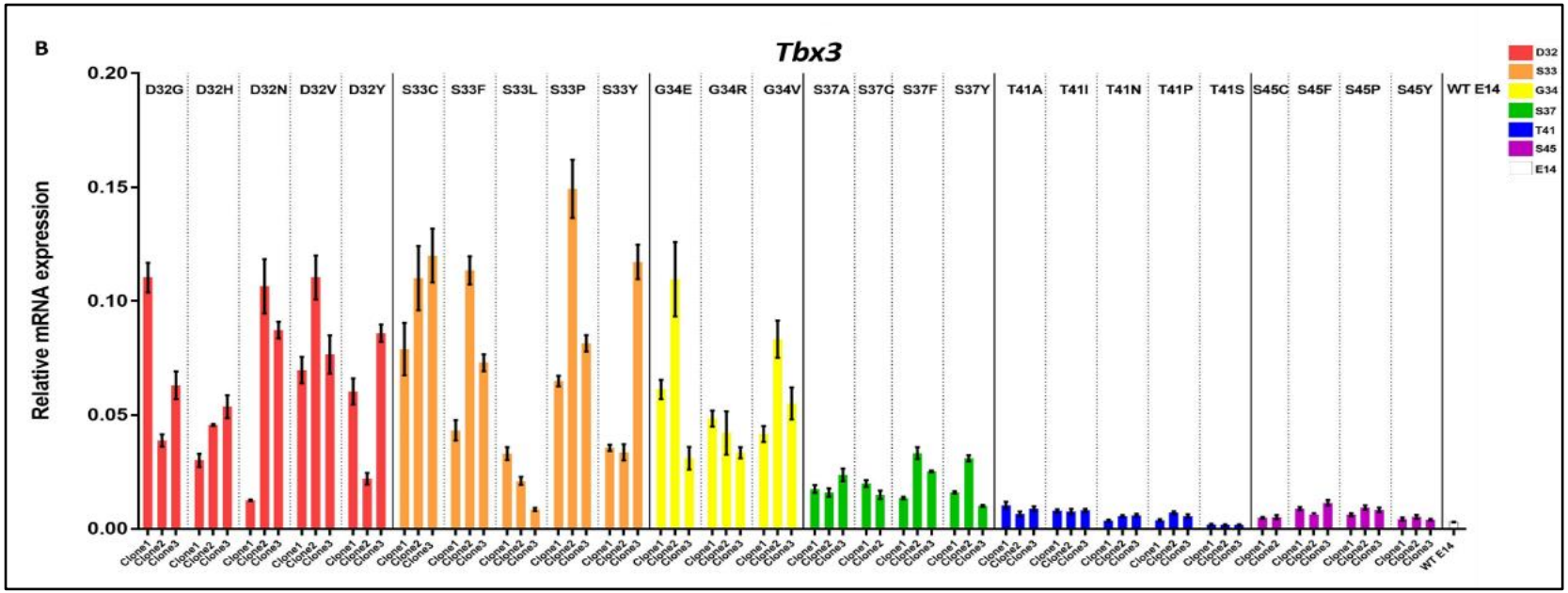
The T box protein *Tbx3* that promotes mesendoderm specification and is also important for maintaining pluripotency of mESCs, and the homeobox protein *Cdx1* known for its role in axial/vertebral patterning with expression later localized to intestinal endoderm, are both known to be direct transcriptional targets of canonical Wnt signaling (Subramanian, Meyer and Gruss, 1995; Silberg *et al.*, 2000; Pilon *et al.*, 2007; Renard *et al.*, 2007; Weidgang *et al.*, 2013; Russell *et al.*, 2015). Both these transcriptional targets displayed heterogeneous expression across the various mutants, with comparatively

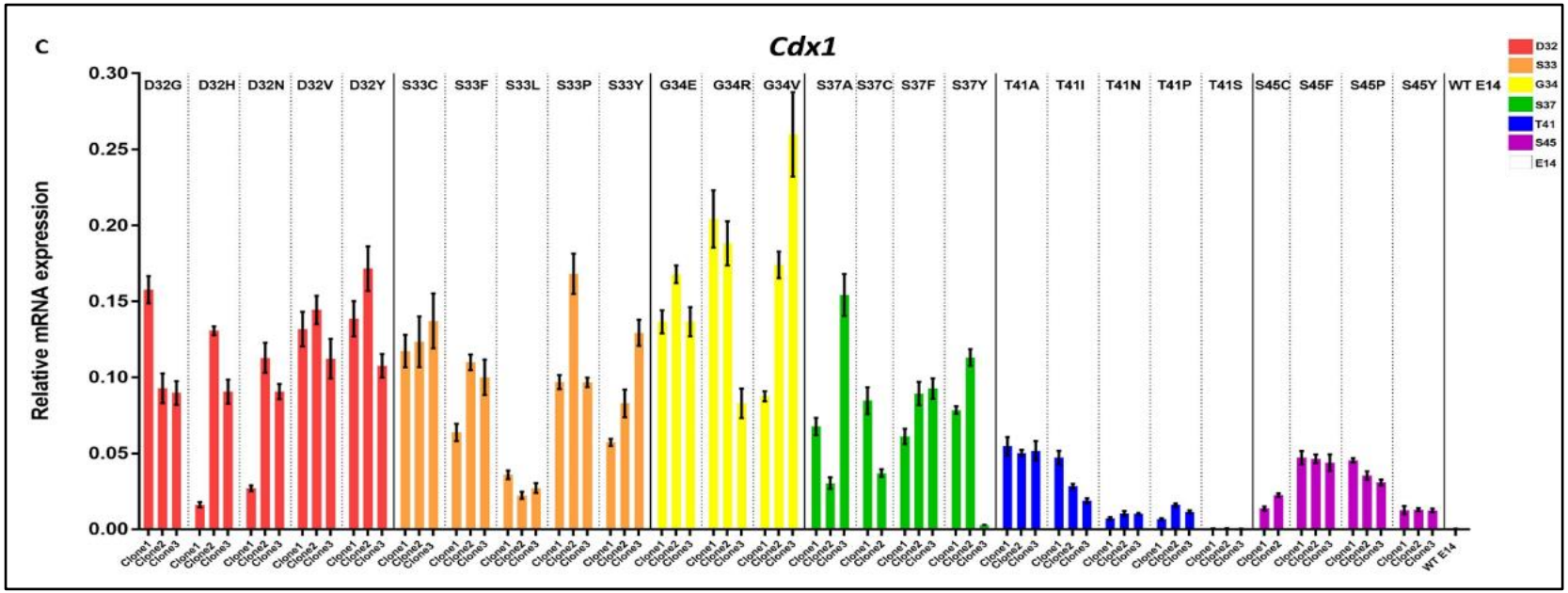
lower expression along the S37, T41 and S45 residues (Fig 5-7 B and C). The binding of TCF/β-catenin to the *Cdx1* promoter mediated by Wnt stimulation is shown to induce its expression (Lickert *et al.*, 2000). However the S37 T41I/A clones with increased TCF based reporter activity do not show a corresponding induction in *Cdx1* expression.

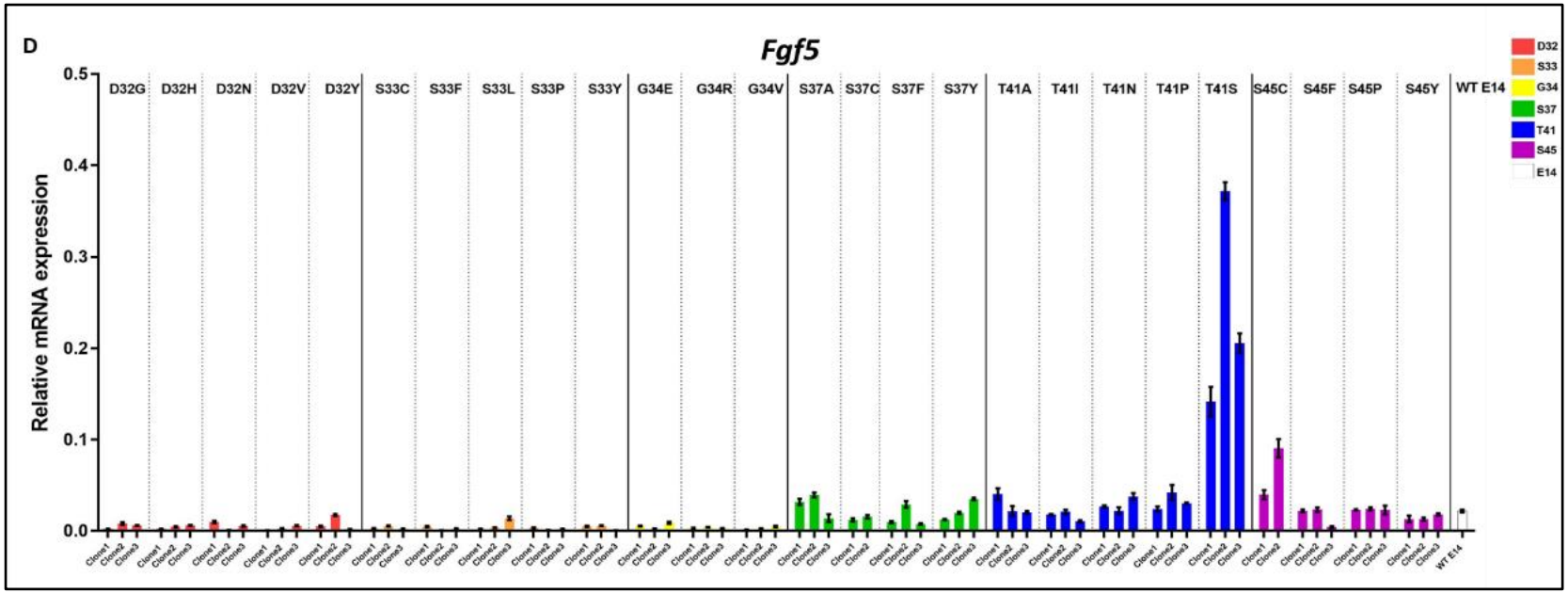
Furthermore, the expression of *Fgf5*, a marker for neural precursor that is expressed in the embryonic ectoderm was higher in T41S clones compared to the rest of the mutants, indicating a possibly differentiated clonal population (Fig 5-7D)(Hébert, Boyle and Martin, 1991). *Gata4,* a known marker of yolk sac endoderm and later known for its role in in the specification and differentiation of cardiac lineage, once again showed variable expression (Stefanovic and Christoffels, 2015). However, the canonical Wnt β-catenin signaling negatively regulates the expression of Gata transcription factors imparting an antagonistic effect on cardiogenesis (Afouda *et al.*, 2008). A similar reciprocal correlation between β-catenin activity and *Gata4* expression was observed in some of the mutant clones, with higher transcript levels observed for clones of T41S and S45C variants that were among the mutants with the lowest TCF reporter activity (Fig 5-7E).

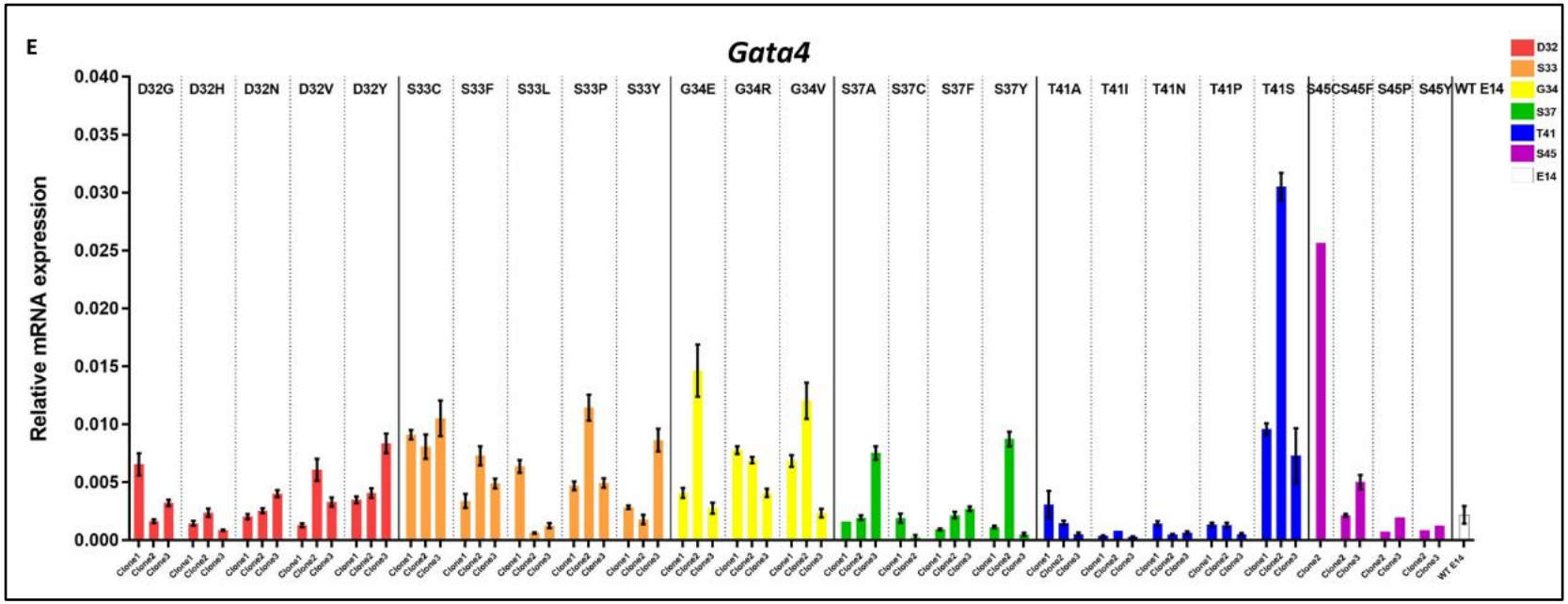
The classic pluripotency markers *Pou5f1* (encoding Oct4) and *Nanog* showed considerable but heterogeneous expression among all the mutant clones in comparison with the WT control (Fig 5-7F and G). In addition, the transcript levels of *Klf2*, another gene known to be involved in pluripotency also fluctuated among the mutants (Fig 5-7H). Finally, *Cdh1* (encoding E-cadherin) expression was slightly higher in D32 and S33 mutants compared to the other mutants and control (Fig 5-7I). Although there exists heterogeneity in expression, overall, the results of Taqman analysis indicate the presence of differential expression among the mutant clones for a number of genes analysed.

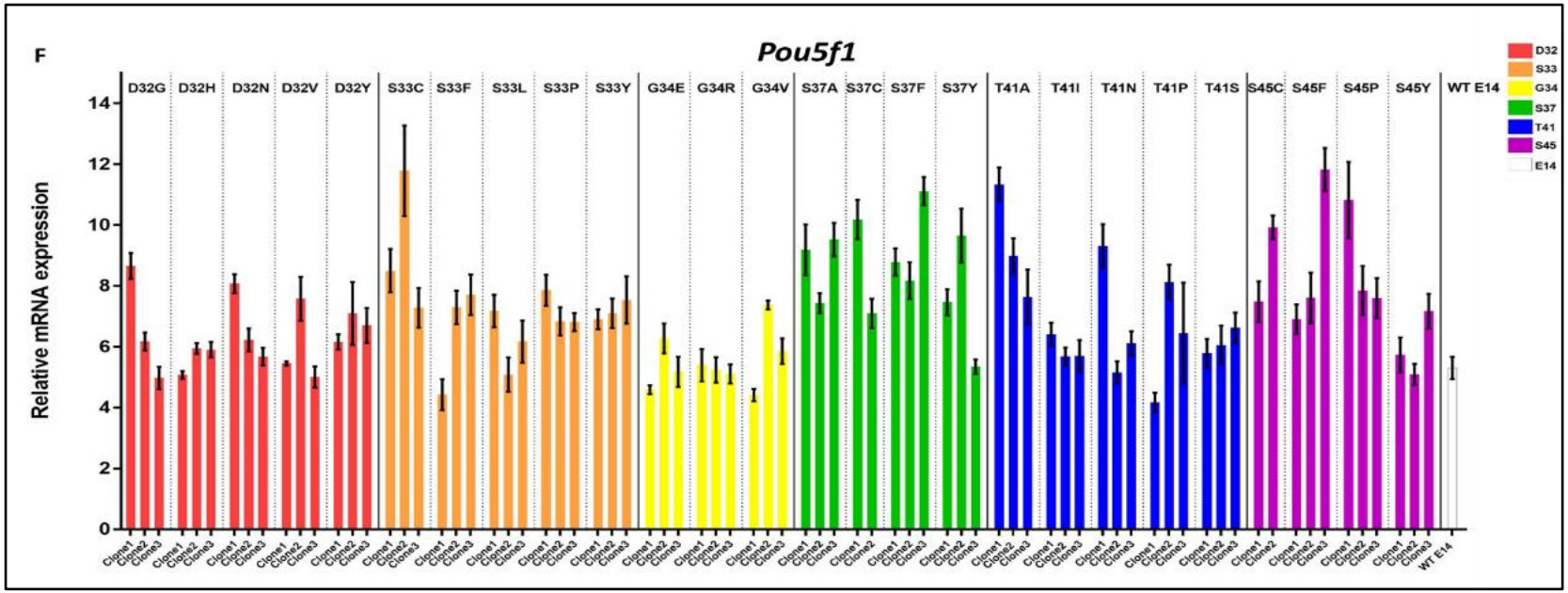

A
T/Brachyury
Relative mRNA expression
0.25
0.20
0.15
0.10
0.05
0.00
D32G D32H D32N D32V D32Y S33C S33F S33L S33P S33Y G34E G34R G34V S37A S37C S37F S37Y T41A T41I T41N T41P T41S S45C S45F S45P S45Y WT E14
D32
S33
G34
S37
T41
S45
E14

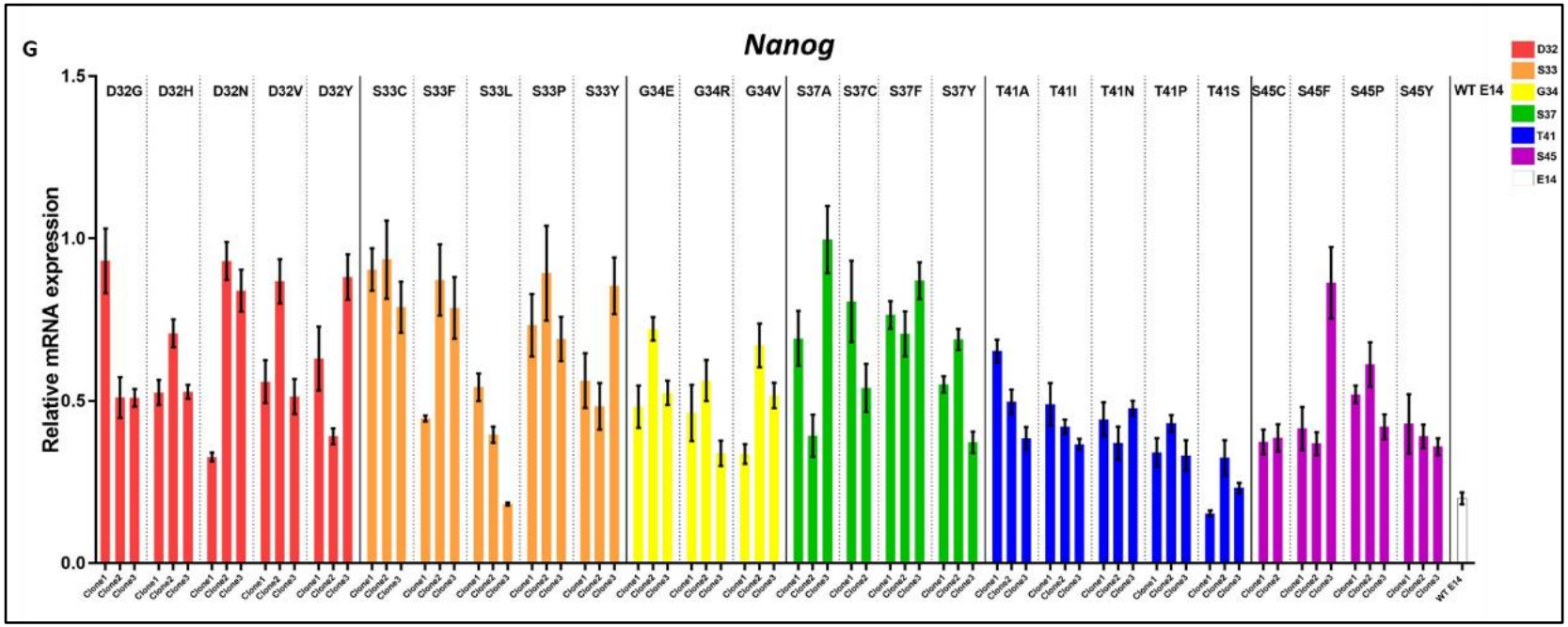

B
Tbx3
Relative mRNA expression
0.20
0.15
0.10
0.05
0.00
D32G
D32H
D32N
D32V
D32Y
S33C
S33F
S33L
S33P
S33Y
G34E
G34R
G34V
S37A
S37C
S37F
S37Y
T41A
T41I
T41N
T41P
T41S
S45C
S45F
S45P
S45Y
WT E14
D32
S33
G34
S37
T41
S45
E14

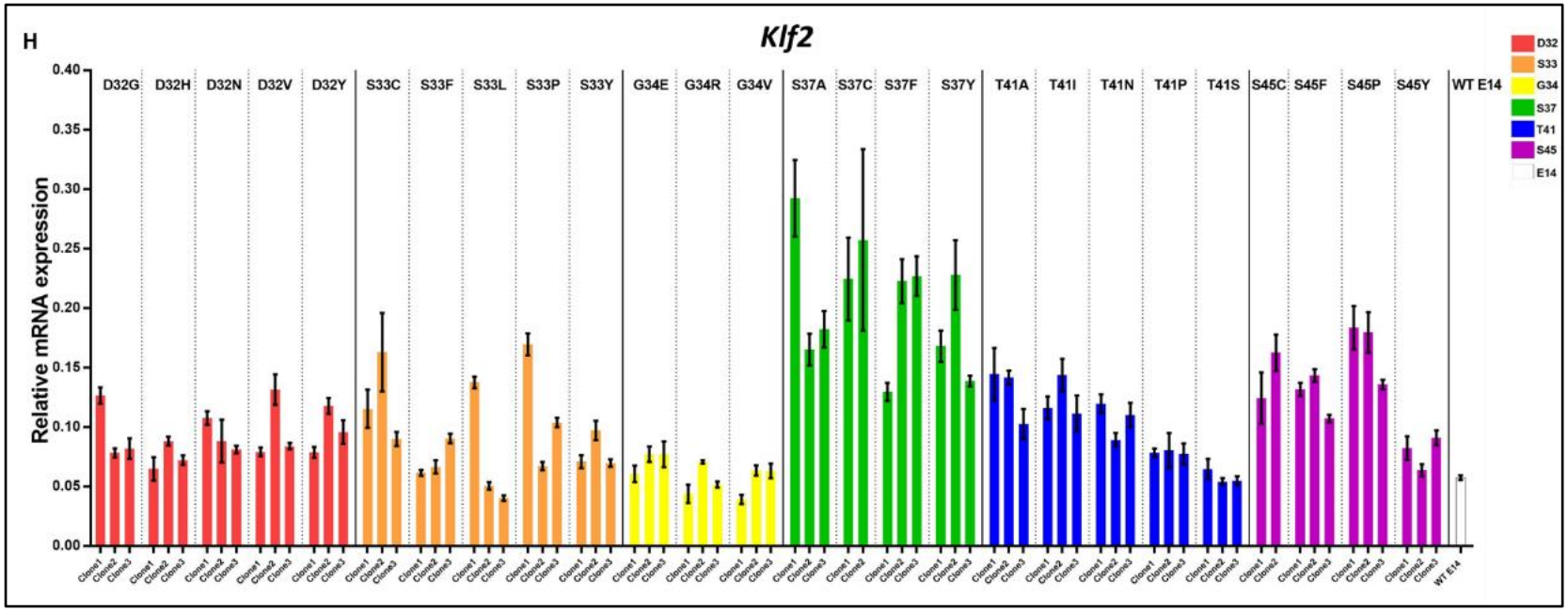

C
Cdx1
Relative mRNA expression
0.30
0.25
0.20
0.15
0.10
0.05
0.00
D32G
D32H
D32N
D32V
D32Y
S33C
S33F
S33L
S33P
S33Y
G34E
G34R
G34V
S37A
S37C
S37F
S37Y
T41A
T41I
T41N
T41P
T41S
S45C
S45F
S45P
S45Y
WT E14
D32
S33
G34
S37
T41
S45
E14


D
Fgf5
Relative mRNA expression
0.5
0.4
0.3
0.2
0.1
0.0
D32G
D32H
D32N
D32V
D32Y
S33C
S33F
S33L
S33P
S33Y
G34E
G34R
G34V
S37A
S37C
S37F
S37Y
T41A
T41I
T41N
T41P
T41S
S45C
S45F
S45P
S45Y
WT E14
D32
S33
G34
S37
T41
S45
E14


E
Gata4
Relative mRNA expression
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
D32G
D32H
D32N
D32V
D32Y
S33C
S33F
S33L
S33P
S33Y
G34E
G34R
G34V
S37A
S37C
S37F
S37Y
T41A
T41I
T41N
T41P
T41S
S45C
S45F
S45P
S45Y
WT E14
D32
S33
G34
S37
T41
S45
E14


F
Pou5f1
Relative mRNA expression
D32G
D32H
D32N
D32V
D32Y
S33C
S33F
S33L
S33P
S33Y
G34E
G34R
G34V
S37A
S37C
S37F
S37Y
T41A
T41I
T41N
T41P
T41S
S45C
S45F
S45P
S45Y
WT E14
D32
S33
G34
S37
T41
S45
E14


G
Nanog
Relative mRNA expression
1.5
1.0
0.5
0.0
D32G
D32H
D32N
D32V
D32Y
S33C
S33F
S33L
S33P
S33Y
G34E
G34R
G34V
S37A
S37C
S37F
S37Y
T41A
T41I
T41N
T41P
T41S
S45C
S45F
S45P
S45Y
WT E14
D32
S33
G34
S37
T41
S45
E14


H
Klf2
Relative mRNA expression
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
D32G
D32H
D32N
D32V
D32Y
S33C
S33F
S33L
S33P
S33Y
G34E
G34R
G34V
S37A
S37C
S37F
S37Y
T41A
T41I
T41N
T41P
T41S
S45C
S45F
S45P
S45Y
WT E14
D32
S33
G34
S37
T41
S45
E14

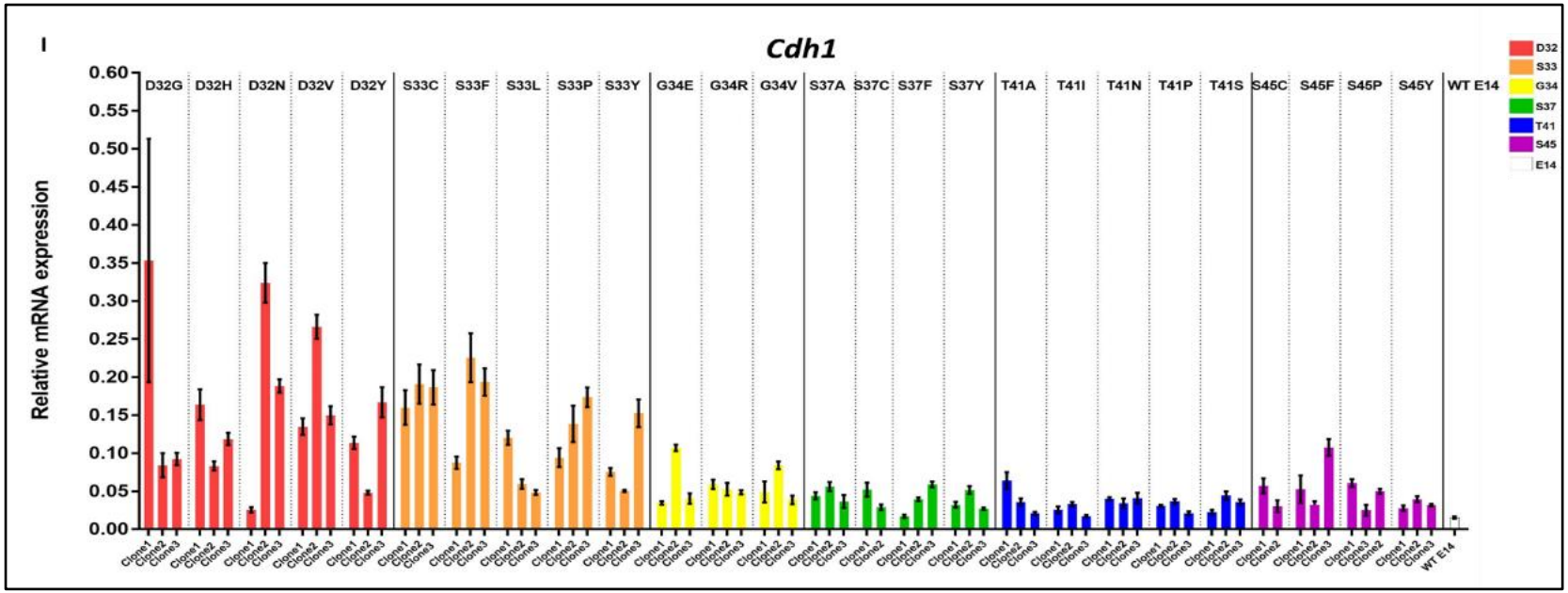


Figure 5-7: Taqman analysis of mRNA expression for markers of differentiation and pluripotency genes. (A-I) The multiplex clones were analysed for differential gene expression of selected markers. Triplicate (in some cases duplicate) clones were analysed for the various amino acid variants across the top six residues. The mRNA expression of the target gene was obtained by normalizing it to the relative expression of the housekeeping gene β-actin. The WT E14 cell line was used as a control. Error bars represent ±SEM.

5.3 Discussion

Mutations in the β-catenin protoncogene have long been studied in various tumours. However, generation of *in vitro* endogenous mutant cell lines has been a laborious task, hindering the possibilities of functional analysis of the various observed genetic anomalies. The advancements in gene editing techniques, and the ease of manipulating the endogenous loci with CRISPR/Cas9 system, provided us the opportunity to obtain a preliminary overview of the phenotypic consequences of few of the β-catenin mutations chosen from our analysis of COSMIC database.

One of the many advantages of using the CRISPR/Cas9 system is the convenience of combinatorial editing of multiple loci, or as in our case multiplex editing of a single locus of interest. However, some targeting strategies still require the use of large targeting vectors instead of single stranded short oligos, which is more laborious and less time efficient. The generation of 26 targeting vectors required for this project was a big challenge, however, with the strategy described in chapter 3 and 4, I achieved a cloning efficiency of 100 percent, making every colony picked correctly cloned. Using our strategy, within a very short space of time and effort, the required tools were generated.

Like in the saturation editing assay, I used a heterozygous β-catenin KO E14 cell line with CRISPR/Cas9 technology. Having guides specific to the selection cassette ensured keeping the WT β-catenin allele intact, resulted in generation of heterozygous mutant cell lines with an HDR efficiency of over 80 percent. I also used a multiplex approach for each residue, which meant that in only six targetings, I generated a total of 78 cell line for 26 mutants. The presence of clonal heterogeneity in mESCs required the analysis of multiple clones, and hence I chose to analyse triplicate clones for the majority of the mutants.

Initially, the endogenous mutant cell lines were analysed for differences in β-catenin activity by the widely used TOP/FOP luciferase assay. Although in an ideal scenario this analysis should have been performed all together in a single experiment, it was not a practical option for this case due to the large number of samples. Therefore the experiment was done in 6 batches, however, I included the same WT E14 control cell line in every experiment. The values from all samples were normalized to the control cell line, which made it possible to compare the results across all the mutant clones. All of the

mutants, except T41S, showed an increase β-catenin activity in comparison to the control. The Gsk3 being an S/T kinase explains the lower activity level conferred by the T41S variant. Similar to the saturation analysis, the activity from each residue was clearly different, with S37 being the highest activating mutation, the difference between S37 variants and S45 variants were very significant, confirming the phenotypical differences in these phosphorylation sites. In addition, T41 residue showed big variation among the amino acid variants we tested.

In addition, comparable results obtained from two different reporter systems TOP flash luciferase and TCF/Lef:H2B-GFP based on the same principle, performed in two separate mESC cell lines E14 and TCF, validates the differential Wnt activity conferred by the various mutants. These observed differences in degree of TCF dependent activation by the different β-catenin mutant alleles, strengths our hypothesis of genotype-phenotype correlation among β-catenin mutations. Next, the treatment with DKK1 inhibitor resulted in an unexpected increased activity level in certain clones, indicating a possible compensatory mechanism in D32, S37 and clones of T41. These mutant clones being heterozygous, we would expect the DKK1 to have an effect at least on the WT β-catenin, causing it to degrade in the absence of Wnt signal and thus resulting in reduced β-catenin activity. However, the mutant allele may be constitutively expressed and unresponsive to Wnt dependent activation. It can be speculated that the WT β-catenin in every clone might still be subjected to degradation in response to DKK1 treatment, resulting in reduced activity, similar to that seen in the WT EI4 controls. However, the decrease in β-catenin activity in the mutant clones might be compensated by the activity of the expressed mutant protein. This compensatory effect, may occur at varying rates and clones such as D32, S37 and T41 I and A with overall higher activity thus may have slightly higher compensatory effect resulting in the increased β-catenin activity observed in these clones in response to DKK1 treatment. However, the exact mechanism as to how these clones act differently to DKK1 treatment remains to be investigated.

Analysis of the target genes and the differentiation markers in these mutant cells would be a way to understand the effect of the individual mutations. Ideally this requires a much more detailed analysis, such as single cell RNA seq, however, the limitations in time and resources did not allow to do this. So I decided to perform a small pilot experiment using

Taqman assay, to investigate how the stem cell markers and the well-known targets of this pathway were affected,

Although β-catenin activity is required for stem cell renewal in mouse ES cells, it has also been shown that it promotes mesendodermal differentiation (Bakre *et al.*, 2007). Adding recombinant Wnt protein or GSK3β inhibitor to these cells, to increase β-catenin activity, resulted in differentiation towards mesenchymal lineage, even though the cells were kept in culture media which was shown to keep them in pluripotent undifferentiated state, and high β-catenin activity from these mutant allele could already be sufficient to result in some degree of differentiation. In addition to mesendodermal differentiation, increasing the β-catenin activity by overexpression of full length or the N terminal truncated forms of β-catenin, has been shown to induce neural differentiation in mES cells (Otero et al. 2004). Given the importance of β-catenin in facilitating lineage commitment, I therefore added some differentiation markers in my study.

The Taqman analysis of the differentiation markers that are also known to be the downstream targets of β-catenin showed variable transcript levels, not particularly correlating with the β-catenin activity levels. Furthermore, the marker genes showed a heterogeneous expression profile with fluctuations existing even among triplicates clones harboring the same mutations. However, comparison of the overall expression profile between the variants points towards a probably variability and existence of a differential expression that is either directly or indirectly regulated by the mutant β-catenin allele. The N terminal regulatory domain has been attributed to contributing to the stability of β-catenin and no known signaling or transcriptional regulatory activity is attributed to this domain. If not based on the stability dependent activation (with downstream target gene expression not particularly correlating with the activity levels), it remains to be explored how each of these β-catenin mutant alleles contribute to the observed variation in gene expression. In addition to the differentiation markers, the analysis of *Pouf51*, *Nanog* and *Klf2* revealed a considerable but heterogeneous expression of these pluripotency markers among all the mutant clones in comparison to the WT E14.

The Taqman analysis although provided clues on the existing differential gene expression among mutant clones, however to draw definitive conclusion on the role of each of the mutant alleles requires the analysis of the complete transcriptional profile, preferably at

single cell resolution. Also, each of the mutant alleles may have different or specific additional roles compared to that of the wild type (probably contributed by the changing binding specificities) and hence it is difficult and also inappropriate to directly extrapolate the phenotypic observations of the WT alleles onto their mutant counterparts, without performing a detailed analysis. The functional assays performed here only provide a preliminary perspective of the existing genotype-phenotype differences among the various mutants with prospects for in-depth analysis.

Chapter 6 Discussion

Discussion

Mutations in β-catenin gene have long been reported in multiple cancer types. These activating mutations result in the aberrant expression of Wnt target genes, many of which are known to contribute to various aspects of the tumourigenic process. Although being one of the most extensively studied pathways, there is however very little effort made towards understanding the molecular consequences of β-catenin mutations observed in these tumours. Nevertheless, there is compelling evidence in literature, suggesting that the current model for the regulation of β-catenin may not be sufficient to explain this complex pathway, and that there may be phenotypical differences among these β-catenin mutations. To be able to gain a better perspective of the phenotypic consequences of the observed mutational variants, in this thesis, I generated an *in vitro* system modelling these mutations, tested their potential in activating the pathway, and studied the genotype-phenotype correlation in β-catenin in cancer.

Before the *in vitro* analysis of the genotype phenotype consequences of β-catenin mutations, a comprehensive analysis of the spectrum of oncogenic *CTNNB1* mutations across the various cancer types was necessary. With this in mind, I initially analysed the various β-catenin mutations occurring across the different types of cancers using data compiled from the COSMIC database. As expected, the exon3 region of β- was observed to be the focal point of mutations (specifically between residues L31-G50), with the highest frequency of mutations at the phosphorylatable serine and threonine residues and the adjacent residues D32, G34. When I analysed the frequency of each of the mutated residue in every cancer type, I saw that there was a preferential selection for mutations at different residues in different cancers. I also observed that the bias was not restricted to the residues, but also present among the different amino acid substitutions across multiple different cancer types. This specific selection of mutations among different types of cancers present in statistically significant proportions already pointed towards the existence of a fundamental difference between these mutations, that was worthy of further investigation. To explore the observed genotype-phenotype correlation in β-catenin mutations two complementary approaches, saturation editing and multiplex targeting, were adapted.

Although overexpression studies have been widely used, and have contributed significantly in uncovering the functional importance of multiple oncogenes under various conditions in different in vitro systems, they still fail to recapitulate the native context of expression. For both our complementary approaches, we therefore decided to study the mutants in their endogenous context. The ease and simplistic adaptability of the CRISPR/Cas9 system, together with the added advantage to study the effect of the mutations in an endogenous context, made it the technique of choice for this investigation. However, at the start of this project not many details were available regarding the optimal conditions required for efficient targeting in different systems using the CRISPR Cas9 technology. Hence, I started with the optimization of the technique.

We and others, quickly realized that the generation of random insertions and deletions were much easier than the precise genome editing due to the active NHEJ pathway. The low efficiency of HDR was a set-back, especially for large scale genome editing projects such as ours, which required generation of multiple mutants. Although mESCs are known to have a higher HDR efficiency than other somatic cells, our initial targeting experiments yielded a low rate of HDR editing. By trying different transfection methods, and use of HDR enhancing drugs, I found the most optimal condition to edit mES cells.

One of the other advantages of using CRISPR/Cas9 technology, is the ability to use short single stranded oligos as HDR template. This eliminates the need to generate targeting vectors (TVs), which is usually the most time consuming part, especially in large scale projects. Therefore, this was also my first choice of DNA template in our targetings. However, as I started to test various guide RNAs for HDR efficiency, I noticed that, it was essential to have the guides cutting site as close as possible to the desired edit. The efficiency of HDR was decreasing to almost zero percent only after about 8bp far from the cutting site. As I was aiming to edit every amino acid in 20 amino acid region, using ssDNA as the DNA template did not seem to be a plausible option. I therefore tested using vector as HDR template, and achieved a good efficiency of HDR in the distant residues as well.

Due to the bialleleic nature of CRISPR cutting and an active NHEJ repair mechanism, generating clean heterozygous mutations proved to be a challenging task as well. I wanted to perform saturation editing in heterozygous condition, not only because this

would be more physiologically relevant, but also would ensure that the activity difference is due to one specific mutation and not due to additional indel mutations in the other allele. This problem was finally resolved by generating a heterozygous β-catenin KO cell line in which one of the β-catenin allele was replaced by puDeltatk selection cassette. Using this counter selection strategy along with TVs as HDR template proved to be a successful combination, and helped us achieve a high rate of HDR throughout the region in both approaches.

The CRISPR nuclease system although provides a highly efficient on target activity, they are also known to have off-target effects. Due to time constraints I was unable to check the off-target effects of the guides used in this project, which will need to be done in the future. Since the introduction of the nuclease mediated genome editing, various methods have been adopted to identify the off-targets. The guide designing software have included off-target prediction score or rank and based on these scores the loci with predicted off-targets can be selected, PCR amplified and sequenced using Sanger sequencing to check for off-target activities of the guides. However, the computational prediction methods are not absolutely fool proof in precise detection of every possible off-target, and better options include whole genome sequence techniques for detection of off-targets. Various methods have been developed including GUIDE-seq, Digenome sequencing and BLESS (Zischewski, Fischer and Bortesi, 2017). Since the same guides were used for both saturation and multiplex strategies, testing of either the pooled DNA from saturation assay or the DNA from few of the independent clonal cell lines generated by multiplex targeting will provide evidence of the specificity of the guides used in our experimental system.

The first approach of saturation editing of the β-catenin hot spot, was performed to analyse the β-catenin activity levels of all the amino acid variants across L31-G50, allowing us to compare the activity levels of both cancerous and non-cancerous mutations. The analysis of the proportion of variants for each residue across the different sorted population for the first time provided evidence of the of allele-specific activity levels conferred by the different endogenous β-catenin variants. The differential activity response were mostly confined to the phosphorylatable serine and threonine residues and the residues D32 and G34 that are known to be a part of the degron motif. These six

residues also being the most frequently observed residues in our analysis of COSMIC database, strongly highlights the β-catenin activity levels to be a major criterion for selection of these mutations in the tumourigenic process.

As expected, the synonymous substitutions of residues (D32D, S33S, G34G etc) gave no increase in the activity levels. The phosphorylatable S and T when interchanged to T or S were seen to produce no increase in β-catenin activity levels. This is in accordance with the priming and sequential model of phosphorylation. However, according to the priming model the substitution of S45 to other amino acid that prevents its phosphorylation should block the GSK3 dependent prevent phosphorylation of the other three residues in this sequential cascade, and in turn prevent the ubiquitin ligase degradation, leading to increased β-catenin activity. Although the S45 variants are capable of increasing the activity levels, the overall strength of activity is much lower when compared to the activity levels conferred by T41, S33 and S37 residues, indicating the presence of additional mechanism of regulating β-catenin activity even in the absence of priming. The phosphorylation of both S37 and S33 have been shown to form a docking site for the E3 ligase, however the activity difference of mutations at these two residues suggests the possibility of the two residues being differently regulated. In addition to the phosphorylatable residues, the D32 and G34 variants were able to increase the activity levels, which confirms the importance of these two residues in β-catenin regulation.

Furthermore, variability in activity levels was evident among the amino acid variants each of the six residues. For example although the overall activity of residue T41 was lower than that of S37, individual amino acid variants such as T41I and T41A were capable of increasing the activity levels in the highest intensity spectrum. Using the regression analysis to assign a score of overall activity for each of the amino acid variants, we were able to confirm the observed differences in the activity levels for the individual substituents for each of the residues from L31-G48. This allele-specific variability indicates a differential binding strength of the amino acid variants with their interacting protein partners. Overall, these results draw focus on the molecular and biochemical properties governing the regulation of TCF/Lef dependent β-catenin activity that still remain unexplained.

Next, the analysis of the mutational effect vs the frequency of mutations (compiled from COSMIC database) of the individual tumour types showed a distinct pattern indicating a tissue-specific selection of mutations based on the mutational effect. Majority of the tumour types selected for mutations that include a variable range of mutational effect. The difference in activity based on tumour subtype, presence of additional mechanism that could increase the β-catenin effect and association of specific β-catenin mutations with particular genetic alterations might be some of the possible factors contributing to the selection of distinct pattern of mutations with low, medium and high mutational effect observed in the individual tumour types. Although in this study we did not categorize the cancers according to the sub type, doing so would determine if the β-catenin activity is indeed subtype specific. In addition, analysis of the whole genome sequencing data for each of the tumour samples would help us understand if particular β-catenin mutations are associated with specific genetic alteration.

The analysis of the background mutational rate for the endometrial and liver tumours showed that likelihood of amino acid substitution does play an important role in determining the presence of a mutation. However, the tissue specific selectivity of mutations in these tumour types does not entirely correlate with the likelihood of amino acid substitution, suggesting that the Darwinian selection based on functional significance has a much larger influence in contributing to the observed tumour type specific mutational bias. This analysis of the background mutational rate was restricted to liver and endometrial tumour due low number of sample size for other tumour types in the TCGA database. In the future, as more and more whole genome or whole exome sequences of tumour samples are available, it will be possible to analyse the background mutational profile of the various other tumour types, which will give a better perspective of the contribution of the mutational processes in the observed tissue specific mutational bias.

These mutations being specific to human cancers, the saturation mutagenesis screen could be extended to human cell lines. Similar to mESCs the role of Wnt signaling in self renewal and differentiation have been extensively studied in hESCs, making it a suitable in vitro model for studying these β-catenin mutations. Wnt Reporter hESCs such as those generated by Nusse lab with a similar TCF-GFP system that allows FACS based sorting

could be used to perform a similar saturation mutagenesis assay (Blauwkamp et al 2012). In addition, tissue specific cell lines can be used to generate the mutations observed in that particular tissue type, which will not only provide a more physiologically relevant analysis of phenotypic differences in terms of β-catenin activity levels, but will also help in understanding the differences in tumourigeneic potential of mutational variants. These mutational variants when modelled in specific tissue type could be used for performing various conventional *in vitro* transformation assays including soft agar assay (for analysis of anchorage independent growth), focus formation assay (for analysis of loss of contact inhibition), scratch assay/transwell migration assay (for analysis of migration and invasion potential) and BrdU assay for cell proliferation. Understanding the tumourigenic potential of these β-catenin variants through such experiments will provide better insights into the disease pathogenesis.

In addition to *in vitro* analysis, it would be essential to perform *in vivo* analysis by generating conditional mouse models. Few interesting mutations could initially be selected to generate mouse models. For example, based on our screen, the two I35 variants selected for by salivary gland and liver were seen to have variable mutational effect. The I35T variant observed in the salivary gland conferred a low mutational effect compared to I35S variant selected for liver cancers that yielded a higher mutational effect. Conditional expression of I35T and I35S in salivary gland and liver using tissue specific CRE would provide better insights into the phenotypic advantage leading to the observed tissue specific selectivity. Traditionally, mouse models with N terminally truncated β-catenin forms (involving exon 3 deletion) have been used to replicate the constitutively active form of β- catenin (Brault *et al.*, 2001; Huelsken *et al.*, 2001) . However, the allele specific activity levels of the various β-catenin variants observed in our study questions the credibility of using these deletion mutants as a general model for studying the various disease phenotypes, and emphasizes on the necessity of modelling the specific mutation (also taking into account the amino acid substitution) to draw more physiologically relevant conclusions.

In the second complementary approach the endogenous β-catenin heterozygous clonal cell lines were generated to perform functional analysis. Even before the functional analysis of the mutant clones, there were several observable differences in morphology.

Although due to time constraints the detailed analysis was not possible, the initial microscopic observation of the multiplex clones revealed a difference in morphology, specifically evident among the T41 and S45 variants. Majority of the mutants from D32, S33, G34 and S37 resembled cells being cultured in 2i with a more compact and homogeneous morphology, however T41N/P and S45C/Y mostly resembled the wild type E14 phenotype with a more scattered and heterogenous appearance. The T41S variants started to differentiate and increased cell death was observed after few passages of being cultured in normal ES media and had to be shifted to 2i supplemented media for continued survival and growth. In addition, the more compact clones seem to grow at a slower rate when compared to the clones with a heterogeneous morphology. β-catenin is known to be important for both maintenance of stemness and self-renewal of ESCs and further experiments to include a detailed analysis of these clones by performing alkaline phosphatase staining for analysis of the differentiation state and also analysis of the proliferative index and rate of apoptosis of these clones will be needed for better characterization of these observed differences.

The Luciferase assay on multiplex clones once again showed an allele specific β-catenin activity level with a pattern similar to those observed by saturation assay. The S37 variants were among those that conferred the highest activity, and the activity levels exhibited by the variants of the priming S45 residue were much lower in comparison. In addition, the β-catenin activity between the various amino acid variants for a given residue also varied and a significant difference in the activity levels was specifically observed among the T41 variants. The T41S variant was the only mutant among the multiplex clones not yielding an increase in activity level, and its activity was similar to that observed in the WT E14 cell line. Given that GSK3 is a serine/threonine kinase, this was expected. Although having activity levels comparable to WT, the inability of these T41S mutants to survive in culture without being supplemented by 2i, is unexpected and cannot be explained based on the current knowledge of regulation.

The Taqman analysis of multiplex clones resulted in a heterogeneous expression among the triplicate clones for majority of the genes analysed, however there was still evidence of differential expression among the β-catenin mutants. For example, the gene expression profile of differentiation markers T/Brachyury was specifically enriched in the

S37 variants. Differentiation assays towards mesodermal lineage would show whether or not these S37 mutants have a preference to differentiate into mesodermal lineage. In addition, the β-catenin mutants could be differentiated into neuronal or cardiomyocyte to understand if the different mutations confers selective advantage to differentiate towards specific lineages.

Furthermore, it is necessary to analyse the expression of additional β-catenin target genes either by performing Taqman assay or by using other conventional techniques such as microarray, both of which allow analysis of gene expression profile of known genes. However, analysis of the complete transcriptome profile using assays based on the second and third generation sequencing platforms including RNA-seq and CAGE-seq, preferably at single cell level will provide higher sensitivity quantification of expression. The analysis of the complete expression profile, will not only help to quantitate the expression of known genes, but will also help uncover novel genes regulated by these β-catenin variants.

Both the saturation screen and multiplex assay was performed in heterozygous condition, however the saturation vectors and the puDeltatk system have been designed so as to allow generation of hemizygous mutant clones. Using g9B CRISPR and targeting the WT β-catenin allele and keeping the puDeltatk selection cassette intact in the heterozygous β-catenin KO clones, would provide a second option of generating hemizygous clones. Although the DKK1 on multiplex clones showed that the activity of the mutant allele to be dominant over the WT allele, analyzing the variants in hemizygous condition would further help in analyzing the strength of activity confined to the mutant allele. Also, the presence of the WT allele in the heterozygous clone might be a complication for downstream experimental approaches, and hence, would require the generation of either hemizygous or homozygous mutants.

In the intial round of multiplex targeting using ssODN, I have successfully generated homozygous S45 F, C, Y and P mutants, and these can be used as a starting point. These S45 homozygous mutants could be used to test the phosphorylation state of mutant protein. Wang et al have shown that the T41, S37 and S33 residues can still be phosphorylated in S45 mutant cells (Wang, Vogelstein and Kinzler, 2003). This can be tested by performing western blot using phospho-specific antibodies. In addition, it has

been observed that the S45 variants differ in their localization (Austinat *et al.*, 2008a). For the purpose of analyzing the sub-cellular localization, during the course of my PhD I had optimized the subcellular fractionation protocol and was successful in separating the nuclear cytoplasmic and membrane fragments, but due to time constraints was unable to complete the analysis. These clones can also be used for confocal analysis by performing β-catenin antibody staining to visualize the localization. These analysis can then be extended for other residues, and now that we have all the tools ready it would be very easy to generate both homozygous or hemizygous mutants for the rest of the residues.

The results from both the saturation and multiplex clones have shown allele-specific activity levels that cannot be explained by the current signaling model based on the sequential cascade, and hence, would require further analysis of the mechanism of regulation. In addition to assays to understand the localization and transcriptome profile, it is necessary to perform proteomic analysis. The analysis of the protein-protein interaction in the β-catenin variants can be done by performing mass spectrometry analysis on the hemi/homozygous mutants which will help in identifying the differential binding partners. These experiments together will provide insights into mechanism of regulation of the differential activity response by the mutational variants, which will help in understanding their role in the tumourigenic process.

Concluding Remarks

In conclusion, using two complementary approaches, I provide evidence confirming the genotype-phenotype correlation among the various β-catenin mutational variants. These results not only emphasize the importance of understanding the allele specific variation in β-catenin activity, in contributing to the tumourigenic response, but also highlights the drawbacks in the current model of β-catenin signaling, and thus underscores the need to further examine the underlying mechanistic process involved in the observed differential phenotypic response.

Chapter 7 Materials and Methods

7.1 General buffers and solutions

The general buffers and solutions including phosphate buffered saline (PBS), 3M NaAc, 50X TAE, TE, 5M NaCl and Tris Hcl were prepared and autoclaved by the Central Support Unit (CSU) of the Roslin institute.

7.2 Molecular Biology

7.2.1 DNA isolation techniques

7.2.1.1 Genomic DNA isolation

DNA isolation was done using mainly two different methods based on the quality of DNA required for downstream applications. For genotyping purpose, DNA was isolated using QuickExtract (QE) DNA extraction solution (Epicentre). The cells were washed with PBS followed by the addition of appropriate amount of QE to cover the cell surface, typically for cells cultured in 96 well plates 50µl of QE was added and incubated for 1-2min. The cell lysates were then transferred to 96 well PCR plates and incubated at 68°C for 15 min followed by 98°C for 8 min in a thermocycler. The crude lysates were then directly used for PCR purpose. For saturation assay, where better quality of DNA was required, DNeasy blood and tissue kit (Qiagen) was used and DNA was isolated according to the manufacturer's protocol.

7.2.1.2 Plasmid DNA isolation

Plasmid DNA was isolated using either QIAprep spin miniprep kit or maxiprep kit (Qiagen) depending on the quantity of DNA required. For restriction digestion experiments and sequencing purpose where smaller quantity of DNA is sufficient, miniprep kit was used, and for transfection experiments that require larger quantity of DNA, the plasmid DNA was isolated using maxiprep kit, following the manufacturer's protocol.

7.2.2 Quantification – Nanodrop

DNA was quantified using Nanodrop 2000 UV VIS spectrophotometer (ThermoScientific).

7.2.3 DNA clean up and ethanol precipitation for transfection

Before the use of maxiprep plasmid DNA for transfection, the DNA was precipitated and made ethanol free. For this, 50µg of plasmid DNA was brought to a total volume of 200µl in TE buffer. First, 1/20 3M Sodium acetate was added and mixed well by vortexing. This was followed by the addition of three volumes of 100 percent ethanol, and the tubes were placed overnight at -20°C or at -80°C for 15min. The tubes were centrifuged at 13k rpm for 15min to pellet the DNA and then transferred to T/C hood. The supernatant was aspirated and the pellet was allowed to air dry for 15min. The DNA was then reconstituted in 50µl PBS to obtain a final concentration of 1µg/µl and mixed well to completely resuspend the DNA in the solution. The ethanol free DNA was stored at 4°C and used for transfection purpose.

7.2.4 PCR components

7.2.4.1 Primers

All primers were designed using either Primer3Plus (www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) or Geneious software (https://www.geneious.com/), and ordered from Sigma, in desalted and dried form. The primers were reconstituted to 100µM concentration with dH2O. Aliquots of 12.5µM concentration of working solution were made and used for PCR at a final concentration of 0.25 µM.

7.2.4.2 PCR Master-mix components

Different PCRs were optimized using different polymerases, and were used along with the available compatible buffers/master mixes. High fidelity DNA polymerases including Q5 high fidelity DNA polymerases and Q5 hot start high fidelity DNA polymerases (NEB), Phusion high fidelity DNA polymerases (Thermo scientific) were available with their own ready to use compatible master mixes with all components added. The master mix for standard taq polymerase/platinum taq (Thermo scientific) used for general genotyping purpose was supplemented with dNTPs and Mgcl2are given below. The components for the reaction mix are given in table 7-1.

7.2.4.2.1 dNTPs

100mM dNTPs (Invitrogen) each of dATP, dTTP, dCTP, dGTP were mixed in equal ratios and diluted with appropriate amount of dH2O to obtain a working concentration of 2mM, and was used for PCR at a final concentration of 200µM.

7.2.4.2.2 MgCl2

50mM Mgcl2 was used for PCR at a final concentration of 2.5mM.

Reaction components	**Volume**
10X master mix	2.5µl
dNTP (2mM)	2.5µl
MgCl2 (50mM)	1.25µl
Forward primer (12.5µM)	0.5µl
Reverse primer (12.5µM)	0.5µl
Taq polymerase- Thermo scientific (5u/µl)	0.2µl
DNA (10-100ng)	1.0µl
dH2O	Made upto 25.0µl

Table 7-1: Components for PCR using taq polymerase.

7.2.5 Agarose gel electrophoresis

Separation of DNA fragments based on their size was done using agarose gel electrophoresis. The concentration of agarose was determined based on the size of the DNA being separated. Appropriate amount of agarose was weighed and dissolved in 1X TAE buffer by heating the mixture in a microwave oven. Next, gel red dye (Cambridge Biosciences) (an alternative to EtBr that acts as intercalating agents and allows visualization of DNA under UV) was added to the agarose mixture, mixed well and poured onto a gel casting tray containing a comb and allowed to set.

The DNA was mixed with loading dye and loaded into the wells along with a DNA ladder. Hyper ladder 1kb or 100bp (Bioline) were used to determine the size of the DNA. The

electrophoresis tank was filled with 1X TAE and the gel was run at 100-120V for 1hr. The gel was visualized under a UV transilluminator.

7.2.5.1 Elution of DNA from agarose gel

Elution of DNA from agarose gel was performed using GeneJet gel extraction kit (Thermo Scientific) according to the manufacturer's protocol.

7.2.6 CRISPR design and assembly

The online CRISPR design tool provided by Feng Zhang's lab (http://crispr.mit.edu/) was mostly used to identify suitable target site for recognition of sgRNA and to design the 20nt guide oligo. In addition, the recently available Sanger Institute design tool – WGE (http://www.sanger.ac.uk/htgt/wge/) and Benchling (https://benchling.com/) were also used to design guide oligos.

7.2.6.1 Ordering of guide oligo

Processing by U6 promoter is significantly enhanced by the presence of G nucleotide at the beginning of the guide (Ran et al. 2013). Hence an additional G nucleotide was added at the start of the oligo (if G is not the first base of the oligo). In addition bases complementary to the 5' and 3' overhangs formed following restriction digestion of the backbone pX458 nuclease vector, were added on either ends of the oligo. Using these guidelines the complementary top and bottom strand of guide oligos were ordered separately from Sigma. The sequence of the top and bottom oligos for each of the designed guides used in this project is given in table 7-2

Target region	Guide	Top oligo	Bottom oligo
Exon 3	g3	CACCGCTGGCAGCAGCAGTCTTACT	AAACAGTAAGACTGCTGCTGCCAGC
	g5	CACCGCAGCAGTCTTACTTGGATTC	AAACGAATCCAAGTAAGACTGCTGC
	g19	CACCGCTGTGGTGGTGGCACCAGAA	AAACTTCTGGTGCCACCACCACAGC
	g35	CACCgACCACAGCTCCTTCCCTGAG	AAACCTCAGGGAAGGAGCTGTGGTC
	g9B	CACCGAGCTCCTTCCCTGAGTGGCA	AAACTGCCACTCAGGGAAGGAGCTC
	g8B	CACCGCTCCTTCCCTGAGTGGCAA	AAACTTGCCACTCAGGGAAGGAGC
	g6	CACCGAGTGGCAAGGGCAACCCTG	AAACCAGGGTTGCCCTTGCCACTC
Start codon	Scg3	CACCGCGTGGACAATGGCTACTCA	AAACTGAGTAGCCATTGTCCACGC
Intron 1	bcat KO g2	CACCGTCTGCCTTTTGACGGACATT	AAACAATGTCCGTCAAAAGGCAGAC
	bcat KO gSanger	CACCGCACCCTCCAGGGCTGCTGTG	AAACCACAGCAGCCCTGGAGGGTGC

Intron 6	bcat KO gB36	CACCGAAAGCCTCACAGGATCCACC	AAACGGTGGATCCTGTGAGGCTTTC
	bcat KO g1	CACCGTGTTAGAGTTGGGCTAAGGC	AAACGCCTTAGCCCAACTCTAACAC
bGH polyA	PKO 5' g2	CACCGTGGGGATGCGGTGGGCTCTA	AAACTAGAGCCCACCGCATCCCCAC
	PKO 5' g4	CACCGCCACCGCATCCCCAGCATGC	AAACGCATGCTGGGGATGCGGTGGC
PGK	PKO 3' g1	CACCGCCTCCCCTACCCGGTAGTG	AAACCACTACCGGGTAGGGGAGGC
	PKO 3' g2	CACCGCCTCACTACCGGGTAGGGG	AAACCCCCTACCCGGTAGTGAGGC

Table 7-2: Sequence of the designed guides.

*CACC/AAAC – complementary sequences (sticky ends) required for ligation with BbsI digested pX458 vector.

G/complemtary C– additional G nucleotide required for U6 promoter when the guide sequence does not begin with a G and complementary C added to the bottom oligo for base pairing with G nucleotide in the top oligo.

7.2.6.2 Backbone vector for cloning sgRNA

The mammalian codon optimized Cas9 nuclease from *S. pyogenes* along with T2A EGFP has been cloned in a mammalian vector system by Feng Zhang's lab (Ran *et al.*, 2013). In addition, it consist of insertion site for sgRNA of interest and allows single step cloning of the guide based on Golden gate cloning strategy detailed in section 7.2.9.2.3, thus allowing expression of both Cas9 and sgRNA required for targeted induction of DSB. The EGFP further provides a selection strategy for fluorescence based separation of

successfully transfected cells. pSpCas9(BB)-2A-GFP (pX458) was a gift from Feng Zhang (Addgene plasmid # 48138).

In addition to GFP pX458, a mCherry version of pX458 nuclease vector was cloned as follows:

A PCR was performed (using platinum taq) to amplify the mCherry sequence from an existing mCherry vector with primers mCherry F and R (Table 7-3) having regions overlaping with the pSpCas9 (BB)-2A-(pX458) backbone vector. The PCR was performed using the reaction mix given in table 7-4 and thermocycler parameters given in table 7-5. Following amplification, DpnI digestion was performed to get rid of the template plasmid using the reaction mix given in table 7-6. Next, backbone vector was generated by restriction digestion. The GFP pX458 CRISPR vector (used as backbone vector) has an EcoRI site on either side of the GFP tag and could be cleaved out with restriction digestion with EcoRI, and then mCherry sequence was inserted into the pX458 CRISPR vector in a single step Gibson assembly by incubating the reaction mix consisting of mCherry amplicon, the backbone vector and home-made GA master mix for 1 hour at 50°C, as described in section 7.2.9.2.1. The assembled vector was then transformed into competent Stbl3 cells and cultured on LB plates containing ampicillin selection. A colony PCR was performed to shortlist clones with the mCherry insert and the positive clones were sequenced by Sanger sequencing.

Primer	**Sequence**
mCherry F	AGGCAAAAAAGAAAAAGGAAGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAAC ATGCGGTGACGTCGAGGAGAATCCTGGCCCAGTGAGCAAGGGCGAGGAG
mCherry R	CGAGCTCTAGTTAGAATTTTACTTGTACAGCTCGTCCATGC

Table 7-3: Primers used to amplify mCherry insert.

Reaction components	Volume	Incubation parameter
DNA (GFP pX458) (1 µg/ µl)	2.0µl	37°C for 2hrs
Buffer H	2.0µl	
EcorI enzyme (Promega)	1.0µl	
dH20	Made up to 20.0 µl	

Table 7-4: Reaction mix and incubation parameter for EcoRI Restriction digestion of GFP pX458 vector.

Step	Temperature	Time
1	94°C	2min
2	94°C	30s
3	59°C	30s
4	72°C	60s
5	72°C	10min
6	16°C	∞

Table 7-5: PCR parameters for amplification of mCherry insert for cloning and mCherry colony PCR.

Reaction components	Volume	Incubation parameter
DNA (mCherry PCR product)	3.0µl	37°C for 2hrs
Acetylated 1/10 BSA	2.0µl	
Buffer B	2.0µl	
DpnI enzyme (Promega)	1.0µl	
dH2O	Made up to 20.0 µl	

Table 7-6: Reaction mix and incubation parameter for DpnI digestion of GFP pX458 vector.

7.2.6.3 Annealing and Phosphorylation of guide oligos

Initially, the top and bottom strands of the guide oligos were annealed and phosphorylated in a single reaction. The components of the reaction mix and thermocycler parameters are shown in table 7-7 and table 7-8.

Reaction components	Volume
sgRNA TOP (100µM)	1.0µl
sgRNA Botom (100µM)	1.0µl
T4 ligase buffer with 10mM ATP – NEB (10X)	1.0µl
T4 Polynucleiotide Kinase - NEB	1.0µl
dH2O	Made up to 10.0µl

Table 7-7: Reaction mix for sgRNA annealing and phosphorylation.

Step	Temperature	Time
1	37°C	5min
2	94°C	5min
3	25°C	5min
4	Ramp down to 25C at 0.1°C/s	

Table 7-8: Thermocycler parameters for sgRNA annealing and phosphorylation

7.2.6.4 Insertion of guide oligo into pX458

Next, the annealed and phosphorylated oligo was cloned into the pX458 nuclease vector. The cloning of the inserts into the pX458 nuclease vector is based on Golden Gate cloning strategy making use of the typeIIS restriction enzyme BbsI that cuts outside the recognition site, hence allowing the restriction digestion of the pX458 nuclease vector using BbsI and ligation of ds oligo by base pairing with the complementary overhangs to be carried out in a single reaction, as described in section 7.2.9.2.3. The reaction mix and thermocycler parameters are shown in table 7-9 and table 7-10.

Reaction components	Volume
pX458 nuclease (100ng)	1.0µl
1:20 diluted oligo	2.0µl
T4 ligase buffer with 10mM ATP (NEB)	2.0µl
ATP (1mM)	1.0µl
Fast digest BbsI (Thermo scientific)	1.0µl
Quick ligase (NEB)	0.5µl
dH20	Made up to 20.0 µl

Table 7-9: Reaction mix for insertion of guide oligos into pX458.

Step	Temperature/Cycle condition	Time
1	37°C	5min
2	21°C	5min
3	Cycle to step 1, 5 times	

Table 7-10: Thermocycler parameters for insertion of guide oligos into pX458.

7.2.6.5 PlasmidSafe nuclease treatment

To remove any residual linearized DNA that may result in false positive, the ligation reaction was treated with plasmidSafe nuclease using the reaction mix and incubation temperatures shown in table 7-11 and table 7-12.

Reaction Components	Volume
Ligation reaction	11.0µl
ATP (1mM)	1.5µl
Plasmid safe buffer – 10X	1.5µl
Plasmid safe ATP dependent DNase - Epicentre	1.0µl
dH20	1.0µl

Table 7-11: Reaction mix for PlasmidSafe nuclease treatment.

Step	Temperature	Time
1	37°C	30min
2	70°C	30min

Table 7-12: Reaction mix for PlasmidSafe nuclease treatment

Following plasmid safe treatment 2µl of the vector was transformed into Stbl3 (20µl) competent cells using the manufacturers instruction and plated on ampicillin LB plates and incubated at 37°C overnight.

All the designed guides given in table 7-2 were cloned into either GFP pX458 or mCherry pX458. Next day, following transfection and plating, 3 clones were picked for each guide and inoculated for miniprep plasmid isolation. The isolated plasmids were sequenced using U6 F primer given in table 7-13, and one vector with correct guide sequence was chosen for further targeting. For the purpose of transfection, plasmid DNA was isolated by maxiprep isolation and cleaned by ethanol precipitation and reconstituted to 1µg/µl in sterile PBS as described in section 7.2.3.

Sequencing target	Forward primer
U6	GAGGGCCTATTTCCCATGATTCC

Table 7-13: Primer for sequencing the CRISPR guides.

7.2.7 Sanger sequencing

Sanger sequencing was done at the sequencing unit at MRC, HGU and the ABI files were analysed using geneious software.

7.2.8 T7 Endonuclease I assay

The efficiency of CRISPRs to induce strand breaks and the resulting indels (for guides targeting the exon 3 region) was analysed by performing T7 assay. The CRISPR plasmids with the specific guides inserted were transfected into mouse embryonic cells E14 with lipofectamine 2000 using transfection protocol given in section 7.4.1.5, and the transfected cells were FACS sorted and collected 24 hours post transfection. The sorted cells were put back in culture for 2 days, genomic DNA was isolated, and PCR amplification using Phusion high fidelity DNA polymerase (Thermo Scientific) was done with primers flanking the exon 3 region given in table 7-14 (β-catenin exon 3 forward and reverse primers). The PCR reaction mix and thermocycler parameters are shown in table 7-15 and table 7-16. Following amplification, 25µl of the PCR product was purified using QIAquick PCR purification kit (Qiagen) according to the manufacturer's protocol, and eluted in 50µl of elution buffer.

The purified PCR product was then denatured and reannealed to form heteroduplexes. For this, 2µl of NEB buffer 2 (10X) was mixed with 18 µl of the purified PCR product and the reaction mix was incubated using the thermocycler parameters given in table 7-17. The denatured and reannealed product was finally digested with T7 endonuclease I (T7EI) enzyme and run on an agarose gel. The reaction mix and incubation parameters for T7E1 digestion are given in table 7-18 .The indels formed by NHEJ following induction of strand break are visible as extra bands on the agarose gel. The bands were then quantified based on the intensity using the software ImageJ. The percentage of indels was calculated using the formula $100 \times (1 - (1 - (b + c)/(a + b + c))^{1/2})$, where a is the integrated intensity of the undigested PCR product and b and c are the integrated intensities of the products cleaved by T7E1 (Ran *et al.*, 2013).

Target	Forward primer	Reverse primer
β-catenin exon 3	TCTCCTTGGCTGCCTTTCTA	GTCACACAGCCCTGTC

Table 7-14: Primers used for amplifying β-catenin exon 3 region.

Reaction components	Volume
HF Buffer (5X)	10.0µl
10mM dNTP	5.0µl
Exon 3 F primer	1.3µl
Exon 3 R primer	1.3µl
Phusion HF DNA polymerase – Thermo scientific	0.5µl
DNA	4.0µl
dH2O	Made upto 50µl

Table 7-15: Thermocycler parameters for amplifying exon 3 region for T7E1 assay.

Step	Temperature	Time
1	98°C	2min
2	98°C	10s
3	58°C	15s
4	72°C	30s
Cycle to step 2 for 30 more times		
5	72°C	5min
6	16°C	∞

Table 7-16: PCR reaction mix for amplifying exon 3 region for T7E1 assay.

Step	Temperature	Time
1	95°C	7 min
2	85°C Ramp to 85°C at 2.0°C / second	30s
3	85°C Decrease by 5.0°C every cycle	30s

4	80°C Ramp to 80°C at 0.3°C/second Decrease by 5.0°C every cycle	30s
4	Cycle to step 3 for 11 more times	
5	4°C	∞

Table 7-17: Thermocycler parameters for denaturation and reannealing for T7E1 assay.

Reaction components	**Volume**	**Incubation parameter**
Buffer2 – NEB (10x)	0.5µl	37°C for 15 min
T7 E1 - NEB	0.5µl	
Denatured and reannealed product	20.0µl	
dH2O	4.0µl	

Table 7-18: Reaction mix and incubator parameters for T7E1 Restriction Digestion.

7.2.9 HDR templates

7.2.9.1 Design of short single strand oligodinucleotide

Short (70-130bp) single stranded oligonucleotide with the desired mutation were ordered from IDT as ultramers and used as templates for CRISPR mediated HDR. Sequence of 4mut PAM oligo, S33Y oligo, ΔS45 oligo, S45 and T41 multiplex oligos used for targeting is given in table 7-19.

ssODN	**Sequence**
4mut PAM oligo	GCTGCTGTCAGCCACTGGCAGCAGCAGTCTTACTTAGATTCTGGAATACATT CTGGTGCCACCACCACAGCTCCTTCTCTGAGTGGCAAAGGCAACCCTGAGG AAGAAGATGTTGACACCTCCCA
S33Y oligo	TGTCAGCCACTGGCAGCAGCAGTCTTACTTGGATTATGGAATTCATTCT-GGTGCCACCACCACAGCTCCTTCCCTGA
ΔS45 oligo	TCTGGAATCCATTCTGGTGCCACCACCACAGCTCCTCTGAGTGGTAAAGGCA ACCCTGAGGAAGAAGATGTTGACACCTCCCA
S45 ssODN used for multiplex targeting	

S45Y	TCTGGAATCCATTCTGGTGCCACCACCACAGCTCCTTACCTGAGTGGTAAAG GCAACCCTGAGGAAGAAGATGTTGACACCTCCCAA
S45P	TCTGGAATCCATTCTGGTGCCACCACCACAGCTCCTCCCCTGAGTGGTAAAG GCAACCCTGAGGAAGAAGATGTTGACACCTCCCAA
S45F	TCTGGAATCCATTCTGGTGCCACCACCACAGCTCCTTTCCTGAGTGGTAAAG GCAACCCTGAGGAAGAAGATGTTGACACCTCCCAA
S45C	TCTGGAATCCATTCTGGTGCCACCACCACAGCTCCTTGCCTGAGTGGTAAAG GCAACCCTGAGGAAGAAGATGTTGACACCTCCCAA
T41 ssODN used for multiplex targeting	
T41N	AGCCACTGGCAGCAGCAGTCTTACTTGGATTCTGGAATTCATTCTGGTGCCA CCAACACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGAGG
T41I	AGCCACTGGCAGCAGCAGTCTTACTTGGATTCTGGAATTCATTCTGGTGCCA CCATCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGAGG
T41A	AGCCACTGGCAGCAGCAGTCTTACTTGGATTCTGGAATTCATTCTGGTGCCA CCGCCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGAGG
T41S	AGCCACTGGCAGCAGCAGTCTTACTTGGATTCTGGAATTCATTCTGGTGCCA CCTCCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGAGG
T41P	AGCCACTGGCAGCAGCAGTCTTACTTGGATTCTGGAATTCATTCTGGTGCCA CCCCCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGAGG

Table 7-19: ssODNs used as repair templates.

7.2.9.2 Design and cloning of targeting vectors

Targeting vectors to be used as HDR templates were generated mainly using three efficient cloning approaches: Gibson assembly, TOPO TA cloning and Golden Gate cloning.

7.2.9.2.1 Gibson assembly

Gibson assembly is a simple and efficient method of cloning multiple fragments in a single step. It makes use of the 5' exonuclease activity and requires overlap sequences between the inserts and the backbone vector. The digestion of the 5' end by exonuclease is followed by the addition of nucleotides by the polymerase and nick joining by ligases at the region of annealed overlap (Gibson *et al.*, 2009). The online Tool NEBuilder (https://nebuilder.neb.com/) was used for Gibson assembly design. By providing the

backbone and insert sequences, the tool generates primers with overlap sequences to amplify insert fragments, and backbone can be generated either by PCR or restriction digestion based approaches. The tool also allows addition of spacers at the 5' end of the primers that allows various modifications to be made. On generating the inserts and backbone, the fragments can be easily assembled in a single step reaction. The reaction mix consisting of homemade Gibson assembly master mix and appropriate proportions of insert and backbone vector fragments were incubated for 1 hour at 50°C as shown in table 7-20.

Reaction components	**Volume**	**Incubation parameters**
Insert	Vector : Insert was used at ratio of either 1:2 or 1:3	50°C for 1hr
Vector Backbone		
Gibson assembly master mix	10µl	
dH2O	Made upto 20µl	

Table 7-20: Reaction mix for Gibson assembly reaction

7.2.9.2.1.1 Home-made Gibson assembly master mix

Gibson assembly master mix was made according to the modified Gibson assembly protocol provided by Brand Lab and 10 µl aliquots Gibson assembly master mix were stored at -20°C. The reaction mix is given in table 7-21.

Reaction components	**Volume**	**Final concentration**
1M Tris Hcl pH 7.5	40.0µl	100mM
1M MgCl2	4.0µl	10mM
Q5 high fidelity DNA polymerse (NEB)	5.0µl	0.5U/reaction
100Mm dNTPs- dATPs, dTTPs, dGTPs, dCTPs (Promega)	1.6µl (0.4 µl each from 100mM)	0.2mM each
5' T5 exonuclease (Epicentre)	0.32µl	0.16U/reaction
dH2O	149.08µl	

Table 7-21: Reaction components for home-made Gibson assembly master mix.

7.2.9.2.2 TOPO cloning

TOPO cloning is another efficient approach for cloning of PCR generated products. TOPO cloning is based on the property of terminal transferase activity of the taq polymerase which incorporates an additional A at the end of the non-template strand of every PCR product. The amplified product anneals to the T overhangs of the TOPO vector followed by the formation of phosphodiester bond by the action of topoisomerase I, which is covalently attached to the TOPO vector.

To avoid mismatches that are generated while amplifying large fragments using taq polymerase, PCR was performed using Q5 high fidelity DNA polymerase. The PCR product was purified to remove the residual polymerase and then incubated with taq polymerase. The removal of Q5 is crucial as its proofreading 3'-5' activity prevents addition of A overhangs (A tailing) by taq polymerase. Following addition of A overhangs using the reaction components given in table 7-22, the PCR product was then cloned into TOPO 4 vector. TOPO cloning was performed using TOPO TA cloning kit (Invitrogen) according to the manufacturer's protocol.

Reaction components	**Volume**	**Incubation parameter**
Cleaned PCR product	5.0µl	72°C for 20min
10X Buffer	2.5µl	
1mM ATP	5.0µl	
MgCl2	1.25µl	
Taq polymerase	0.2µl	
dH20	Made upto 25.0µ	

Table 7-22: Reaction mix and incubation parameters for A tailing.

7.2.9.2.3 Golden gate cloning

The Golden Gate cloning, described as precision cloning method by Engler et al, is yet another simple and efficient cloning strategy that is based on the properties of typeIIS restriction enzymes (Engler, Kandzia and Marillonnet, 2008). This method harnesses the property of typeIIS restriction enzymes to cut outside their recognition site, and in

combination with T4 DNA ligase, provides a seamless cloning approach with the convenience of performing both restriction digestion and ligation reaction in a single step.

7.2.9.2.4 Cloning of targeting vector with 1Kb homology arm

For cloning of the 1Kb homology arm vector, a previously targeted E14 clone with homozygous S45 deletion in exon 3 and V5 tag inserted immediately after the start codon in the exon 2 region in the endogenous *Ctnnb1* gene was used. The genomic DNA isolated from these S45 mutant cells were used as template to amplify the region covering 1kb homology arm on the 5' side and 3' end of the region of interest, using primers given in table 7-23. A Touch down PCR was performed with Q5 high fidelity master mix using the reaction components and thermocycler parameters given in table 7-24 and 7-25. The PCR product was purified and TOPO cloned. Following TOPO cloning, the vector was transformed into competent cells and plasmid DNA was sequence verified.

Target	**Forward primer**	**Reverse primer**
β-catenin region constituting 5' and 3' 1Kb Homology arms	TGGGCTTTAGAGGGAACAGT	ACCATTTTCTGCAGTCCACC

Table 7-23: Primer sequence of 1Kb homology arm vector.

Step	**Temparature**	**Time**
1	98°C	2min
2	98°C	10s
3	72°C	15s
4	Decrease by 1°C every cycle	
5	72°C	1.30min
Cycle to step2 9 more times		
6	98°C	10s
7	66.5°C	15s
8	72°C	1.30min

Cycle to step 6 for 20 more times		
9	72°C	5min
10	16°C	∞

Table 7-24: Thermocycler parameters for amplifying β-catenin homology arm insert.

7.2.9.2.5 Cloning of 5.5Kb WT β-catenin TOPO vector

Initially, a 5.5kb region of β-catenin was amplified using WT genomic DNA from E14 cells as template with primers given in table 7-25. A touch down PCR was performed using Q5 2X high fidelity master mix using the thermocycler parameters given in table 7-26. Following PCR clean up, A- tailing was done using taq polymerase and then the PCR product was cloned into TOPO 4 vector using TOPO cloning strategy. The vector was then transformed into competent cells and plasmid DNA was sequence verified.

Target	**Forward Primer**	**Reverse Primer**
5.5Kb region of β-catenin	GGTTGATACTACCTTGAGTACTC	GATTCACAGGGCTGCTAGTG

Table 7-25: Primers for amplifying 5.5kb region of β-catenin.

Step	**Temperature**	**Time**
1	98°C	2min
2	98°C	10s
3	72°C	15s
4	Decrease by 1°C every cycle	
5	72°C	3min
Cycle to step2 9 more times		
6	98°C	10s
7	67°C	15s
8	72°C	3min
Cycle to step 6 for 20 more times		

9	72°C	5min
10	16°C	∞

Table 7-26: PCR Reaction mix for amplifying 5.5kb region of β-catenin.

7.2.9.2.6 Cloning of PuDeltatk TV

A Gibson assembly was designed for cloning of puDeltatk vector. Primers were designed to amplify the puDeltatk selection cassette and having overlap with β-catenin region in 5.5Kb WT β-catenin TOPO vector. The selection cassette was amplified using a puDeltatk vector (kindly provided by Ailbhe Brazel) as template. The β-catenin region that would constitute the 3' and 5' arms were amplified along with the vector backbone from the 5.5kb WT β-catenin TOPO vector. The primers for amplifying the insert and the vector backbone are given in table 7-27. A touchdown PCR was performed using Q5 2X high fidelity master mix to amplify both the fragments. The thermocycler parameters are given in table 7-28 and 7-29. The two amplicons were gel eluted and the fragments were cloned using Gibson assembly as described in section 7.2.9.2.1. The vector was then transformed into competent cells and plasmid DNA was isolated using miniprep kit. Next, double digestion of the plasmid DNA was performed using XhoI and XbaI enzymes (Roche) using the reaction mix given in table 7-30, and plasmids with fragments of correct size were sequence verified.

Target	**Forward primer**	**Reverse primer**
puDeltatk selection cassette	CTGTTTTCATTCTGCCTTTTGACCATAGAGCCCACCGCATCC	GCCAACAAAGAAAGCCTCACTACCGGGTAGGGGAGGCG
5' and 3' β-catenin arms and vector backbone	GTGAGGCTTTCTTTGTTGGC	GTCAAAAGGCAGAATGAAAACAG

Table 7-27: Primers for cloning puDeltatk vector.

Step	Temperature	Time
1	98°C	2min
2	98°C	10s
3	72°C	15s
4	Decrease by 1°C every cycle	
5	72°C	1.30min
Cycle to step2 9 more times		
6	98°C	10s
7	69°C	15s
8	72°C	1.30min
Cycle to step 6 for 20 more times		
9	72°C	5min
10	16°C	∞

Table 7-28: Thermocycler parameters for amplifying puDeltak selection cassette.

Step	Temperature	Time
1	98°C	2min
2	98°C	10s
3	65°C	15s
4	Decrease by 1°C every cycle	
5	72°C	3min
Cycle to step2 9 more times		
6	98°C	10s

7	59°C	15s
8	72°C	3min
Cycle to step 6 for 20 more times		
9	72°C	5min
10	16°C	∞

Table 7-29: Thermocycler parameters for amplifying β-catenin backbone vector for PuDeltatk cloning.

Reaction components	Volume µl	Incubation parameter
XhoI	1.0µl	37°C for 2 hours
XbaI	1.0µl	
Sure cut buffer H	2.0µl	
Miniprep plasmid DNA	1.0µl	
Nuclease free dH20	15.0µl	

Table 7-30: Reaction mix and incubation parameters for identification of correctly cloned puDeltatk vector.

7.2.9.2.7 Cloning of β-catenin Golden gate vector

For the purpose of cloning, all the TVs for multiplex targeting and saturation editing using Golden Gate cloning, initially a β-catenin designation vector was cloned by Gibson assembly by incorporating two typeIIS restriction enzyme BbsI recognition sites flanking the region of interest. The β-catenin region constituting the 5' and 3' arm and the backbone vector was amplified using the 5.5Kb WT β-catenin TOPO vector as template. However, this template already had a BbsI site in the 5' end of β-catenin region, and hence primers were designed such that the F primer included a synonymous mutation in BbsI recognition site (as a spacer) and the R primer deleted the region of our interest and inserted two BbsI sites in the opposite orientation with two extra bases between the two recognition sites. The primer sequences are given in table 7-31.

Both the insert and backbone vector PCRs were performed using Q5 2X master mix (NEB) using the reaction mix and thermocycler parameters given in table 7-32. The two

amplicons were gel eluted and the fragments were cloned using Gibson assembly as described in section 7.2.9.2.1. Next, the vector was transformed into competent cells and plasmid DNA was isolated using miniprep kit. Then, double digestion of the plasmid DNA was performed using Fast digest BbsI and NotI enzymes (Thermo Scientific) using the reaction mix given in table 7-33 and plasmids with fragments of correct size were sequence verified.

Target	Forward primer	Reverse primer
Vector backbone	GGGTCTTCCAGAAGACCTTGGGAGCAAGGCTTTTCC	CTCATTTTGGTTTTACTGTATAATATTCAAGAAAAC
β-catenin 5' and 3' arm insert	TACAGTAAAACCAAAATGAGGACATAATTTAGACTAAAGTTCACCAG	AAAAGCCTTGCTCCCAAGGTCTTCTGGAAGACCCCAGCTTTTCTGTCCGGCT

Table 7-31: Primers used for generation of β-catenin golden gate vector.

Step	Temperature	Time
1	98°C	30s
2	98°C	10s
3	64°C	30s
4	72°C	3min
5	72°C	5min
6	16°C	∞

Table 7-32: Thermocycler parameters for amplifying β-catenin backbone and insert for β-catenin Golden gate vector cloning.

Reaction components	Volume	Incubation parameter
Fast digest BbsI	1.0µl	37°C for 40 min
Fast digest NotI	1.0µl	
Fast digest buffer	2.0µl	
Miniprep plasmid DNA	1.0µl	
Nuclease free dH20	15.0µl	

Table 7-33: Reaction mix and incubation parameters for identification of correctly cloned β-catenin golden gate vector.

7.2.9.2.8 Golden gate cloning of vectors for multiplex targeting

For the purpose of cloning the TVs for multiplex targeting, ds oligos were ordered from Geneart Strings DNA Libraries (Thermo Fisher Scientific). The ds oligos were ordered in 6 sets, one set (consisting of all chosen amino acid variants) for each of the top six chosen residues from COSMIC database. Each library consisting of the required degenerate nucleotides (depending on the required amino acid variants) was synthesized by random distribution of variants based on IUPAC nomenclature, and hence contained additional variants. The sequence of the ds oligos is given in table 7-34.

The libraries were delivered in dried form, and were reconstituted to a concentration of 0.1µM in dH2O according to the manufacturer's protocol. Next, ds library was cloned into the β-catenin backbone vector using Gibson assembly protocol using 1µl of the reconstituted product in the ligation step of Ran et al protocol described in section 7.2.6, followed by plasmidSafe treatment and transformation in Stbl3. The six sets of transformants were plated on LB plates with kanamycin selection and incubated overnight at 37°C.

Multiplex dsDNA Library	Sequence
D32 dsDNA Library	GCCATGGAGCCGGACAGAAAAGAAGACCCGCTGCTGTCAGCCA CTGGCAGCAGCAGTCTTACTTGNDTTCTGGAATCCATTCTGGTGC CACCACCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGAGG AAGAAGATGTTGACACCTCCCAAGTCCTTTATGAATGGGCCGTCT TCAGCAAGGCTTTTCCCAGTCCT

S33 dsDNA Library	GCCATGGAGCCGGACAGAAAAGAAGACCCGCTGCTGTCAGCCA CTGGCAGCAGCAGTCTTACTTGGATYNTGGAATCCATTCTGGTGC CACCACCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGAGG AAGAAGATGTTGACACCTCCCAAGTCCTTTATGAATGGGCCGTCT TCAGCAAGGCTTTTCCCAGTCCT
G34 ds DNA Library	GCCATGGAGCCGGACAGAAAAGAAGACCCGCTGCTGTCAGCCA CTGGCAGCAGCAGTCTTACTTGGATTCTRDAATCCATTCTGGTGC CACCACCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGAGG AAGAAGATGTTGACACCTCCCAAGTCCTTTATGAATGGGCCGTCT TCAGCAAGGCTTTTCCCAGTCCT
S37 dsDNA Library	GCCATGGAGCCGGACAGAAAAGAAGACCCGCTGCTGTCAGCCA CTGGCAGCAGCAGTCTTACTTGGATTCTGGAATCCATKNTGGTGC CACCACCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGAGG AAGAAGATGTTGACACCTCCCAAGTCCTTTATGAATGGGCCGTCT TCAGCAAGGCTTTTCCCAGTCCT
T41 dsDNA Library	GCCATGGAGCCGGACAGAAAAGAAGACCCGCTGCTGTCAGCCA CTGGCAGCAGCAGTCTTACTTGGATTCTGGAATCCATTCTGGTGC CACCNHCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGAGG AAGAAGATGTTGACACCTCCCAAGTCCTTTATGAATGGGCCGTCT TCAGCAAGGCTTTTCCCAGTCCT
S45 dsDNA Library	GCCATGGAGCCGGACAGAAAAGAAGACCCGCTGCTGTCAGCCA CTGGCAGCAGCAGTCTTACTTGGATTCTGGAATCCATTCTGGTGC CACCACCACAGCTCCTYNCCTGAGTGGCAAGGGCAACCCTGAGG AAGAAGATGTTGACACCTCCCAAGTCCTTTATGAATGGGCCGTCT TCAGCAAGGCTTTTCCCAGTCCT

Table 7-34: Sequence of the six ds libraries synthesized by Geneart Strings DNA library for cloning of multiplex TVs

GAAGACCC/CCGTCTTC BbsI restriction site, IUPAC nomenclature of nucleic acids Y-C/T, N-A/C/G/T, R-A/G, D-A/G/T, K-G/T, H-A/C/T

Next day, colonies were picked from each plate to perform colony PCR (in 96 well PCR format) and simultaneously inoculated in 96 well plates containing LB broth with kanamycin selection that was used to make glycerol stock which were then stored at -80°C. A small amount of the amplicons from colony PCR was checked for insertion by running on agarose gel, and the remaining products were sent for Sanger sequencing. As the synthesized library contained extra variants 3X96 well plates (288 clones) were sequenced and the required variants were selected. Further, maxiprep was performed separately for each of the 26 TVs and were cleaned and reconstituted in PBS as detailed in section 7.2.3.

7.3 Bacterial work

All microbiology work was carried out using aseptic techniques. Nutrient media, both liquid and solid- LB broth and LB agar were prepared and autoclaved by the Central Support Unit of Roslin Institute. For culturing of LB broth containing bacterial inoculum (usually bacteria transformed with plasmids with resistance gene), the media was supplemented with appropriate antibiotics – mostly either ampicillin (100mg/ml) or kanamycin (50mg/ml) just before inoculation and cultured overnight at 37°C in a shaker incubator. Usual volumes of either 500ml or 5ml LB broth was inoculated for maxiprep and miniprep plasmid isolation, respectively.

For culturing of bacteria on LB agar, the solidified media was heated and defrosted in a microwave and allowed to cooldown to 50°C and then supplemented with the appropriate antibiotic and poured on to petri dishes and allowed to solidify. Bacteria was cultured on petri dishes using spread plate technique to obtain a good spread of single cell colonies. The petri plates were incubated (placed inverted) overnight in an incubator at 37°C.

7.3.1 Bacterial transformation

Various chemically competent *E.coli* strains (Stbl3, DH5α, C2987H, TOP 10) were used for transformation of plasmid DNA using heat shock method. Briefly, the plasmid DNA was incubated with competent cells on ice (the incubation times vary for different competent cells) after which the tubes are transferred to 42°C water bath (for 30-45s again depending on the strain being used) allowing to create pores for uptake of the DNA, and then the tubes were placed on ice for 2 min for the pores to close. The transformed bacteria was immediately plated if containing ampicillin resistance, but kanamycin resistance requires an outgrowth period before plating (on LB agar with appropriate antibiotics).

ES cell targeting and screening

7.3.2 Cell culture

7.3.2.1 Sterility

A high standard of aseptic conditions was maintained in ES lab to reduce the risk of contamination, as ES cells are culture without antibiotics. Clean lab coats and nitrile gloves were worn at all times. Before start of work, the TC hood was sprayed thoroughly with 70 percent ethanol and wiped clean. All reagent bottles, falcons, pipettes etc were sprayed with 70 percent ethanol and wiped clean before placing in the T/C hood. The T/C hoods were kept uncluttered to increase space, and to avoid risk of infection. In addition, the T/C hoods were UV irradiated at the end of each day.

7.3.2.2 Cell lines and culture media

The mouse embryonic feeder free stem cell line E14IVtg2a (E14) and the TCF cell line were used in this project. The TCF cell line was derived from blastocyst from established TCF/Lef:H2B-GFP reporter mouse strain in the lab of Kat Hadjkonaikis (Ferrer-Vaquer *et al.*, 2010) .

Both E14 and TCF cell lines were grown on gelatinized flask in the presence LIF supplemented media. Both normal ES media (for culturing E14) and R2i media (for culturing R2i) consisted of the following components:

Homemade LIF was prepared by transfecting LIF expressing constructs into Cos7 cells and the expressed LIF is released by the cells. The LIF containing supernatant was tested by titrating different concentrations, and the optimal concentration was selected based on the morphology of E14 cells.

Fetal Bovine Serum (FBS) was batch tested for quality suitable for ES cell culture at MRC, and used at either 10 or 15 percent. However due to batch to batch variability, the Knock out serum replacement (KOSR - Thermo Scientific) was substituted for FBS and used at 15 percent at later stages of the project.

The media was supplemented with sodium pyruvate, non-essential amino acids, 2-mercaptoethanol. In addition, R2i media consisted of GSK3β inhibitor CHIR99021 and MEK inhibitor PD0325901 and both these components are known to maintain cells ground state pluripotency of ES cells. The exact composition of normal ES media and R2i media are given in table 7-35 and table 7-36.

Normal ES media components	**Volume**	**Final Concentration**
Glasgow MEM (BHK 21) Life Technologies	500ml	
FBS	50-75ml	10-15%
MEM Non-essential amino acids (Life Technologies)	5.0ml	0.1Mm
Sodium pyruvate (Life Technologies)	5.0ml	1Mm
2-mercaptoethanol (Life Technologies)	1.0ml	0.1Mm
L-Glutamine (Life Technologies)	5.0ml	2Mm
Homemade LIF (Life Technologies)	1.0-2.0ml depending on titre amount	

Table 7-35: Normal ES cell media composition.

R2i media components	**Volume**	**Final Concentration**
Knockout DMEM Life Technologies	500ml	
FBS	50-75ml	10-15%
MEM Non-essential amino acids (Life Technologies)	5.0ml	0.1Mm
2-mercaptoethanol (Life Technologies)	1.0ml	0.1Mm
L-Glutamine (Life Technologies)	5.0ml	2Mm
Homemade LIF (Life Technologies)	1.0-2.0ml depending on titre amount	
CHIR99021 (10 mM stock made up in DMSO)	0.150ml	3μM
PD0325901 (10 mM stock made up in DMSO)	0.050ml	1μM

Table 7-36: R2i media composition

7.3.2.3 Passaging of cells

Both E14 and TCF cells were generally grown in gelatinized T75 flask and were passaged on alternate days at 1/5- 1/8 ratio depending on the confluency. Prior to splitting culture flasks were gelatinized by adding 0.1 percent gelatin, enough to cover the entire flask, and placed in the incubator for a minimum of 15 minutes. For splitting a T75 flask, the media was aspirated and the cells were washed with 5ml of PBS to remove any residual serum. After aspiration of PBS, the flask was coated thoroughly with 1ml trypsin and placed in the incubator for few minutes, until the cells detach and become rounded. Next, 9ml of media was added to stop the activity of trypsin and the cells were mixed few times. The cell suspension was transferred to a 15ml falcon tube and centrifuged at 1000rpm for 5 min. Meanwhile, the gelatin was aspirated from flask and immediately replaced with 9ml of culture media. Next, the supernatant was aspirated and the cell pellet was resuspended in media by gentle pipetting to obtain a single cell suspension. The cell suspension was plated at appropriate density (1:4-1:8) in a gelatinized flask, in a total volume of around 10ml of media, and the flask was shaken well to get an equal distribution of cells throughout the flask.

7.3.2.4 Cryopreservation and thawing

For cryopreservation, cells were washed with PBS, trypsinized and centrifuged as described above. The cell pellet was resuspended with freezing media constituting 10percent DMSO in FBS, and then transferred to cryovials. The cryovials were immediately placed in Styrofoam box, to ensure steady decrease in temperature and placed in -80°C overnight. The cryovials were then transferred to liquid nitrogen for long term storage.

Prior to thawing of cells, 9ml of media was added to a 15 ml falcon and a Pasteur pipette was kept ready in the T/C hood. The cryovial immediately after removal from liquid nitrogen storage, was held in 37°C water bath for 1min until the frozen media just begins to melt. The cells along with freezing media was transferred to the falcon containing media using the Pasteur pipette. The tubes were centrifuged at 1000rpm for 5min. The cell pellet was then resuspended in media and usually plated in a single flask.

7.3.2.5 Transfection

The cells were transfected with either nucleofection or lipofection method.

Prior to transfection using nucleofection method, cells were cultured and split the day before transfection. Electroporation using nucleofection was carried out according to the manufacturer's protocol (Lonza). For nucleofection a total of 2µg of DNA (1µg each of CRISPR vector and template DNA) was used to transfect $2X10^6$ cells.

Transfection using lipofection method was carried out using suspension cells. For lipofection $8X10^5$ cells were counted and plated in 2ml media in 6 well plate, 9µL of Lipofectamine 2000 (Thermo) was diluted in 191µl of OPTIMEM media (reduced serum media) and DNA (2µg each of CRISPR vector and template DNA) was diluted in OPTIMEM media to a total volume of 200µl. Both DNA and lipofectamine were briefly vortexed and incubated for 5min at RT after which 200µl of lipofectamine was mixed with 200µl DNA vortexed briefly and incubated at RT for (a minimum of) 20min. Meanwhile $8X10^5$ cells were counted and plated in 2ml media (without penstrep) in 6 well plate and 400µl of DNA:lipofectamine suspension was added, mixed well and incubated. 4-6 hours after transfection the media was replaced and cultured overnight.

Small molecule enhancers: Two drugs were tested for their ability to enhance HDR efficiency. The drugs SCR7 and L755507 were added to the media at a final concentration of 1µM and 5 µM, respectively. In each case, the cells prior to plating for transfection were resuspended in media containing the drug.

The combination of transfection by lipofection and addition of L755507 was used for all the targeting performed post optimization.

7.3.2.5.1 Selection of transfected clones

The fluorescence markers either GFP/mCherry in the pX458 nuclease vectors allowed selection of clones successfully transfected with the CRISPR Cas9 vector that would enrich for clones edited by DSB induction. The cells transfected by either nucleofection or lipofectamine were sorted into single cells 24 hours post transfection by FACS, and collected into a 15ml falcon. The pooled cells were diluted and approximately 500 cells

were plated in each of the 10cm dishes. This method was used for all targeting experiments performed for the purpose of optimization.

The generation of heterozygous β-catenin mutant cell lines for both multiplex targeting and saturation editing was based on positive negative selection strategy. Hence, instead of FACS sorting, the transfected cells were selected based on the antibiotic resistance conferred by the introduced puDeltatk selection cassette.

7.3.2.5.2 Picking, archiving and sequencing for selection of correctly targeted mutant clones

The cells were incubated at 37°C for about 10 days or until visible colonies appear. In preparation for picking clones, 96 well flat bottom plates were gelatinized, and also 30µl of trypLE was added to U bottom 96 well plates. The media was aspirated from petri dishes, washed with 5ml of PBS, and replaced with 5ml fresh PBS. The single cell clones were picked (along with small amount of PBS) into individual wells of the U bottom 96 well plate containing trypLE using sterile passettes. The plates were then incubated to break up the cells in the colony, and 100µl of media was added and mixed by pipetting to get single cell suspension. Meanwhile, the gelatin from flat bottom 96 well plates was aspirated and replaced with 100 µl of media, and the entire cell suspension from U bottom 96 well plate was transferred to this flat bottom 96 well plate. Next day, the media was replaced, and the cells were cultured until they reach confluency. On reaching 80 percent confluency, the clones were split into two replicates. One plate was cultured further to freeze down, and the other plate was used to extract genomic DNA to test for correct editing. The crude DNA lysates were prepared using QE DNA extraction solution as described in section 7.2.1. The DNA from the clones was PCR amplified using β-catenin exon 3 F/R primers given in section 7.2.8 and the PCR product was sequenced using Sanger sequencing.

7.3.2.5.3 Freezing of 96 well plates and restarting of mutant clones

For freezing, the media was aspirated from 96 well plates, and washed with 200µl of PBS. After removing the PBS, 50µl of trypsin was added and placed in the incubator. Once the cells detach and become rounded, 70µl of FBS (fetal bovine serum) was added and pipetted thoroughly to make a single cell suspension. This was followed by the addition

of 70µl of 2X Freezing media constituting 20 percent DMSO in FBS and mixed quickly. The plate was then placed in styrofoam box and stored at -80°C.

Following sequence confirmation of mutant clones, the correctly targeted clones were started up from the frozen stocks. The 96 well plate was quickly removed from -80°C, and warm media was added to the appropriate wells containing the mutant clones, and allowed to thaw. The thawed cells were transferred to 24 well plate and cultured.

7.3.3 Generation of heterozygous β-catenin KO cell line

Both TCF and E14 cell lines were used to generate heterozygous β-catenin KO mutants. The TCF and E14 cells were transfected with 2µg of CRISPR vectors (targeting the intron 1 and intron 6 region of β-catenin) and 2µg of the puDeltatk TV using suspension cell protocol as described in section 7.4.1.5. The transfections were performed in R2i media for TCF cells and in normal ES media for E14s. Next day post transfection, the cells were trypsinized and plated in 10cm dishes at a low density in their respective media. Eight hours after plating, the media was supplemented with puromycin (positive selection antibiotic) at a final concentration of 1µg/ml. Ten days later colonies were picked into 96 well plates and processed as described in section 7.4.1.5.2.

7.3.3.1 PCR based selection of rightly targeted clones

The DNA from the targeted clones was used to perform four PCRs to identify the 5' and 3' region of puDeltatk and WT alleles. The PCRs were performed using Q5 2X master mix (NEB). The PCR parameters and primer sequences for each of the four PCRs is given in table 7-37 to 7-41. The PCR products were then sent for sequencing.

Step	Temperature	Time
1	98°C	2min
2	98°C	10s
3	72°C	15s
4	Decrease by 1°C every cycle	
5	72°C	2.30min

Cycle to step2 9 more times		
6	98°C	10s
7	69°C	15s
8	72°C	2.30min
Cycle to step 6 for 20 more times		
9	72°C	5min
10	16°C	∞

Table 7-37: PCR parameters for puDeltatk allele 5' arm PCR

Step	**Temperature**	**Time**
1	98°C	2min
2	98°C	10s
3	72°C	15s
4	Decrease by 1°C every cycle	
5	72°C	1.30min
Cycle to step2 9 more times		
6	98°C	10s
7	70°C	15s
8	72°C	1.30min
Cycle to step 6 for 20 more times		
9	72°C	5min
10	16°C	∞

Table 7-38: PCR parameters for puDeltatk allele 5' arm PCR

Step	**Temperature**	**Time**
1	98°C	2min
2	98°C	10s

3	72°C	15s
4	Decrease by 1°C every cycle	
5	72°C	2min
Cycle to step2 9 more times		
6	98°C	10s
7	66°C	15s
8	72°C	2min
Cycle to step 6 for 20 more times		
9	72°C	5min
10	16°C	∞

Table 7-39: PCR parameters for β-catenin WT allele 5' arm PCR

Step	**Temperature**	**Time**
1	98°C	2min
2	98°C	10s
3	72°C	15s
4	Decrease by 1°C every cycle	
5	72°C	2min
Cycle to step2 9 more times		
6	98°C	10s
7	66°C	15s
8	72°C	2min
Cycle to step 6 for 20 more times		
9	72°C	5min
10	16°C	∞

Table 7-40: PCR parameters for β-catenin WT allele 3' arm PCR

Target region	Forward primer	Reverse primer
puDeltatk allele 5' arm	GTGGACATCAGAGGACAACTTG	GGACCGAGTACAAGCCCAC
puDeltatk allele 3' arm	TGGATGTGGAATGTGTGCGAGGC	TGTCTCACCCTCAGCACCGTCC
β-catenin WT allele 5' arm	GTGGACATCAGAGGACAACTTG	AGAAGGGAAGAGAACAAAGGCA
β-catenin WT allele 3' arm	AGTTGTTTGTACAGAGTGTGGAGT	GCACCGTCCTCTACATGATG

Table 7-41: Primers used for amplification of puDeltatk and WT β-catenin alleles of heterozygous β-catenin KO cell lines

7.3.4 Generation of fluorescence tagged S33Y ∆S45 and WT heterozygous and hemizygous pool of cells.

The BFP and RFP tagged S33Y, ∆S45 and WT β-catenin vectors were available in the lab. The fluroscence tag in these vectors was present next to the start codon, the S33Y/ΔS45/WT in the exon 3 region, and each of these vectors had 1Kb homology arm. The scg3 and g9B guides were cloned into pX459 plasmid. Since the TCF cells had a GFP reporter and I was using BFP and RFP tagged TVs, the fluorescence based CRISPR plasmids could not be used. Hence, I cloned the two CRISPR guides in pX459 nuclease plasmid based on puromycin selection. For generating the S33Y/∆S45 heterozygous pool, the TCF cells were transfected with BFP tagged WT β-catenin vector and RFP tagged S33Y/∆S45 vector along with the two guides using suspension cell protocol described in section 7.4.1.5. In addition, another transfection was performed where in the BFP and RFP tagged WT β-catenin vectors were transfected along with the CRISPR guides. For generation of hemizygous pool, the β-catenin heterozygous KO TCF cell line was transfected with S33Y BFP/ΔS45 BFP/WT BFP along with two CRISPR guides scg3 and g9B, keeping the puDeltaK allele intact.

Second day post transfection, pools of BFP RFP red double positive cells were sorted from heterozygous condition and BFP positive cells were sorted for hemizygous condition, and the sorted cells were plated back in culture. Since the TCF cells were being cultured in the presence of 2i with one of the components being CHIRON, a GSK3β

inhibitor that would affect β-catenin activity. To avoid the effect of GSK3 β inhibitor, next day post sorting R2i media was replaced with normal ES media and the cells were further cultured for 2 days in normal ES media and GFP activity was analyzed using Flow cytometry.

7.3.5 Multiplex targeting

The heterozygous β-catenin KO E14 cell line was transfected with 2µg of CRISPR vectors targeting the selection cassette and 2µg of the multiplex TV variants of a particular residue using suspension cell protocol previously described in section 7.4.1.5. Since the heterozygous β-catenin KO E14 cell line was being cultured in R2i media, the cells were transfected in the same media. Next day post transfection, the cells were trypsinized and plated in 10cm dishes at a low density in normal ES media. Following 8hrs of plating, the media was supplemented with FIAU negative selection analogue at a final concentration of 0.2µM. Ten days later colonies were picked into 96 well plates and processed as described in section 7.4.1.5.2. The DNA from multiplex clones was used to perform exon 3 PCR which was then used for sequencing for identification of correctly targeted clones. Six sets of multiplex targeting was performed separately for each of the top six residues (D32, S33, G34, S37, T41 and S45) including all the selected corresponding amino acid substitutions.

7.4 RNA isolation

Total RNA was isolated from cells using Qiagen mini RNA kit according to the manufacturer's protocol. RNA being very sensitive to degradation by ubiquitously present ribonucleases, appropriate care was taken at every step of isolation and also while further handling and storage, including cleaning of bench top and gloves using RNA ZAP while handling RNA, using RNase free filter tips while pipetting and use of RNase free dH20 for isolation and reconstitution. Following isolation, RNA was immediately placed on ice, quantified, aliquoted and stored at -80°C until future use.

7.5 cDNA synthesis

Synthesis of cDNA from total RNA was carried out using AMV reverse transcriptase kit (Promega). A reaction mix comprising of AMV Reverse transcriptase buffer, dNTP's, oligo (dT) primer, AMV Reverse trasnscriptase enzyme (Promega) was added to each tube

(+RT) containing 2µg of RNA in a total volume of 20µl as given in table 7-42. Control – RT reactions consisting of the same reaction mix but without RTase were also set up to detect genomic DNA amplification giving false positive results. +RT reactions were set up for all E14 multiplex clones and two controls E14 and E14 β-catenin KO clone and – RT reactions were set up for S37, T41, S45 E14 multiplex clones and the two controls. The + and – RT tubes were then incubated at 50°C for 1hour. The cDNA synthesized was diluted to obtain a final concentration of 10ng/µl.

Reaction components	**Volume**
RNA	2µg
dNTPs	1.25µl
RNase inhibitor (Roche)	0.25µl
Oligo d(T) primer	1.0µl
AMV Reverse transcriptase 5X master mix (Promega)	4.0µl
AMV Reverse transcriptase (Promega)	1.0µl
dH2O	Made up to 20µl

Table 7-42: Reaction mix for cDNA synthesis

7.6 Taqman assay

Primers and probes for Taqman analysis were designed using Universal Probe library system assay design tool (Roche life sciences) (https://lifescience.roche.com/en_in/brands/universal-probe-library.html#assay-design-center). Where possible, primers were designed spanning exon-exon boundary. A reaction mix consisting of Light cycler 480 master mix, target specific primer and probe, reference primer and probe (Universal ProbeLibrary Mouse ACTB Gene Assay) along with cDNA equivalent to 20ng of RNA in a total volume of 10µl was set up in 384 well (barcoded LC480 plates) plates. The reaction components used for Taqman assay is given in table 7-43.PCR reactions were carried out in LightCycler 480 instrument (Roche Life sciences) using the inbuilt dual colour hydrolysis probe program (that allows detection of fluorescence signal from both target and reference probe). The Reference

gene β-actin expression was used for normalization of Taqman data. The primers and probes used for Taqman assay is given in table 7-44.

Reaction components	**Volume**
Light cycler 480 master mix (Roche)	5µl
Target specific primer F	0.2µl
Target specific primer R	0.2µl
Target probe	0.2µl
Reference primers	0.1µl
Reference probe	0.1µl
cDNA (10ng/µl)	2.0µl
dH2O	Made up to 10µl

Table 7-43: Reaction mix for Taqman assay

Target gene	**Forward primer**	**Reverse primer**	**Probe#**
Oct4	GTTGGAGAAGGTGGAACCAA	CTCCTTCTGCAGGGCTTTC	95
Nanog	TTCTTGCTTACAAGGGTCTGC	AGAGGAAGGGCGAGGAGA	110
Cdx1	ACGCCCTACGAATGGATG	CTTGGTTCGGGTCTTACCG	70
Fgf5	AAAACCTGGTGCACCCTAGA	CATCACATTCCCGAATTAAGC	29
Gata4	GGAAGACACCCCAATCTCG	CATGGCCCCACAATTGAC	13
Cdh1	TGTCTACCAAAGTGACGCTGA	CTCTGGGTTGGATTCAGAGG	77
T	CAGCCCACCTACTGGCTCTA	GAGCCTGGGGTGATGGTA	100

Tbx3	TTGCAAAGGGTTTTCGAGAC	TGCAGTGTGAGCTGCTTTCT	51
Klf2	AGGCCTGTGGGTTCGCTATAAA	GGCAAATTATGGCTCAAAGTAGCAG	99

Table 7-44: Primers used for Taqman assay of multiplex clones.

7.7 Protein isolation

For protein isolation, the cells were trypsinized and the pelleted down. The pellet was then washed twice with ice cold PBS. Following PBS washes, the cells were lysed by addition of appropriate amount of Nonidet P40 (NP-40) cell lysis buffer (Thermo Fisher Scientific) and the lysate was processed according to the manufacturer's protocol. To avoid degradation by protease, NP-40 was supplemented with Protease inhibitor cocktail and PMSF (Thermo Fisher Scientific). The isolated total protein was aliquoted and stored at -80°C.

7.7.1 Protein Quantitation

The protein was quantitated using BCA protein assay kit (Thermo Fisher Scientific) according to the manufacturer's protocol. The concentration of the protein was measured at 562nm using a plate reader. The assay was performed in duplicates and along with the protein of interest, Bovine Serum Albumin (BSA) protein standards ranging from 0.2-2mg/ml were included in every experiment. Following measurement of optical density at 562nm, the total protein concentration was determined by comparing it to the standard protein.

7.7.2 SDS PAGE and western blot

20µg of total protein was mixed with SDS loading dye (6X) and incubated at 90°C for 5 min and centrifuged briefly and the samples were placed on ice. The samples were resolved by running on 10 percent Tris-glycine SDS Polyacrylamide gel in 1X Tris glycine Buffer (Running buffer) for 2 hours at 100V. Following separation by electrophoresis, the protein was transferred onto PVDF membrane (GE Amersham Hybond –P) using transfer buffer in a Biorad transfer apparatus at 100V for 1 hour. The membrane was then blocked in 10 percent skimmed milk powder in TBST for 1 hour at room temperature on a shaker.

Next, the membrane was probed with 1:1000 primary antibody diluted in 5percent skimmed milk powder in TBST and incubated overnight at 4°C on a shaker. Next day, the membrane was washed twice for 20 min each with TBST and incubated in 1:10,000 HRP conjugated secondary antibody in 5 percent skimmed milk powder in TBST for 1 hour at room temperature on a shaker. The membrane was again washed twice for 20 min each with TBST, and incubated in Super signal west dura extended duration substrate (Thermo Scientific) according to the manufacturers protocol and then exposed on an X-ray film (Amersham hyperfilm ECL – GE Healthcare),developed and fixed. The composition of the buffers and antibody concentrations used are given in table 7-45 to 7-48.

Components	**Amount**	
Tris	3.1g	Made upto 1L using dH2O
Glycine	14.4g	
SDS	1g	

Table 7-45: Composition of running buffer.

Components	**Amount/Volume**	
Tris	2.9g	Made upto 1L using dH2O
Glycine	14.4g	
Methanol	200ml	

Table 7-46: Composition of transfer buffer.

Components	**Volume**
Tris (pH 7.5)	25ml
NaCl (5M)	15ml
Tween	0.5ml
dH2O	400ml

Table 7-47: Composition of TBST.

Primary antibody	Secondary antibody
Purified mouse anti β-catenin – BD transduction Lab (1:1000)	HRP conjugate anti mouse (GE Healthcare) 1:10000
PCNA – mouse monoclonal Santa Cruz Biotech (1:1000)	HRP conjugate anti mouse (GE Healthcare) 1:10000

Table 7-48: Antibody concentrations used for western blot.

7.8 FACS sorting and flow cytometry

BD FACS Aria III (BD Biosciences) was used for sorting cells based on their fluorescence spectrum. Prior to FACS sorting cells were trypsinized, resuspended in PBS substituted with 10 percent FBS and centrifuged. The supernatant was discarded and pellet was resuspended in approximately 300-500µl of PBS substituted with 10 percent FBS, mixed thoroughly by pipetting and transferred to FACS tubes (BD falcon tubes with cell strainer caps) by passing through the cell strainer cap to avoid clumps and make a single cell suspension. The tubes were placed on ice until sorting.

Filters used for FACS were: 530/30 GFP (FITC), 450/40 TagBFP, 585/42 RFP/mCherry (PE).

7.9 Compilation of *CTNNB1* mutation data from COSMIC database and statistical analysis

The *CTNNB1* mutation data was compiled from COSMIC database. Tissues types harbouring more than 10 samples with missense mutation in *CTNNB1* were selected for further analysis. A basic tabulation of the frequency distribution of mutation in various residues and also of the amino acid variants for each residue across different cancer types was performed. Statistical analysis of the compiled data was performed by Helen Brown (senior statistician at the Roslin Institute).

To test if there were significant differences in the proportion of mutation (of a specific residue/amino acid substitution) across all the tumour types: here a simulation approach was used to test whether the chi-squared test statistic obtained for the observed data is significantly greater than expected given a null hypothesis of no difference in residue proportions between the sites. One thousand frequency tables were randomly simulated

under the assumption that the residue proportions were the same at all sites. The size of chi-squared test statistics generated from the simulated tables was compared to the chi-squared statistic observed, and a p-value was obtained from the rank of the observed chi-squared statistic.

To test which sites have an overall significant difference in mutation (of a specific residue / amino acid substitution) between tumour types: here, it was necessary to test for each site whether the observed frequency of an individual mutation (of a particular residue/ amino acid substitution) is different than expected, given its expected frequency across all sites. This was addressed by comparing the observed proportion of mutation at a site to its expected proportion, assuming null hypothesis that the proportion of the observed mutation is no different to that expected.

7.10 Luciferase assay

The TOP/FOP and Renilla vectors were purchased from Addgene. The transfections were performed in triplicates using suspension cells in 96 well plates. Following plating of approximately 25,000 cells/well onto gelatinized 96 well plates, the cells in each well were transfected with 122.7ng of either TOP/FOP and 25ng of Renilla vector control. Following 4-6hrs of transfection the media was replaced by fresh media with penstrep. Next day the media was again replenished. After 36hrs post transfection the cells were processed using Dual luciferase reporter assay system (Promega) according to the manufacturers protocol, and the luciferase signal was measured by MicroLumatPlus LB96V microplate luminometer (EG and G Berthold).

For DKK1 treatment, 6 hrs post transfection the media was replaced by fresh media supplemented with 300ng/ml of Recombinant human DKK1 (Merk Millipore). Next day, the media was replenished with fresh DKK1 supplemented media and the signal was measured as described above.

7.11 Saturation assay detailed protocol

7.11.1 Cloning of Targeting vectors

7.11.1.1 Template for Twist library synthesis

Ds oligo library for golden gate cloning of 400 targeting vectors was ordered from Twist Biosciences. 200bp ds oligo consisting our region of interest which included BbsI sites at either ends, synonymous mutation in PAM region and 19th bp of g9b CRISPR guide and 3 additional synonymous mutations that act as handle for identification of specifically the HDR targeted clones was ordered as gblock from IDT (given in table 7-49) and cloned into topo4 vector as described in section 7.2.9.2.2 and sequenced.

gBlock	Sequence
gBlock Twist template	GCCATGGAGCCGGACAGAAAAGAAGACCCGCTGCTGTCAGCCACTGGCAG CAGCAGTCTTACTTGGATTCTGGAATCCATTCTGGTGCCACCACCACAGCT CCTTCCCTGAGTGGTAAAGGCAATCCCGAAGAAGAAGATGTTGACACCTCC CAAGTCCTTTATGAATGGGCCGTCTTCAGCAAGGCTTTTCCCAGTCCT

Table 7-49: gblock used for cloning of template for Twist library synthesis.

*GAAGACCC/CCGTCTTC BbsI restriction site; T/C/A synonymous handle mutation; A synonymous PAM mutation of g9B CRISPR; T synonymous mutation in 19thbp of g9B CRISPR.

7.11.1.2 Twist Library synthesis of ds oligos

The TOPO vector was used as template and ds oligos with all amino acid substitutions at each of the 20 residues between L31 and G50 were synthesized by Twist Biosciences. We received the ds oligo as 20 pools with each pool consisting of all the 20amino acid substitutions for a specific residue.

7.11.1.3 Cloning of Targeting vectors for saturation assay

The ds oligo pools for each residue were cloned into the β-catenin backbone Golden Gate vector with BbsI sites. The ds oligos were reconstituted to a concentration of 5ng/µl in dH2O and cloning and transformation was done as described in section 7.2.9.2.8. However, instead of plating on LB petri plates, the transformed bacteria was inoculated in LB broth supplemented with kanamycin and incubated overnight in a shaker at 37°C.

Next day, equal volumes of inoculum were combined from each of the 20 pools and used as starter culture for maxiprep plasmid isolation. A single maxiprep of the pooled vectors was performed followed by ethanol purification and reconstituted in PBS at 1µg/µl concentration as described in section 7.2.3.

7.11.2 Transfection for saturation assay

$200X10^6$ heterozygous β-catenin KO TCF clone cultured in R2i media were used for transfection. 312µg of each of the four CRISPR/ Cas9 and 312 µg of pooled Twist HDR vector template was used. Transfection was done using lipofection and with the addition of L755507 as described in section 7.4.1.5 and the cells were plated in 26 six well culture plates in R2i media (day1). Next day post transfection (day2) the cells from all 26 six well plates were trypsinized and plated back in T75 plates in R2i media. On day 3 the media was replaced with normal ES media along with FIAU negative selection. Further on day 4, the media was replaced with fresh normal ES media supplemented with FIAU. Finally, on day 5, the cells from all 26 flask were trypsinized and prepared for FACS sorting batch wise as described in section 7.9.

7.11.3 FACS sorting

The cells were FACS sorted based on GFP activity into six segments and collected in separate tubes. The number of cells collected from each segment is given in table 7-50. In addition a small sample of cells (pooled) was collected prior to sorting.

Segment	Replicate1 cell numbers	Replicate2 cell numbers
P2	200,000	200,000
P3	200,000	200,000
P4	200,000	200,000
P5	80,000	1,130,000
P6	445,427	942,000
P7	8,309	11,641

Table 7-50: Number of cells sorted from different segments of Replicate1 and Replicate2.

7.11.3.1 DNA isolation

The entire cell pellet from each segment was used to isolate DNA. DNA isolation was done on the same day of FACS sort using Qiagen DNeasy blood and tissue kit. Next, DNA was quantitated using nanodrop 2000.

7.11.3.2 Long range PCR

All PCRs for saturation assay was performed in TC hoods. Initially a long range PCR for P2-P6 and pooled samples from both replicates was done with F primer (B cat 5' WT/F) outside the homology arm of the targeting vector (to prevent amplification of random integration) and Handle specific reverse primer (Handle test R). The primer sequences are given in table 7-51.

Primers	**Sequence**
B cat 5' WT/F	GTGGACATCAGAGGACAACTTG
Handle test R	TGTCAACATCTTCTTCTTCGGGA

Table 7-51: Primers used for long range PCR.

Both PCRs were done using Q5 Hot Start High-Fidelity 2X Master Mix (NEB) with same reaction mix and PCR parameters with touchdown protocol. The reaction mix and PCR parameters are shown in table 7-52 and table 7-53. The entire DNA obtained was used to perform PCR, except for pool samples and P5 segment of replicate 2 due to excess DNA available.

Reaction components	**Volume**
Q5 Hot Start High-Fidelity 2X Master Mix	12.5µl
F Primer	1.0µl
R Primer	1.0µl
DNA	Different amounts of DNA (0.5-3.0µl) used for each segment depending on the concentration
PCR grade Water (sigma)	Made upto 25.0µl

Table 7-52: PCR Reaction mix for long range PCR.

Step	Temperature	Time
1	98°C	2min
2	98°C	10s
3	72°C	15s
4	Decrease by 1°C every cycle	
5	72°C	2.30min
Cycle to step2 9 more times		
6	98°C	10s
7	67°C	15s
8	72°C	2.30min
Cycle to step 6 for 21 more times		
9	72°C	5min
10	16°C	∞

Table 7-53: Thermocycler parameters for long range PCR.

7.11.3.3 DpnI digestion

Following the first round of PCR, all samples were digested with Fast digest DpnI (Thermo Scientific) to remove any residual vector. The components of the reaction mix are given in table 7-54.

Reaction Components	Volume	Incubation parameter
3Kb PCR product	25.0µl	37°C for 1hr
FastDigest DpnI	2.0 µl	
10X Fast digest buffer	3.0 µl	

Table 7-54: Reaction mix and incubation parameters for DpnI digestion.

7.11.3.4 Gel elution

Following DpnI digestion, the products from the first round of PCR was run on 0.8 percent gel at 120V for 3 hours 30min to get maximum resolution of bands. The bands were then cut using a UV trans illuminator and gel extraction was performed using GeneJet gel extraction and DNA micro cleanup kit (Thermo Scientific) according to the manufacturer's

protocol, and finally the DNA was eluted in 10µl of dH20. Several PCR bands from each segment were gel extracted individually and later pooled into a single sample.

PCR products from each segment were run individually in a single tank to avoid mixing up of PCR products from the different segments. In addition, after each run the gel tank, gel tray and combs were thoroughly washed and soaked overnight in 0.5M NaOH. NaoH is a good denaturing agent used to remove residual DNA, RNA and protein contamination. Next day, the tanks, trays and combs were rinsed with water and each time the gels were made and run in fresh 1X TAE.

7.11.3.5 Second short PCR

To allow sequencing using Illumina Miseq, a second round PCR was done using the gel eluted first round 3.2kb amplicon as the template to amplify a short fragment consisting of only the sequence of interest. The primers for the second round of PCR in addition to the target specific sequence, had the Illumina specific barcode, primer pad, linker and adaptors necessary for Illumina sequencing platform. The Illumina specifc primers were designed based on the earth microbiome project (http://press.igsb.anl.gov/earthmicrobiome/protocols-and-standards/16s/). The primers were ordered as ultramers of 4nM from IDT. Each segment (P2-P7 and pooled 3Kb handle PCR) was amplified using a Forward primer with specific barcode and handle specific reverse primer.

The short PCR was also done on the targeting vectors used for transfection using the barcoded Forward primer and handle specific reverse primer. The primer sequences are given in table 7-55. Three PCR reactions were performed for each of the 16 segments (P2-P7, pooled, plasmid from both replicate 1 and 2) using Q5 Hot Start High-Fidelity 2X Master Mix (NEB) with touch down protocol. The reaction mix and PCR parameters are given in table 7-56 and table 7-57.

Name	Sequence	Sample
bcatbc0	AATGATACGGCGACCACCGAGATCTACACGCTAGCCTTCGTCGCTATGGTAATTGTATGGCCATGGAGCCGGA	pool_1
bcatbc1	AATGATACGGCGACCACCGAGATCTACACGCTTCCATACCGGAATATGGTAATTGTATGGCCATGGAGCCGGA	P2 _1
bcatbc2	AATGATACGGCGACCACCGAGATCTACACGCTAGCCCTGCTACATATGGTAATTGTATGGCCATGGAGCCGGA	P3_ 1
bcatbc3	AATGATACGGCGACCACCGAGATCTACACGCTCCTAACGGTCCATATGGTAATTGTATGGCCATGGAGCCGGA	P4 _1
bcatbc4	AATGATACGGCGACCACCGAGATCTACACGCTCGCGCCTTAAACTATGGTAATTGTATGGCCATGGAGCCGGA	P5 _1
bcatbc5	AATGATACGGCGACCACCGAGATCTACACGCTTATGGTACCCAGTATGGTAATTGTATGGCCATGGAGCCGGA	P6 _1
bcatbc6	AATGATACGGCGACCACCGAGATCTACACGCTTACAATATCTGTTATGGTAATTGTATGGCCATGGAGCCGGA	P7 _1
bcatbc7	AATGATACGGCGACCACCGAGATCTACACGCTAATTTAGGTAGGTATGGTAATTGTATGGCCATGGAGCCGGA	Plasmid _1
bcatbc9	AATGATACGGCGACCACCGAGATCTACACGCTGCCTCTACGTCGTATGGTAATTGTATGGCCATGGAGCCGGA	pool _2
bcatbc10	AATGATACGGCGACCACCGAGATCTACACGCTACTACTGAGGATTATGGTAATTGTATGGCCATGGAGCCGGA	P2_ 2
bcatbc11	AATGATACGGCGACCACCGAGATCTACACGCTAATTCACCTCCTTATGGTAATTGTATGGCCATGGAGCCGGA	P3 _2
bcatbc12	AATGATACGGCGACCACCGAGATCTACACGCTCGTATAAATGCGTATGGTAATTGTATGGCCATGGAGCCGGA	P4 _2
bcatbc13	AATGATACGGCGACCACCGAGATCTACACGCTATGCTGCAACACTATGGTAATTGTATGGCCATGGAGCCGGA	P5 _2
bcatbc14	AATGATACGGCGACCACCGAGATCTACACGCTACTCGCTCGCTGTATGGTAATTGTATGGCCATGGAGCCGGA	P6 _2
bcatbc15	AATGATACGGCGACCACCGAGATCTACACGCTTTCCTTAGTAGTTATGGTAATTGTATGGCCATGGAGCCGGA	P7 _2
bcatbc16	AATGATACGGCGACCACCGAGATCTACACGCTCGTCCGTATGAATATGGTAATTGTATGGCCATGGAGCCGGA	Plasmid _2

bcatsat/R	CAAGCAGAAGACGGCATACGAGATAGTCAGCCAGCCTGTC AACATCTTCTTCTTCGGGA	used for 16 samples

Table 7-55: Primers used for second short PCR.

Reaction components	Volume
Q5 Hot Start High-Fidelity 2X Master Mix (NEB)	12.5µl
Forward primer	0.7µl
Reverse primer	0.7µl
DNA	1.0µl
dH2O	10.1µl

Table 7-56: PCR Reaction mix for second short PCR

Step	Temperature	Time
1	98°C	2min
2	98°C	10s
3	67°C	15s
4	72°C	25s
5	Cycle to step 2 for 13 more times	
6	72°C	1min
7	16°C	∞

Table 7-57: Thermocycler parameters for second short PCR

7.11.3.6 Purification of PCR products

The second short PCR products from each segment was cleaned separately using magnetic beads (Agencourt AMPure XP, Beckman Coulter Life Sciences) using the manufacturers protocol.

7.11.3.7 Quantitation and final pooling of samples for deep sequencing

For sequencing purpose required accurate pooling of the 16 segments. Hence the purified second short PCR products were quantitated using Qubit 3.0 fluorometer. The

samples were processed using Qubit dsDNA HS Assay kit (Life Technologies) according to the manufacturer's protocol.

Finally, equimolar ratios of the PCR products from all the 16 segments were pooled to obtain a final concentration of 5nM. Following pooling of the samples, the integrity of DNA was assessed using a bioanalyzer. A high sensitivity DNA assay was performed, and the quality of the amplified product was analysed. The Bioanalyzer assay was done at the MRC sequencing facility.

7.11.4 Deep sequencing

Sequencing was done by Edinburgh genomics. A paired end sequencing (200bp read1 and 100 read2) was performed using the Illumina MiSeq system. The sequencing primers are given in table 7-58. The primers were ordered as ultramers of 4nM from IDT and reconstituted in TE buffer and submitted at concentration of 100µM.

QC-Before passing the samples for the MiSeq run another QC check was performed by Edinburgh genomics which included a qPCR and another round of Bioanalyser- high sensitivity DNA assay.

Primer	Sequence
Read 1 seq	TATGGTAATTGTATGGCCATGGAGCCGGA
Read 2 seq	AGTCAGCCAGCCTGTCAACATCTTCTTCTTCGGGA
Index seq	AATGATACGGCGACCACCGAGATCTACACGCT

Table 7-58: Sequencing primers used for MiSeq.

7.11.5 Processing and analysis of Deep sequencing data

The processing of raw sequencing data and analysis was done by Martijn Kelder PhD student in Andrew Wood's lab (MRC Human Genetics Unit, IGMM, The University of Edinburgh). Quality control of reads in raw fastq files was done using fastqc (see sheet ‘FastQC’ in BCat_fastqc_summary.xlsx). As forward and reverse reads were respectively, 200nt and 100nt in length, the forward reads were stripped down to 100nt using Python.

Reads were then trimmed for adapter sequences (<1 percent of reads) using Trimgalore and aligned to the reference sequence using BowTie2. Forward and reverse reads were merged into a single 100nt contig using Python. Only contigs without indels and with consensus between forward and reverse read for each position were passed on for further analysis. These files were then used for further analysis described below.

Amino acid substitutions were counted for each saturated codon using Python and the reads were then normalized to the pool sample.

7.11.6 Regression analysis

The statistical regression analysis of the samples normalized to pool was performed by Helen Brown (senior statistician at The Roslin Institute) and the slope (m), SE and p values were calculated. The heatmap was plotted using heatmapper (http://www.heatmapper.ca/).

7.11.7 Analysis of background mutation rate

The analysis of the background mutational rate of β-catenin was performed by Ailith Ewing in Colin Semple's group in IGMM. The background mutational profiles for liver and endometrial cancers were calculated as follows:

7.11.7.1 Mutational data

The somatic mutation calls taken from the whole exome sequence for each cancer were calculated by the TCGA project as follows: the tumour sequence was aligned to the reference exome (GRCh38) and the normal sequence was aligned to the reference exome. Then, any mutations that occur in the tumour but not the normal were called as single nucleotide variants (SNVs). TCGA use four SNV callers: muse, mutect2, varscan and somaticsniper. These calls can be downloaded as .maf files. Next, an ensemble approach of these calls were taken and only mutations that had been called by at least 2 of these callers were considered. Then, the samples were filtered to only include those with a mutation in the β-catenin region in question. Then, for the purposes of calculating relative frequencies and probabilities non-synonymous mutations (missense and nonsense) were excluded in an effort to make these mutations mostly selectively neutral.

After these steps there were 4996 mutations across the liver HCCs and 33,223 mutations across the endometrial cancers.

7.11.7.2 Relative trinucleotide contexts

Then, each nucleotide change was annotated with its trinucleotide context using the reference exome (GRCh38). With this the relative frequencies of each of the 96 possible mutations accounting for trinucleotide context were calculated. For each of the 96 types the relative frequency is: (number of that type)/ (total number of mutations). Therefore, the relative frequencies sum to 1.

7.11.7.3 Calculation of the likelihood of particular amino acid substitution

For a given amino acid substitution, the likelihood of that amino acid substitution in a given cancer type was calculated combining the three possible scenarios (only requires one SNV, requires several SNVs, can be achieved by 2+ paths) along with the SNV rates in their relative trinucleotide context discussed above.

Appendix

Appendix 1A Statistical analysis result of test performed to determine if there are significant differences between the proportions of mutation at each of the top 18 residues across the cancer types

residue	pvalue	test_stat	Significant?
T41	0.001	1705.41965	Yes
S45	0.001	1084.57036	Yes
I35	0.001	790.162202	Yes
S37	0.001	722.367373	Yes
D32	0.001	600.984042	Yes
S33	0.001	589.631303	Yes
K49	0.001	516.482237	Yes
T42	0.001	516.482237	Yes
E53	0.001	452.509655	Yes
G34	0.001	313.366331	Yes
T40	0.001	311.100107	Yes
A43	0.001	189.779571	Yes
P44	0.001	157.942864	Yes
G48	0.002	146.221487	Yes
G24	0.002	144.049664	Yes
H36	0.001	105.544157	Yes
K33	0.054	52.1975665	No
S47	0.12	42.6454752	No

Appendix 1B Statistical analysis result of test performed to determine if there are significant differences between the proportions of the observed amino acid variants for each of the top 6 residues across the cancer types

residue	pvalue	mutation	test_stat	Significant?	
D32	0.008	D32E	123.051823	Yes	
D32	0.001	D32Y	113.326286	Yes	
D32	0.001	D32H	80.8909966	Yes	
D32	0.001	D32N	52.9578491	Yes	
D32	0.001	D32G	52.852131	Yes	
D32	0.039	D32V	33.0106119	Yes	
D32	0.597	D32A	13.1892563	No	
D32	0.318	D32d	7.30693307	No	
G34	0.001	G34R	105.753345	Yes	
G34	0.001	G34V	71.3603046	Yes	
G34	0.001	G34E	65.6808879	Yes	
G34	0.092	G34I	25.5804416	No	
S33	0.001	S33C	168.653133	Yes	
S33	0.002	S33L	146.013947	Yes	
S33	0.001	S33Y	110.981388	Yes	
S33	0.001	S33F	98.0592472	Yes	
S33	0.058	S33T	67.5005577	No	
S33	0.079	S33N	58.6667995	No	
S33	0.004	S33P	47.2505151	Yes	
S33	0.492	S33A	19.8924563	No	
S37	0.001	S37A	227.067767	Yes	
S37	0.001	S37F	217.260869	Yes	
S37	0.001	S37C	161.974547	Yes	
S37	0.048	S37Y	37.305067	Yes	

S37	0.277	S37P	25.078841	No	
S45	0.001	S45F	785.344011	Yes	
S45	0.001	S45d	250.491087	Yes	
S45	0.006	S45C	79.5602774	Yes	
S45	0.001	S45P	77.7361193	Yes	
S45	0.006	S45Y	74.9261189	Yes	
S45	0.065	S45A	49.1220549	No	
S45	0.202	S45T	12.4740222	No	
S45	0.198	S45E	5.76061455	No	
T41	0.001	T41A	1867.6365	Yes	
T41	0.001	T41I	476.556462	Yes	
T41	0.001	T41P	273.371509	Yes	
T41	0.001	T41S	183.188424	Yes	
T41	0.001	T41N	85.3267468	Yes	

Appendix 1C Statistical analysis result of test performed to determine whether the observed frequency of an individual residue (top 18 and other) is different than expected, given its expected frequency across all tumour types

residue	site	total	observed	obs_propn	expected_propn	ratio_propns	p
A43	Haematopoietic and lymphoid	49	5	0.102040816	0.003070175	33.2361516	4.65E-07
A43	Pituitary	110	3	0.027272727	0.003070175	8.883116883	0.00489
A43	Large intestine	371	3	0.008086253	0.003070175	2.63380824	0.107293
A43	Thyroid	62	1	0.016129032	0.003070175	5.253456221	0.173573
A43	Endometrium	326	1	0.003067485	0.003070175	0.999123576	0.633009
A43	Soft tissue	1325	1	0.000754717	0.003070175	0.245822102	0.982995
D32	CNS	295	85	0.288135593	0.100219298	2.875050996	0

D32	Pancreas	143	48	0.335664336	0.100219298	3.349298404	1.87E-14
D32	Liver	892	134	0.150224215	0.100219298	1.498954971	1.81E-06
D32	Stomach	182	37	0.203296703	0.100219298	2.028518527	2.48E-05
D32	Pituitary	110	25	0.227272727	0.100219298	2.267754128	7.52E-05
D32	Endometrium	326	55	0.168711656	0.100219298	1.683424843	9.4E-05
D32	Prostate	25	7	0.28	0.100219298	2.793873085	0.009588
D32	Testis	10	4	0.4	0.100219298	3.991247265	0.012893
D32	Ovary	115	17	0.147826087	0.100219298	1.475026163	0.066872
D32	Skin	143	20	0.13986014	0.100219298	1.395541002	0.079659
D32	Urinary tract	19	3	0.157894737	0.100219298	1.575492341	0.295737
D32	Oesophagus	11	1	0.090909091	0.100219298	0.907101651	0.687029
D32	Small intestine	12	1	0.083333333	0.100219298	0.831509847	0.718395
D32	Bone	25	2	0.08	0.100219298	0.798249453	0.729958
D32	Biliary tract	45	2	0.044444444	0.100219298	0.443471918	0.948098
D32	Haematopoietic and lymphoid	49	1	0.020408163	0.100219298	0.203635065	0.994342
D32	Lung	71	2	0.028169014	0.100219298	0.281073751	0.995063
D32	Large intestine	371	6	0.016172507	0.100219298	0.161371183	1
D32	Soft tissue	1325	7	0.005283019	0.100219298	0.052714587	1
E53	Haematopoietic and lymphoid	49	8	0.163265306	0.002850877	57.26844584	1.77E-12
E53	Large intestine	371	4	0.010781671	0.002850877	3.781878499	0.02254
E53	Soft tissue	1325	1	0.000754717	0.002850877	0.264731495	0.977241
G245	Large intestine	371	13	0.035040431	0.002850877	12.29110512	1.04E-10
G34	Liver	892	122	0.1367713	0.069517544	1.967435742	1.24E-12
G34	CNS	295	45	0.152542373	0.069517544	2.19430038	6.61E-07
G34	Endometrium	326	47	0.144171779	0.069517544	2.073890577	1.93E-06
G34	Pancreas	143	22	0.153846154	0.069517544	2.213055084	0.000376

G34	Ovary	115	14	0.12173913	0.069517544	1.75120011	0.028866
G34	Stomach	182	20	0.10989011	0.069517544	1.580753631	0.028886
G34	Testis	10	2	0.2	0.069517544	2.876971609	0.150031
G34	Skin	143	12	0.083916084	0.069517544	1.207120955	0.292095
G34	Bone	25	2	0.08	0.069517544	1.150788644	0.526575
G34	Prostate	25	2	0.08	0.069517544	1.150788644	0.526575
G34	Lung	71	5	0.070422535	0.069517544	1.013018172	0.554171
G34	Breast	15	1	0.066666667	0.069517544	0.958990536	0.66067
G34	Biliary tract	45	2	0.044444444	0.069517544	0.639327024	0.829566
G34	Pituitary	110	5	0.045454546	0.069517544	0.653857184	0.887066
G34	Large intestine	371	12	0.032345014	0.069517544	0.465278427	0.999359
G34	Soft tissue	1325	4	0.003018868	0.069517544	0.043425987	1
G48	Haematopoietic and lymphoid	49	4	0.081632653	0.002850877	28.63422292	1.26E-05
G48	Stomach	182	4	0.021978022	0.002850877	7.709213863	0.001953
G48	Large intestine	371	3	0.008086253	0.002850877	2.836408874	0.090949
G48	Thyroid	62	1	0.016129032	0.002850877	5.657568238	0.162226
G48	Skin	143	1	0.006993007	0.002850877	2.452931684	0.335192
H36	Liver	892	31	0.034753363	0.010087719	3.445116007	6.04E-09
H36	Kidney	197	9	0.045685279	0.010087719	4.528801589	0.000203
H36	Thyroid	62	1	0.016129032	0.010087719	1.59887798	0.46667
H36	Stomach	182	2	0.010989011	0.010087719	1.089345437	0.549019
H36	Skin	143	1	0.006993007	0.010087719	0.693219824	0.765399
H36	Large intestine	371	1	0.002695418	0.010087719	0.267197937	0.976752
H36	Soft tissue	1325	1	0.000754717	0.010087719	0.074815423	0.999999
I35	Salivary gland	14	12	0.857142857	0.014254386	60.13186813	0
I35	Liver	892	39	0.043721973	0.014254386	3.067264574	1.71E-09

I35	Adrenal gland	65	2	0.030769231	0.014254386	2.158579882	0.237039
I35	Lung	71	1	0.014084507	0.014254386	0.98808234	0.639168
I35	Pituitary	110	1	0.009090909	0.014254386	0.637762238	0.793873
I35	Ovary	115	1	0.008695652	0.014254386	0.610033445	0.808152
I35	CNS	295	2	0.006779661	0.014254386	0.475619296	0.923774
I35	Kidney	197	1	0.005076142	0.014254386	0.356110894	0.940888
I35	Endometrium	326	2	0.006134969	0.014254386	0.430391694	0.946998
I35	Large intestine	371	2	0.005390836	0.014254386	0.37818785	0.969058
I35	Soft tissue	1325	2	0.001509434	0.014254386	0.105892598	1
K335	Liver	892	14	0.015695067	0.003289474	4.771300448	2.48E-06
K335	Kidney	197	1	0.005076142	0.003289474	1.543147208	0.477481
K49	Thyroid	62	10	0.161290323	0.003070175	52.53456221	6.88E-15
K49	Large intestine	371	4	0.010781671	0.003070175	3.51174432	0.028496
Other	Large intestine	371	102	0.274932615	0.067982456	4.044170072	0
Other	Thyroid	62	25	0.403225807	0.067982456	5.93132154	7.84E-14
Other	Haematopoietic and lymphoid	49	20	0.408163265	0.067982456	6.003949967	1.81E-11
Other	Lung	71	24	0.338028169	0.067982456	4.972285325	2.13E-11
Other	Stomach	182	33	0.181318681	0.067982456	2.667139312	2.39E-07
Other	Bone	25	7	0.28	0.067982456	4.118709677	0.001081
Other	Breast	15	5	0.333333333	0.067982456	4.903225807	0.002445
Other	Oesophagus	11	4	0.363636364	0.067982456	5.348973607	0.004779
Other	Ovary	115	10	0.086956522	0.067982456	1.279102384	0.255894
Other	Prostate	25	2	0.08	0.067982456	1.176774194	0.51427
Other	Salivary gland	14	1	0.071428571	0.067982456	1.050691244	0.626804
Other	Urinary tract	19	1	0.052631579	0.067982456	0.774193548	0.737543
Other	Skin	143	7	0.048951049	0.067982456	0.720054139	0.860398

Other	Endometrium	326	17	0.052147239	0.067982456	0.767069068	0.897528
Other	Biliary tract	45	1	0.022222222	0.067982456	0.32688172	0.957919
Other	Kidney	197	7	0.035532995	0.067982456	0.522678893	0.982472
Other	Pancreas	143	1	0.006993007	0.067982456	0.102864877	0.999958
Other	Liver	892	30	0.033632287	0.067982456	0.494720093	0.999997
Other	Adrenal gland	310	3	0.009677419	0.067982456	0.142351717	1
Other	CNS	295	2	0.006779661	0.067982456	0.099726627	1
Other	Soft tissue	1325	8	0.006037736	0.067982456	0.088813147	1
P44	Thyroid	62	5	0.080645161	0.004385965	18.38709677	8.53E-06
P44	Kidney	197	5	0.025380711	0.004385965	5.786802031	0.001902
P44	Adrenal gland	20	2	0.1	0.004385965	22.8	0.003468
P44	Haematopoietic and lymphoid	49	1	0.020408163	0.004385965	4.653061225	0.193769
P44	Large intestine	371	3	0.008086253	0.004385965	1.843665768	0.223478
P44	Skin	143	1	0.006993007	0.004385965	1.594405594	0.466647
P44	Stomach	182	1	0.005494506	0.004385965	1.252747253	0.550672
P44	Liver	892	2	0.002242153	0.004385965	0.511210762	0.90228
S33	CNS	295	111	0.376271186	0.107894737	3.487391484	0
S33	Pituitary	110	28	0.254545455	0.107894737	2.359201774	1.19E-05
S33	Endometrium	326	58	0.17791411	0.107894737	1.648960048	0.000103
S33	Ovary	115	23	0.2	0.107894737	1.853658537	0.002589
S33	Liver	892	123	0.137892377	0.107894737	1.278026906	0.00302
S33	Pancreas	143	25	0.174825175	0.107894737	1.620330889	0.010522
S33	Skin	143	24	0.167832168	0.107894737	1.555517653	0.019203
S33	Stomach	182	29	0.159340659	0.107894737	1.476815867	0.021255
S33	Biliary tract	45	9	0.2	0.107894737	1.853658537	0.048523
S33	Breast	15	4	0.266666667	0.107894737	2.471544715	0.070031

S33	Bone	25	5	0.2	0.107894737	1.853658537	0.125247
S33	Prostate	25	4	0.16	0.107894737	1.482926829	0.281141
S33	Parathyroid	11	2	0.181818182	0.107894737	1.685144124	0.33625
S33	Urinary tract	19	2	0.105263158	0.107894737	0.975609756	0.623168
S33	Lung	71	7	0.098591549	0.107894737	0.913775335	0.656058
S33	Testis	10	1	0.1	0.107894737	0.926829268	0.680728
S33	Oesophagus	11	1	0.090909091	0.107894737	0.842572062	0.715176
S33	Haematopoietic and lymphoid	49	2	0.040816327	0.107894737	0.378297661	0.974243
S33	Large intestine	371	24	0.064690027	0.107894737	0.599566104	0.998403
S33	Thyroid	62	1	0.016129032	0.107894737	0.149488592	0.999157
S33	Kidney	197	3	0.015228426	0.107894737	0.141141513	1
S33	Soft tissue	1325	6	0.004528302	0.107894737	0.041969627	1
S37	Endometrium	326	83	0.254601227	0.112938597	2.254333194	9.20E-13
S37	Stomach	182	50	0.274725275	0.112938597	2.432518937	1.71E-09
S37	Ovary	115	36	0.313043478	0.112938597	2.771802448	7.42E-09
S37	Lung	71	25	0.352112676	0.112938597	3.11773554	1.08E-07
S37	Biliary tract	45	19	0.422222222	0.112938597	3.738511327	1.30E-07
S37	Parathyroid	11	9	0.818181818	0.112938597	7.244483672	1.33E-07
S37	Small intestine	12	9	0.75	0.112938597	6.640776699	4.77E-07
S37	Skin	143	36	0.251748252	0.112938597	2.229071899	2.85E-06
S37	Pancreas	143	34	0.237762238	0.112938597	2.105234571	1.96E-05
S37	Urinary tract	19	9	0.473684211	0.112938597	4.194174757	9.51E-05
S37	Pituitary	110	23	0.209090909	0.112938597	1.851368049	0.002555
S37	Oesophagus	11	5	0.454545455	0.112938597	4.024713151	0.004714
S37	CNS	295	45	0.152542373	0.112938597	1.350666447	0.023219
S37	Testis	10	3	0.3	0.112938597	2.65631068	0.094191

S37	Liver	892	96	0.107623318	0.112938597	0.952936567	0.707091
S37	Breast	15	1	0.066666667	0.112938597	0.590291262	0.834307
S37	Thyroid	62	4	0.064516129	0.112938597	0.571249609	0.930252
S37	Bone	25	1	0.04	0.112938597	0.354174757	0.950015
S37	Prostate	25	1	0.04	0.112938597	0.354174757	0.950015
S37	Large intestine	371	17	0.045822102	0.112938597	0.4057258	0.999998
S37	Adrenal gland	515	3	0.005825243	0.112938597	0.051578848	1
S37	Soft tissue	1325	6	0.004528302	0.112938597	0.040095256	1
S45	Kidney	197	151	0.766497462	0.239473684	3.200758632	0
S45	Soft tissue	1325	510	0.38490566	0.239473684	1.607298362	0
S45	Large intestine	371	107	0.288409704	0.239473684	1.204348212	0.017278
S45	Urinary tract	19	3	0.157894737	0.239473684	0.659340659	0.868101
S45	Bone	25	3	0.12	0.239473684	0.501098901	0.958825
S45	Prostate	25	3	0.12	0.239473684	0.501098901	0.958825
S45	Skin	143	24	0.167832168	0.239473684	0.700837624	0.985236
S45	Biliary tract	45	2	0.044444444	0.239473684	0.185592186	0.999932
S45	Lung	71	5	0.070422535	0.239473684	0.294072125	0.999958
S45	Liver	892	164	0.183856502	0.239473684	0.767752427	0.999974
S45	Haematopoietic and lymphoid	49	1	0.020408163	0.239473684	0.085220902	0.999999
S45	Thyroid	62	2	0.032258065	0.239473684	0.134704006	0.999999
S45	Ovary	115	5	0.043478261	0.239473684	0.181557573	1
S45	Pituitary	110	4	0.036363636	0.239473684	0.151848152	1
S45	Pancreas	143	5	0.034965035	0.239473684	0.146007838	1
S45	Endometrium	326	26	0.079754601	0.239473684	0.333041192	1
S45	Adrenal gland	1092	73	0.066849817	0.239473684	0.279153081	1
S45	CNS	295	2	0.006779661	0.239473684	0.028310672	1

S45	Stomach	182	2	0.010989011	0.239473684	0.045888178	1
S47	Large intestine	371	5	0.013477089	0.002850877	4.727348124	0.004547
S47	Thyroid	62	2	0.032258065	0.002850877	11.31513648	0.013722
S47	Pancreas	143	1	0.006993007	0.002850877	2.452931684	0.335192
S47	Skin	143	1	0.006993007	0.002850877	2.452931684	0.335192
S47	Stomach	182	1	0.005494506	0.002850877	1.927303466	0.405242
S47	Liver	892	3	0.003363229	0.002850877	1.179717144	0.467361
T40	Thyroid	62	8	0.129032258	0.003508772	36.77419355	6.56E-11
T40	Haematopoietic and lymphoid	49	2	0.040816327	0.003508772	11.63265306	0.012979
T40	Bone	25	1	0.04	0.003508772	11.4	0.084123
T40	CNS	295	1	0.003389831	0.003508772	0.966101695	0.645451
T40	Endometrium	326	1	0.003067485	0.003508772	0.874233129	0.682053
T40	Large intestine	371	1	0.002695418	0.003508772	0.76819407	0.728568
T40	Soft tissue	1325	2	0.001509434	0.003508772	0.430188679	0.946223
T41	Soft tissue	1325	777	0.586415094	0.248903509	2.355993683	0
T41	Breast	15	4	0.266666667	0.248903509	1.071365639	0.534757
T41	Prostate	25	6	0.24	0.248903509	0.964229075	0.616896
T41	Biliary tract	45	10	0.222222222	0.248903509	0.892804699	0.714338
T41	Small intestine	12	2	0.166666667	0.248903509	0.669603524	0.83957
T41	Bone	25	4	0.16	0.248903509	0.642819383	0.901706
T41	Pituitary	110	21	0.190909091	0.248903509	0.767000401	0.938892
T41	Salivary gland	14	1	0.071428571	0.248903509	0.286972939	0.981814
T41	Urinary tract	19	1	0.052631579	0.248903509	0.211453745	0.995653
T41	Haematopoietic and lymphoid	49	5	0.102040816	0.248903509	0.409961341	0.997265
T41	Large intestine	371	64	0.172506739	0.248903509	0.693066721	0.999839
T41	Skin	143	16	0.111888112	0.248903509	0.449524044	0.999988

T41	Ovary	115	9	0.07826087	0.248903509	0.314422524	0.999999
T41	Thyroid	62	2	0.032258065	0.248903509	0.129600682	1
T41	Kidney	197	20	0.101522843	0.248903509	0.407880319	1
T41	Lung	71	2	0.028169014	0.248903509	0.113172427	1
T41	Endometrium	326	36	0.110429448	0.248903509	0.443663685	1
T41	Pancreas	143	7	0.048951049	0.248903509	0.196666769	1
T41	Liver	892	134	0.150224215	0.248903509	0.603543984	1
T41	Adrenal gland	1135	9	0.007929515	0.248903509	0.031857789	1
T41	CNS	295	2	0.006779661	0.248903509	0.02723811	1
T41	Stomach	182	3	0.016483517	0.248903509	0.066224524	1

***Statistically significant scores (p<0.05) are highlighted**

Appendix 1D Statistical analysis result of test performed to determine whether the observed frequency of an individual amino acid variant (of the top 6 residues) is different than expected, given its expected frequency across all tumour types

residue	mutation	site	total	observed	obs_propn	expected_propn	ratio_propns	p
D32	D32A	Liver	134	11	0.082089552	0.041575492	1.974469757	0.02456
D32	D32A	Pancreas	48	3	0.0625	0.041575492	1.503289474	0.322073
D32	D32A	Ovary	17	1	0.058823529	0.041575492	1.414860681	0.51417
D32	D32A	Endometrium	55	2	0.036363636	0.041575492	0.874641148	0.672394
D32	D32A	CNS	85	2	0.023529412	0.041575492	0.565944272	0.873136
D32	D32E	Small intestine	1	1	1	0.008752735	114.25	0.008753
D32	D32E	Skin	20	1	0.05	0.008752735	5.7125	0.161236
D32	D32E	Stomach	37	1	0.027027027	0.008752735	3.087837838	0.277673
D32	D32E	Pancreas	48	1	0.020833333	0.008752735	2.380208333	0.344253
D32	D32G	Large intestine	6	4	0.666666667	0.164113786	4.062222222	0.008219

D32	D32G	Liver	134	33	0.246268657	0.164113786	1.500597015	0.009513
D32	D32G	Lung	2	1	0.5	0.164113786	3.046666667	0.301294
D32	D32G	Pancreas	48	9	0.1875	0.164113786	1.1425	0.388371
D32	D32G	Stomach	37	7	0.189189189	0.164113786	1.152792793	0.405956
D32	D32G	Urinary tract	3	1	0.333333333	0.164113786	2.031111111	0.415961
D32	D32G	Skin	20	4	0.2	0.164113786	1.218666667	0.421048
D32	D32G	Ovary	17	3	0.176470588	0.164113786	1.075294118	0.545144
D32	D32G	Prostate	7	1	0.142857143	0.164113786	0.870476191	0.714878
D32	D32G	Soft tissue	7	1	0.142857143	0.164113786	0.870476191	0.714878
D32	D32G	Pituitary	25	1	0.04	0.164113786	0.243733333	0.988684
D32	D32G	CNS	85	7	0.082352941	0.164113786	0.501803922	0.990703
D32	D32G	Endometrium	55	3	0.054545455	0.164113786	0.332363636	0.996393
D32	D32H	Pituitary	25	12	0.48	0.120350109	3.988363636	1.05E-05
D32	D32H	Stomach	37	11	0.297297297	0.120350109	2.47027027	0.003268
D32	D32H	Testis	4	2	0.5	0.120350109	4.154545455	0.073589
D32	D32H	Haematopoietic and lymphoid	1	1	1	0.120350109	8.309090909	0.12035
D32	D32H	Urinary tract	3	1	0.333333333	0.120350109	2.76969697	0.319341
D32	D32H	Prostate	7	1	0.142857143	0.120350109	1.187012987	0.592461
D32	D32H	Ovary	17	2	0.117647059	0.120350109	0.977540107	0.624013
D32	D32H	Endometrium	55	6	0.109090909	0.120350109	0.906446281	0.662823
D32	D32H	Pancreas	48	4	0.083333333	0.120350109	0.692424242	0.845098
D32	D32H	Skin	20	1	0.05	0.120350109	0.415454546	0.923052
D32	D32H	Liver	134	10	0.074626866	0.120350109	0.620081411	0.967856
D32	D32H	CNS	85	4	0.047058824	0.120350109	0.391016043	0.993862
D32	D32N	CNS	85	29	0.341176471	0.221006565	1.54373908	0.007435
D32	D32N	Oesophagus	1	1	1	0.221006565	4.524752475	0.221007

D32	D32N	Pancreas	48	13	0.270833333	0.221006565	1.225453795	0.249477
D32	D32N	Pituitary	25	7	0.28	0.221006565	1.266930693	0.307088
D32	D32N	Bone	2	1	0.5	0.221006565	2.262376238	0.393169
D32	D32N	Endometrium	55	13	0.236363636	0.221006565	1.069486949	0.443571
D32	D32N	Testis	4	1	0.25	0.221006565	1.131188119	0.631756
D32	D32N	Stomach	37	7	0.189189189	0.221006565	0.856034252	0.739672
D32	D32N	Large intestine	6	1	0.166666667	0.221006565	0.754125413	0.776538
D32	D32N	Prostate	7	1	0.142857143	0.221006565	0.646393211	0.825925
D32	D32N	Liver	134	24	0.179104478	0.221006565	0.810403428	0.900996
D32	D32N	Skin	20	1	0.05	0.221006565	0.226237624	0.993229
D32	D32N	Biliary tract	101	2	0.01980198	0.221006565	0.089599059	1
D32	D32V	Liver	134	23	0.171641791	0.10940919	1.56880597	0.019539
D32	D32V	Prostate	7	2	0.285714286	0.10940919	2.611428571	0.173488
D32	D32V	Soft tissue	7	2	0.285714286	0.10940919	2.611428571	0.173488
D32	D32V	Testis	4	1	0.25	0.10940919	2.285	0.37091
D32	D32V	Skin	20	3	0.15	0.10940919	1.371	0.376852
D32	D32V	Pancreas	48	6	0.125	0.10940919	1.1425	0.430534
D32	D32V	CNS	85	10	0.117647059	0.10940919	1.075294118	0.454226
D32	D32V	Pituitary	25	2	0.08	0.10940919	0.7312	0.775258
D32	D32V	Endometrium	55	1	0.018181818	0.10940919	0.166181818	0.998293
D32	D32Y	Endometrium	55	29	0.527272727	0.332603939	1.585287081	0.002252
D32	D32Y	Ovary	17	11	0.647058824	0.332603939	1.945433437	0.007862
D32	D32Y	Skin	20	10	0.5	0.332603939	1.503289474	0.090714
D32	D32Y	CNS	85	33	0.388235294	0.332603939	1.167260062	0.164938
D32	D32Y	Soft tissue	7	4	0.571428571	0.332603939	1.718045113	0.172178
D32	D32Y	Bone	2	1	0.5	0.332603939	1.503289474	0.554582
D32	D32Y	Lung	2	1	0.5	0.332603939	1.503289474	0.554582

D32	D32Y	Urinary tract	3	1	0.333333333	0.332603939	1.002192983	0.70273
D32	D32Y	Stomach	37	11	0.297297297	0.332603939	0.893847795	0.731491
D32	D32Y	Prostate	7	2	0.285714286	0.332603939	0.859022556	0.735279
D32	D32Y	Large intestine	6	1	0.166666667	0.332603939	0.501096491	0.911631
D32	D32Y	Pancreas	48	12	0.25	0.332603939	0.751644737	0.917233
D32	D32Y	Liver	134	33	0.246268657	0.332603939	0.740426159	0.98815
D32	D32Y	Pituitary	25	3	0.12	0.332603939	0.360789474	0.99642
D32	D32del	Endometrium	55	1	0.018181818	0.002188184	8.309090909	0.113507
G34	G34E	Stomach	20	10	0.5	0.302839117	1.651041667	0.050944
G34	G34E	Skin	12	6	0.5	0.302839117	1.651041667	0.1224
G34	G34E	CNS	45	16	0.355555556	0.302839117	1.174074074	0.267518
G34	G34E	Endometrium	47	16	0.340425532	0.302839117	1.124113475	0.337582
G34	G34E	Prostate	2	1	0.5	0.302839117	1.651041667	0.513967
G34	G34E	Testis	2	1	0.5	0.302839117	1.651041667	0.513967
G34	G34E	Large intestine	12	4	0.333333333	0.302839117	1.100694444	0.516216
G34	G34E	Liver	122	37	0.303278689	0.302839117	1.001451503	0.529881
G34	G34E	Ovary	14	4	0.285714286	0.302839117	0.943452381	0.653449
G34	G34E	Lung	5	1	0.2	0.302839117	0.660416667	0.835311
G34	G34I	Skin	12	1	0.083333333	0.003154574	26.41666667	0.037205
G34	G34R	Pancreas	22	15	0.681818182	0.38170347	1.786250939	0.004201
G34	G34R	Soft tissue	4	4	1	0.38170347	2.619834711	0.021228
G34	G34R	CNS	45	24	0.533333333	0.38170347	1.397245179	0.027567
G34	G34R	Pituitary	5	4	0.8	0.38170347	2.095867769	0.073728
G34	G34R	Ovary	14	7	0.5	0.38170347	1.309917355	0.258808
G34	G34R	Bone	2	1	0.5	0.38170347	1.309917355	0.617709
G34	G34R	Testis	2	1	0.5	0.38170347	1.309917355	0.617709
G34	G34R	Lung	5	2	0.4	0.38170347	1.047933884	0.630716

G34	G34R	Skin	12	3	0.25	0.38170347	0.654958678	0.895237
G34	G34R	Liver	122	40	0.327868853	0.38170347	0.8589622	0.907168
G34	G34R	Stomach	20	5	0.25	0.38170347	0.654958678	0.929481
G34	G34R	Endometrium	47	13	0.276595745	0.38170347	0.724635133	0.951554
G34	G34R	Large intestine	12	2	0.166666667	0.38170347	0.436639119	0.973754
G34	G34V	Biliary tract	2	2	1	0.312302839	3.202020202	0.097533
G34	G34V	Liver	122	45	0.368852459	0.312302839	1.181073025	0.106714
G34	G34V	Large intestine	12	6	0.5	0.312302839	1.601010101	0.138317
G34	G34V	Endometrium	47	18	0.382978723	0.312302839	1.226305609	0.186091
G34	G34V	Breast	1	1	1	0.312302839	3.202020202	0.312303
G34	G34V	Lung	5	2	0.4	0.312302839	1.280808081	0.496941
G34	G34V	Bone	2	1	0.5	0.312302839	1.601010101	0.527073
G34	G34V	Prostate	2	1	0.5	0.312302839	1.601010101	0.527073
G34	G34V	Pancreas	22	7	0.318181818	0.312302839	1.01882461	0.55609
G34	G34V	Stomach	20	5	0.25	0.312302839	0.800505051	0.797337
G34	G34V	Pituitary	5	1	0.2	0.312302839	0.64040404	0.846189
G34	G34V	Ovary	14	3	0.214285714	0.312302839	0.686147186	0.861765
G34	G34V	Skin	12	2	0.166666667	0.312302839	0.533670034	0.92784
G34	G34V	CNS	45	5	0.111111111	0.312302839	0.355780022	0.99962
S33	S33A	Pituitary	28	4	0.142857143	0.040650407	3.514285714	0.025635
S33	S33A	Endometrium	58	5	0.086206897	0.040650407	2.120689655	0.086457
S33	S33A	Soft tissue	6	1	0.166666667	0.040650407	4.1	0.220419
S33	S33A	Liver	123	5	0.040650407	0.040650407	1	0.563155
S33	S33A	Pancreas	25	1	0.04	0.040650407	0.984	0.645658
S33	S33A	Stomach	29	1	0.034482759	0.040650407	0.848275862	0.699856
S33	S33A	CNS	111	3	0.027027027	0.040650407	0.664864865	0.83357
S33	S33C	Ovary	23	16	0.695652174	0.414634146	1.677749361	0.00606

S33	S33C	Lung	7	5	0.714285714	0.414634146	1.722689076	0.111115
S33	S33C	Parathyroid	2	2	1	0.414634146	2.411764706	0.171921
S33	S33C	Breast	4	3	0.75	0.414634146	1.808823529	0.196467
S33	S33C	Soft tissue	6	4	0.666666667	0.414634146	1.607843137	0.200042
S33	S33C	Stomach	29	14	0.482758621	0.414634146	1.164300203	0.286861
S33	S33C	Biliary tract	9	5	0.555555556	0.414634146	1.339869281	0.297874
S33	S33C	Pancreas	25	12	0.48	0.414634146	1.157647059	0.319745
S33	S33C	Oesophagus	1	1	1	0.414634146	2.411764706	0.414634
S33	S33C	Testis	1	1	1	0.414634146	2.411764706	0.414634
S33	S33C	Liver	123	52	0.422764228	0.414634146	1.019607843	0.46151
S33	S33C	Pituitary	28	12	0.428571429	0.414634146	1.033613445	0.512322
S33	S33C	Endometrium	58	23	0.396551724	0.414634146	0.956389452	0.657596
S33	S33C	CNS	111	44	0.396396396	0.414634146	0.956014838	0.685044
S33	S33C	Bone	5	2	0.4	0.414634146	0.964705882	0.687858
S33	S33C	Kidney	3	1	0.333333333	0.414634146	0.803921569	0.799423
S33	S33C	Prostate	4	1	0.25	0.414634146	0.602941177	0.882589
S33	S33C	Large intestine	24	3	0.125	0.414634146	0.301470588	0.99959
S33	S33C	Skin	24	3	0.125	0.414634146	0.301470588	0.99959
S33	S33F	Skin	24	14	0.583333333	0.237804878	2.452991453	0.0003
S33	S33F	CNS	111	39	0.351351351	0.237804878	1.477477478	0.004653
S33	S33F	Urinary tract	2	2	1	0.237804878	4.205128205	0.056551
S33	S33F	Thyroid	1	1	1	0.237804878	4.205128205	0.237805
S33	S33F	Prostate	4	2	0.5	0.237804878	2.102564103	0.241316
S33	S33F	Haematopoietic and lymphoid	2	1	0.5	0.237804878	2.102564103	0.419059
S33	S33F	Stomach	29	7	0.24137931	0.237804878	1.015030946	0.553411
S33	S33F	Kidney	3	1	0.333333333	0.237804878	1.401709402	0.557209

S33	S33F	Endometrium	58	13	0.224137931	0.237804878	0.942528736	0.646172
S33	S33F	Ovary	23	5	0.217391304	0.237804878	0.914158306	0.669442
S33	S33F	Pituitary	28	6	0.214285714	0.237804878	0.901098901	0.685506
S33	S33F	Bone	5	1	0.2	0.237804878	0.841025641	0.742765
S33	S33F	Pancreas	25	5	0.2	0.237804878	0.841025641	0.743511
S33	S33F	Lung	7	1	0.142857143	0.237804878	0.600732601	0.850561
S33	S33F	Liver	123	18	0.146341463	0.237804878	0.615384615	0.995498
S33	S33F	Large intestine	24	1	0.041666667	0.237804878	0.175213675	0.998522
S33	S33L	Biliary tract	9	4	0.444444444	0.012195122	36.44444444	2.65E-06
S33	S33L	Liver	123	2	0.016260163	0.012195122	1.333333333	0.443201
S33	S33N	Kidney	3	1	0.333333333	0.008130081	41	0.024192
S33	S33N	Lung	7	1	0.142857143	0.008130081	17.57142857	0.055541
S33	S33N	Endometrium	58	1	0.017241379	0.008130081	2.120689655	0.377164
S33	S33N	Liver	123	1	0.008130081	0.008130081	1	0.633621
S33	S33P	Liver	123	29	0.235772358	0.128048781	1.841269841	0.000755
S33	S33P	Pancreas	25	7	0.28	0.128048781	2.186666667	0.033371
S33	S33P	Pituitary	28	5	0.178571429	0.128048781	1.394557823	0.284819
S33	S33P	Endometrium	58	8	0.137931035	0.128048781	1.077175698	0.468849
S33	S33P	Bone	5	1	0.2	0.128048781	1.561904762	0.495965
S33	S33P	Stomach	29	3	0.103448276	0.128048781	0.807881773	0.736458
S33	S33P	Large intestine	24	2	0.083333333	0.128048781	0.650793651	0.831197
S33	S33P	Skin	24	2	0.083333333	0.128048781	0.650793651	0.831197
S33	S33P	Ovary	23	1	0.043478261	0.128048781	0.339544514	0.957212
S33	S33P	CNS	111	5	0.045045045	0.128048781	0.351780352	0.999098
S33	S33T	Breast	4	1	0.25	0.006097561	41	0.024168
S33	S33T	Soft tissue	6	1	0.166666667	0.006097561	27.33333333	0.036032
S33	S33T	Liver	123	1	0.008130081	0.006097561	1.333333333	0.528717

S33	S33Y	Large intestine	24	18	0.75	0.152439024	4.92	1.04E-10
S33	S33Y	CNS	111	20	0.18018018	0.152439024	1.181981982	0.242744
S33	S33Y	Haematopoietic and lymphoid	2	1	0.5	0.152439024	3.28	0.28164
S33	S33Y	Skin	24	5	0.208333333	0.152439024	1.366666667	0.298644
S33	S33Y	Prostate	4	1	0.25	0.152439024	1.64	0.483959
S33	S33Y	Bone	5	1	0.2	0.152439024	1.312	0.562624
S33	S33Y	Stomach	29	4	0.137931035	0.152439024	0.904827586	0.664593
S33	S33Y	Endometrium	58	8	0.137931035	0.152439024	0.904827586	0.675862
S33	S33Y	Liver	123	15	0.12195122	0.152439024	0.8	0.85811
S33	S33Y	Ovary	23	1	0.043478261	0.152439024	0.285217391	0.977719
S33	S33Y	Pituitary	28	1	0.035714286	0.152439024	0.234285714	0.990255
S37	S37A	Parathyroid	9	9	1	0.130097087	7.686567164	1.07E-08
S37	S37A	Large intestine	17	12	0.705882353	0.130097087	5.425812116	7.68E-08
S37	S37A	Small intestine	9	8	0.888888889	0.130097087	6.832504146	6.53E-07
S37	S37A	Stomach	50	12	0.24	0.130097087	1.844776119	0.024316
S37	S37A	Prostate	1	1	1	0.130097087	7.686567164	0.130097
S37	S37A	Soft tissue	6	1	0.166666667	0.130097087	1.281094527	0.566664
S37	S37A	Endometrium	83	10	0.120481928	0.130097087	0.926092429	0.651429
S37	S37A	Liver	96	9	0.09375	0.130097087	0.720615672	0.891273
S37	S37A	CNS	45	3	0.066666667	0.130097087	0.512437811	0.943577
S37	S37A	Ovary	36	2	0.055555556	0.130097087	0.427031509	0.95773
S37	S37C	Ovary	36	22	0.611111111	0.32038835	1.907407407	0.000313
S37	S37C	Pituitary	23	15	0.652173913	0.32038835	2.035573123	0.001107
S37	S37C	Biliary tract	19	11	0.578947368	0.32038835	1.807017544	0.017746
S37	S37C	Endometrium	83	33	0.397590361	0.32038835	1.240963855	0.083993
S37	S37C	Liver	96	37	0.385416667	0.32038835	1.202967172	0.105709

S37	S37C	Lung	25	11	0.44	0.32038835	1.373333333	0.143345
S37	S37C	Urinary tract	9	4	0.444444444	0.32038835	1.387205387	0.318213
S37	S37C	Bone	1	1	1	0.32038835	3.121212121	0.320388
S37	S37C	Breast	1	1	1	0.32038835	3.121212121	0.320388
S37	S37C	Pancreas	34	7	0.205882353	0.32038835	0.642602496	0.95186
S37	S37C	CNS	45	9	0.2	0.32038835	0.624242424	0.974882
S37	S37C	Large intestine	17	2	0.117647059	0.32038835	0.367201426	0.987313
S37	S37C	Skin	36	4	0.111111111	0.32038835	0.346801347	0.999171
S37	S37C	Stomach	50	5	0.1	0.32038835	0.312121212	0.999944
S37	S37C	Adrenal gland	165	3	0.018181818	0.32038835	0.056749311	1
S37	S37F	Skin	36	27	0.75	0.411650485	1.821933962	3.97E-05
S37	S37F	Oesophagus	5	5	1	0.411650485	2.429245283	0.011821
S37	S37F	Pancreas	34	21	0.617647059	0.411650485	1.500416204	0.012308
S37	S37F	Thyroid	4	4	1	0.411650485	2.429245283	0.028715
S37	S37F	Testis	3	3	1	0.411650485	2.429245283	0.069757
S37	S37F	Stomach	50	25	0.5	0.411650485	1.214622642	0.130484
S37	S37F	Lung	25	13	0.52	0.411650485	1.263207547	0.184198
S37	S37F	Soft tissue	6	4	0.666666667	0.411650485	1.619496855	0.195694
S37	S37F	Urinary tract	9	5	0.555555556	0.411650485	1.349580713	0.291378
S37	S37F	Endometrium	83	35	0.421686747	0.411650485	1.024380541	0.467815
S37	S37F	Biliary tract	19	8	0.421052632	0.411650485	1.022840119	0.553826
S37	S37F	CNS	45	17	0.377777778	0.411650485	0.917714885	0.727992
S37	S37F	Pituitary	23	7	0.304347826	0.411650485	0.739335521	0.897606
S37	S37F	Ovary	36	11	0.305555556	0.411650485	0.742269392	0.93044
S37	S37F	Liver	96	26	0.270833333	0.411650485	0.657920598	0.998545
S37	S37F	Large intestine	17	1	0.058823529	0.411650485	0.142896781	0.999879
S37	S37P	Liver	96	12	0.125	0.062135922	2.01171875	0.01594

S37	S37P	CNS	45	6	0.133333333	0.062135922	2.145833333	0.058625
S37	S37P	Large intestine	17	2	0.117647059	0.062135922	1.893382353	0.285501
S37	S37P	Pancreas	34	3	0.088235294	0.062135922	1.420036765	0.354678
S37	S37P	Small intestine	9	1	0.111111111	0.062135922	1.788194444	0.438617
S37	S37P	Stomach	50	3	0.06	0.062135922	0.965625	0.607985
S37	S37P	Pituitary	23	1	0.043478261	0.062135922	0.699728261	0.771325
S37	S37P	Endometrium	83	3	0.036144578	0.062135922	0.581701807	0.895586
S37	S37P	Skin	36	1	0.027777778	0.062135922	0.447048611	0.90068
S37	S37Y	CNS	45	10	0.222222222	0.075728155	2.934472935	0.00168
S37	S37Y	Liver	96	12	0.125	0.075728155	1.650641026	0.058897
S37	S37Y	Skin	36	4	0.111111111	0.075728155	1.467236467	0.289138
S37	S37Y	Stomach	50	5	0.1	0.075728155	1.320512821	0.327743
S37	S37Y	Soft tissue	6	1	0.166666667	0.075728155	2.200854701	0.376555
S37	S37Y	Pancreas	34	3	0.088235294	0.075728155	1.165158371	0.480919
S37	S37Y	Lung	25	1	0.04	0.075728155	0.528205128	0.860365
S37	S37Y	Ovary	36	1	0.027777778	0.075728155	0.366809117	0.941279
S37	S37Y	Endometrium	83	2	0.024096386	0.075728155	0.31819586	0.988689
S45	S45A	Liver	164	12	0.073170732	0.014652015	4.993902439	6.56E-06
S45	S45A	Adrenal gland	16	1	0.0625	0.014652015	4.265625	0.210351
S45	S45A	Endometrium	26	1	0.038461539	0.014652015	2.625	0.318712
S45	S45A	Large intestine	107	1	0.009345794	0.014652015	0.637850467	0.793894
S45	S45A	Kidney	151	1	0.006622517	0.014652015	0.451986755	0.892345
S45	S45C	Kidney	151	19	0.125827815	0.029304029	4.293874172	1.16E-07
S45	S45C	Stomach	2	1	0.5	0.029304029	17.0625	0.057749
S45	S45C	Endometrium	26	2	0.076923077	0.029304029	2.625	0.176281
S45	S45C	Liver	164	6	0.036585366	0.029304029	1.24847561	0.349496
S45	S45C	Skin	24	1	0.041666667	0.029304029	1.421875	0.510224

S45	S45C	Large intestine	107	1	0.009345794	0.029304029	0.318925234	0.958514
S45	S45C	Soft tissue	510	2	0.003921569	0.029304029	0.133823529	0.999996
S45	S45E	Kidney	151	1	0.006622517	0.000915751	7.23178808	0.129199
S45	S45F	Soft tissue	510	408	0.8	0.576923077	1.386666667	0
S45	S45F	Large intestine	107	73	0.682242991	0.576923077	1.182554517	0.016534
S45	S45F	Biliary tract	2	2	1	0.576923077	1.733333333	0.33284
S45	S45F	CNS	2	2	1	0.576923077	1.733333333	0.33284
S45	S45F	Haematopoietic and lymphoid	1	1	1	0.576923077	1.733333333	0.576923
S45	S45F	Ovary	5	3	0.6	0.576923077	1.04	0.641971
S45	S45F	Stomach	2	1	0.5	0.576923077	0.866666667	0.821006
S45	S45F	Endometrium	26	13	0.5	0.576923077	0.866666667	0.839541
S45	S45F	Skin	24	11	0.458333333	0.576923077	0.794444444	0.915791
S45	S45F	Urinary tract	3	1	0.333333333	0.576923077	0.577777778	0.924272
S45	S45F	Lung	5	1	0.2	0.576923077	0.346666667	0.986445
S45	S45F	Pancreas	5	1	0.2	0.576923077	0.346666667	0.986445
S45	S45F	Liver	164	59	0.359756098	0.576923077	0.623577236	1
S45	S45F	Adrenal gland	630	18	0.028571429	0.576923077	0.04952381	1
S45	S45F	Kidney	151	36	0.238410596	0.576923077	0.413245033	1
S45	S45P	Liver	164	68	0.414634146	0.26007326	1.594297492	1.2E-05
S45	S45P	Pituitary	4	3	0.75	0.26007326	2.883802817	0.056639
S45	S45P	Thyroid	2	2	1	0.26007326	3.845070423	0.067638
S45	S45P	Bone	3	2	0.666666667	0.26007326	2.563380282	0.167733
S45	S45P	Prostate	3	2	0.666666667	0.26007326	2.563380282	0.167733
S45	S45P	Urinary tract	3	2	0.666666667	0.26007326	2.563380282	0.167733
S45	S45P	Lung	5	2	0.4	0.26007326	1.538028169	0.388428
S45	S45P	Pancreas	5	2	0.4	0.26007326	1.538028169	0.388428

S45	S45P	Endometrium	26	7	0.269230769	0.26007326	1.035211268	0.532066
S45	S45P	Ovary	5	1	0.2	0.26007326	0.769014085	0.778209
S45	S45P	Skin	24	4	0.166666667	0.26007326	0.64084507	0.904677
S45	S45P	Kidney	151	28	0.185430464	0.26007326	0.712993191	0.987816
S45	S45P	Large intestine	107	14	0.130841122	0.26007326	0.503093326	0.999637
S45	S45P	Soft tissue	510	99	0.194117647	0.26007326	0.746396023	0.999806
S45	S45P	Adrenal gland	284	48	0.169014085	0.26007326	0.649871057	0.999902
S45	S45T	Liver	164	3	0.018292683	0.003663004	4.993902439	0.022934
S45	S45T	Large intestine	107	1	0.009345794	0.003663004	2.551401869	0.324742
S45	S45Y	Liver	164	15	0.091463415	0.034798535	2.628369705	0.000664
S45	S45Y	Skin	24	5	0.208333333	0.034798535	5.986842105	0.001246
S45	S45Y	Kidney	151	12	0.079470199	0.034798535	2.283722551	0.006819
S45	S45Y	Pituitary	4	1	0.25	0.034798535	7.184210526	0.132096
S45	S45Y	Ovary	5	1	0.2	0.034798535	5.747368421	0.162297
S45	S45Y	Adrenal gland	38	2	0.052631579	0.034798535	1.512465374	0.383072
S45	S45Y	Endometrium	26	1	0.038461539	0.034798535	1.105263158	0.601831
S45	S45Y	Soft tissue	510	1	0.001960784	0.034798535	0.056346749	1
S45	S45del	Kidney	151	54	0.357615894	0.07967033	4.488696049	0
S45	S45del	Large intestine	107	17	0.158878505	0.07967033	1.994199162	0.004741
S45	S45del	Lung	5	2	0.4	0.07967033	5.020689655	0.053951
S45	S45del	Pancreas	5	2	0.4	0.07967033	5.020689655	0.053951
S45	S45del	Bone	3	1	0.333333333	0.07967033	4.183908046	0.220475
S45	S45del	Prostate	3	1	0.333333333	0.07967033	4.183908046	0.220475
S45	S45del	Skin	24	3	0.125	0.07967033	1.568965517	0.298375
S45	S45del	Endometrium	26	2	0.076923077	0.07967033	0.965517241	0.624585
S45	S45del	Adrenal gland	87	4	0.045977012	0.07967033	0.577090765	0.923154
S45	S45del	Liver	164	1	0.006097561	0.07967033	0.076534903	0.999999

T41	T41A	Soft tissue	777	770	0.990990991	0.8969163	1.104886812	0
T41	T41A	Breast	4	4	1	0.8969163	1.114931238	0.647154
T41	T41A	CNS	2	2	1	0.8969163	1.114931238	0.804459
T41	T41A	Lung	2	2	1	0.8969163	1.114931238	0.804459
T41	T41A	Small intestine	2	2	1	0.8969163	1.114931238	0.804459
T41	T41A	Urinary tract	1	1	1	0.8969163	1.114931238	0.896916
T41	T41A	Biliary tract	10	8	0.8	0.8969163	0.89194499	0.9244
T41	T41A	Large intestine	64	54	0.84375	0.8969163	0.940723232	0.938346
T41	T41A	Kidney	20	16	0.8	0.8969163	0.89194499	0.951726
T41	T41A	Stomach	3	2	0.666666667	0.8969163	0.743287492	0.970312
T41	T41A	Prostate	6	4	0.666666667	0.8969163	0.743287492	0.982766
T41	T41A	Ovary	9	6	0.666666667	0.8969163	0.743287492	0.990716
T41	T41A	Liver	134	111	0.828358209	0.8969163	0.923562443	0.994824
T41	T41A	Bone	4	2	0.5	0.8969163	0.557465619	0.995957
T41	T41A	Haematopoietic and lymphoid	5	2	0.4	0.8969163	0.445972495	0.999482
T41	T41A	Pancreas	7	2	0.285714286	0.8969163	0.318551782	0.999992
T41	T41A	Skin	16	5	0.3125	0.8969163	0.348416012	1
T41	T41A	Endometrium	36	15	0.416666667	0.8969163	0.464554682	1
T41	T41A	Adrenal gland	1018	9	0.008840864	0.8969163	0.009856956	1
T41	T41A	Pituitary	21	1	0.047619048	0.8969163	0.053091964	1
T41	T41I	Pituitary	21	19	0.904761905	0.086343612	10.47862002	0
T41	T41I	Endometrium	36	18	0.5	0.086343612	5.790816327	1.39E-10
T41	T41I	Skin	16	9	0.5625	0.086343612	6.514668367	1.73E-06
T41	T41I	Pancreas	7	5	0.714285714	0.086343612	8.272594752	8.68E-05
T41	T41I	Liver	134	19	0.141791045	0.086343612	1.642171794	0.021962
T41	T41I	Ovary	9	3	0.333333333	0.086343612	3.860544218	0.036361

T41	T41I	Bone	4	2	0.5	0.086343612	5.790816327	0.039748
T41	T41I	Prostate	6	2	0.333333333	0.086343612	3.860544218	0.088468
T41	T41I	Large intestine	64	9	0.140625	0.086343612	1.628667092	0.098186
T41	T41I	Thyroid	2	1	0.5	0.086343612	5.790816327	0.165232
T41	T41I	Biliary tract	10	2	0.2	0.086343612	2.316326531	0.211583
T41	T41I	Stomach	3	1	0.333333333	0.086343612	3.860544218	0.237309
T41	T41I	Haematopoietic and lymphoid	5	1	0.2	0.086343612	2.316326531	0.36333
T41	T41I	Soft tissue	777	7	0.009009009	0.086343612	0.104339033	1
T41	T41N	Skin	16	2	0.125	0.006167401	20.26785714	0.004309
T41	T41N	Haematopoietic and lymphoid	5	1	0.2	0.006167401	32.42857143	0.030459
T41	T41N	Liver	134	3	0.02238806	0.006167401	3.630063966	0.050755
T41	T41N	Kidney	20	1	0.05	0.006167401	8.107142857	0.116382
T41	T41P	Kidney	20	3	0.15	0.005286344	28.375	0.000157
T41	T41P	Salivary gland	1	1	1	0.005286344	189.1666667	0.005286
T41	T41P	Large intestine	64	1	0.015625	0.005286344	2.955729167	0.287677
T41	T41P	Liver	134	1	0.007462687	0.005286344	1.411691542	0.508478
T41	T41S	Endometrium	36	3	0.083333333	0.005286344	15.76388889	0.000926
T41	T41S	Thyroid	2	1	0.5	0.005286344	94.58333333	0.010545
T41	T41S	Haematopoietic and lymphoid	5	1	0.2	0.005286344	37.83333333	0.026154
T41	T41S	Pituitary	21	1	0.047619048	0.005286344	9.007936508	0.105337

***Statistically significant scores (p<0.05) are highlighted**

Appendix 2A Multiple comparison (Tukey's test) of one way ANOVA for luciferase assay on multiplex clones

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value
D32H vs. D32V	-159.9	-349.6 to 29.82	No	ns	0.2179
D32H vs. D32Y	-154.1	-343.8 to 35.61	No	ns	0.2764
D32H vs. D32N	-25.78	-215.5 to 163.9	No	ns	>0.9999
D32H vs. D32G	-135.4	-325.1 to 54.36	No	ns	0.5219
D32H vs. S33Y	94.25	-95.47 to 284	No	ns	0.9658
D32H vs. S33P	90.34	-99.38 to 280.1	No	ns	0.9785
D32H vs. S33C	112.2	-77.53 to 301.9	No	ns	0.8367
D32H vs. S33F	107.6	-82.08 to 297.4	No	ns	0.8815
D32H vs. S33L	165.3	-24.46 to 355	No	ns	0.1722
D32H vs. G34V	5.638	-184.1 to 195.4	No	ns	>0.9999
D32H vs. G34E	43.45	-146.3 to 233.2	No	ns	>0.9999
D32H vs. G34R	-10.71	-200.4 to 179	No	ns	>0.9999
D32H vs. S37C	-319.8	-531.9 to -107.7	Yes	****	<0.0001
D32H vs. S37F	-242.2	-431.9 to -52.44	Yes	**	0.0019
D32H vs. S37A	-241.7	-453.8 to -29.55	Yes	*	0.0104
D32H vs. S37Y	-461.1	-673.2 to -249	Yes	****	<0.0001
D32H vs. S45Y	207.4	17.69 to 397.1	Yes	*	0.0179
D32H vs. S45P	183.5	-6.176 to 373.3	No	ns	0.0699
D32H vs. S45F	145	-44.74 to 334.7	No	ns	0.3868
D32H vs. S45C	233.5	21.4 to 445.6	Yes	*	0.0163
D32H vs. T41I	-446.1	-635.8 to -256.3	Yes	****	<0.0001
D32H vs. T41A	-309	-498.7 to -119.3	Yes	****	<0.0001
D32H vs. T41S	254.4	64.69 to 444.1	Yes	***	0.0008
D32H vs. T41N	156.1	-33.66 to 345.8	No	ns	0.2556
D32H vs. T41P	80.21	-109.5 to 269.9	No	ns	0.9951

D32H vs. E14	254.9	90.63 to 419.2	Yes	****	<0.0001	
D32V vs. D32Y	5.795	-183.9 to 195.5	No	ns	>0.9999	
D32V vs. D32N	134.1	-55.59 to 323.8	No	ns		0.54
D32V vs. D32G	24.55	-165.2 to 214.3	No	ns	>0.9999	
D32V vs. S33Y	254.2	64.43 to 443.9	Yes	***		0.0008
D32V vs. S33P	250.2	60.53 to 440	Yes	**		0.0011
D32V vs. S33C	272.1	82.37 to 461.8	Yes	***		0.0002
D32V vs. S33F	267.5	77.83 to 457.3	Yes	***		0.0003
D32V vs. S33L	325.2	135.4 to 514.9	Yes	****	<0.0001	
D32V vs. G34V	165.5	-24.18 to 355.3	No	ns		0.1701
D32V vs. G34E	203.4	13.64 to 393.1	Yes	*		0.0228
D32V vs. G34R	149.2	-40.53 to 338.9	No	ns		0.3332
D32V vs. S37C	-159.9	-372 to 52.22	No	ns		0.4132
D32V vs. S37F	-82.26	-272 to 107.5	No	ns		0.9931
D32V vs. S37A	-81.76	-293.9 to 130.4	No	ns		0.9987
D32V vs. S37Y	-301.2	-513.3 to -89.06	Yes	***		0.0003
D32V vs. S45Y	367.3	177.6 to 557	Yes	****	<0.0001	
D32V vs. S45P	343.4	153.7 to 533.2	Yes	****	<0.0001	
D32V vs. S45F	304.9	115.2 to 494.6	Yes	****	<0.0001	
D32V vs. S45C	393.4	181.3 to 605.5	Yes	****	<0.0001	
D32V vs. T41I	-286.2	-475.9 to -96.44	Yes	****	<0.0001	
D32V vs. T41A	-149.1	-338.8 to 40.65	No	ns		0.3347
D32V vs. T41S	414.3	224.6 to 604	Yes	****	<0.0001	
D32V vs. T41N	316	126.2 to 505.7	Yes	****	<0.0001	
D32V vs. T41P	240.1	50.4 to 429.8	Yes	**		0.0022
D32V vs. E14	414.8	250.5 to 579.1	Yes	****	<0.0001	
D32Y vs. D32N	128.3	-61.39 to 318.1	No	ns		0.6253

D32Y vs. D32G	18.75	-171 to 208.5	No	ns	>0.9999	
D32Y vs. S33Y	248.4	58.64 to 438.1	Yes	**		0.0012
D32Y vs. S33P	244.5	54.73 to 434.2	Yes	**		0.0016
D32Y vs. S33C	266.3	76.58 to 456	Yes	***		0.0003
D32Y vs. S33F	261.8	72.03 to 451.5	Yes	***		0.0005
D32Y vs. S33L	319.4	129.7 to 509.1	Yes	****	<0.0001	
D32Y vs. G34V	159.7	-29.97 to 349.5	No	ns		0.2194
D32Y vs. G34E	197.6	7.843 to 387.3	Yes	*		0.0321
D32Y vs. G34R	143.4	-46.32 to 333.1	No	ns		0.408
D32Y vs. S37C	-165.7	-377.8 to 46.43	No	ns		0.3456
D32Y vs. S37F	-88.06	-277.8 to 101.7	No	ns		0.984
D32Y vs. S37A	-87.55	-299.7 to 124.6	No	ns		0.9964
D32Y vs. S37Y	-307	-519.1 to -94.86	Yes	***		0.0002
D32Y vs. S45Y	361.5	171.8 to 551.2	Yes	****	<0.0001	
D32Y vs. S45P	337.7	147.9 to 527.4	Yes	****	<0.0001	
D32Y vs. S45F	299.1	109.4 to 488.8	Yes	****	<0.0001	
D32Y vs. S45C	387.6	175.5 to 599.7	Yes	****	<0.0001	
D32Y vs. T41I	-292	-481.7 to -102.2	Yes	****	<0.0001	
D32Y vs. T41A	-154.9	-344.6 to 34.85	No	ns		0.2682
D32Y vs. T41S	408.5	218.8 to 598.2	Yes	****	<0.0001	
D32Y vs. T41N	310.2	120.5 to 499.9	Yes	****	<0.0001	
D32Y vs. T41P	234.3	44.6 to 424	Yes	**		0.0032
D32Y vs. E14	409	244.7 to 573.3	Yes	****	<0.0001	
D32N vs. D32G	-109.6	-299.3 to 80.14	No	ns		0.8634
D32N vs. S33Y	120	-69.69 to 309.7	No	ns		0.7421
D32N vs. S33P	116.1	-73.6 to 305.8	No	ns		0.7917
D32N vs. S33C	138	-51.75 to 327.7	No	ns		0.484

D32N vs. S33F	133.4	-56.3 to 323.1	No	ns		0.5504
D32N vs. S33L	191	1.321 to 380.8	Yes	*		0.0465
D32N vs. G34V	31.42	-158.3 to 221.1	No	ns	>0.9999	
D32N vs. G34E	69.23	-120.5 to 259	No	ns		0.9994
D32N vs. G34R	15.06	-174.7 to 204.8	No	ns	>0.9999	
D32N vs. S37C	-294	-506.1 to -81.9	Yes	***		0.0004
D32N vs. S37F	-216.4	-406.1 to -26.67	Yes	*		0.0102
D32N vs. S37A	-215.9	-428 to -3.77	Yes	*		0.0414
D32N vs. S37Y	-435.3	-647.4 to -223.2	Yes	****	<0.0001	
D32N vs. S45Y	233.2	43.47 to 422.9	Yes	**		0.0034
D32N vs. S45P	209.3	19.6 to 399	Yes	*		0.0159
D32N vs. S45F	170.8	-18.96 to 360.5	No	ns		0.1334
D32N vs. S45C	259.3	47.18 to 471.4	Yes	**		0.0037
D32N vs. T41I	-420.3	-610 to -230.6	Yes	****	<0.0001	
D32N vs. T41A	-283.2	-472.9 to -93.48	Yes	***		0.0001
D32N vs. T41S	280.2	90.47 to 469.9	Yes	***		0.0001
D32N vs. T41N	181.8	-7.879 to 371.6	No	ns		0.0765
D32N vs. T41P	106	-83.73 to 295.7	No	ns		0.8957
D32N vs. E14	280.7	116.4 to 445	Yes	****	<0.0001	
D32G vs. S33Y	229.6	39.89 to 419.3	Yes	**		0.0044
D32G vs. S33P	225.7	35.98 to 415.4	Yes	**		0.0056
D32G vs. S33C	247.5	57.83 to 437.3	Yes	**		0.0013
D32G vs. S33F	243	53.28 to 432.7	Yes	**		0.0018
D32G vs. S33L	300.6	110.9 to 490.3	Yes	****	<0.0001	
D32G vs. G34V	141	-48.72 to 330.7	No	ns		0.441
D32G vs. G34E	178.8	-10.91 to 368.5	No	ns		0.0895
D32G vs. G34R	124.6	-65.07 to 314.4	No	ns		0.6786

D32G vs. S37C	-184.4	-396.5 to 27.68	No	ns		0.1746
D32G vs. S37F	-106.8	-296.5 to 82.91	No	ns		0.8888
D32G vs. S37A	-106.3	-318.4 to 105.8	No	ns		0.9625
D32G vs. S37Y	-325.7	-537.8 to -113.6	Yes	****	<0.0001	
D32G vs. S45Y	342.8	153 to 532.5	Yes	****	<0.0001	
D32G vs. S45P	318.9	129.2 to 508.6	Yes	****	<0.0001	
D32G vs. S45F	280.3	90.62 to 470.1	Yes	***		0.0001
D32G vs. S45C	368.9	156.8 to 581	Yes	****	<0.0001	
D32G vs. T41I	-310.7	-500.4 to -121	Yes	****	<0.0001	
D32G vs. T41A	-173.6	-363.3 to 16.1	No	ns		0.1161
D32G vs. T41S	389.8	200.1 to 579.5	Yes	****	<0.0001	
D32G vs. T41N	291.4	101.7 to 481.1	Yes	****	<0.0001	
D32G vs. T41P	215.6	25.85 to 405.3	Yes	*		0.0108
D32G vs. E14	390.3	226 to 554.6	Yes	****	<0.0001	
S33Y vs. S33P	-3.906	-193.6 to 185.8	No	ns	>0.9999	
S33Y vs. S33C	17.94	-171.8 to 207.7	No	ns	>0.9999	
S33Y vs. S33F	13.39	-176.3 to 203.1	No	ns	>0.9999	
S33Y vs. S33L	71.01	-118.7 to 260.7	No	ns		0.9991
S33Y vs. G34V	-88.61	-278.3 to 101.1	No	ns		0.9828
S33Y vs. G34E	-50.8	-240.5 to 138.9	No	ns	>0.9999	
S33Y vs. G34R	-105	-294.7 to 84.76	No	ns		0.904
S33Y vs. S37C	-414	-626.2 to -201.9	Yes	****	<0.0001	
S33Y vs. S37F	-336.4	-526.1 to -146.7	Yes	****	<0.0001	
S33Y vs. S37A	-335.9	-548 to -123.8	Yes	****	<0.0001	
S33Y vs. S37Y	-555.3	-767.4 to -343.2	Yes	****	<0.0001	
S33Y vs. S45Y	113.2	-76.56 to 302.9	No	ns		0.8261
S33Y vs. S45P	89.29	-100.4 to 279	No	ns		0.9812

S33Y vs. S45F	50.73	-139 to 240.5	No	ns	>0.9999	
S33Y vs. S45C	139.3	-72.85 to 351.4	No	ns		0.6797
S33Y vs. T41I	-540.3	-730 to -350.6	Yes	****	<0.0001	
S33Y vs. T41A	-403.2	-592.9 to -213.5	Yes	****	<0.0001	
S33Y vs. T41S	160.2	-29.56 to 349.9	No	ns		0.2155
S33Y vs. T41N	61.81	-127.9 to 251.5	No	ns	>0.9999	
S33Y vs. T41P	-14.04	-203.8 to 175.7	No	ns	>0.9999	
S33Y vs. E14	160.7	-3.625 to 325	No	ns		0.0628
S33P vs. S33C	21.84	-167.9 to 211.6	No	ns	>0.9999	
S33P vs. S33F	17.3	-172.4 to 207	No	ns	>0.9999	
S33P vs. S33L	74.92	-114.8 to 264.6	No	ns		0.9981
S33P vs. G34V	-84.71	-274.4 to 105	No	ns		0.99
S33P vs. G34E	-46.89	-236.6 to 142.8	No	ns	>0.9999	
S33P vs. G34R	-101.1	-290.8 to 88.66	No	ns		0.9316
S33P vs. S37C	-410.1	-622.3 to -198	Yes	****	<0.0001	
S33P vs. S37F	-332.5	-522.2 to -142.8	Yes	****	<0.0001	
S33P vs. S37A	-332	-544.1 to -119.9	Yes	****	<0.0001	
S33P vs. S37Y	-551.4	-763.5 to -339.3	Yes	****	<0.0001	
S33P vs. S45Y	117.1	-72.65 to 306.8	No	ns		0.7801
S33P vs. S45P	93.2	-96.52 to 282.9	No	ns		0.9696
S33P vs. S45F	54.64	-135.1 to 244.4	No	ns	>0.9999	
S33P vs. S45C	143.2	-68.94 to 355.3	No	ns		0.6293
S33P vs. T41I	-536.4	-726.1 to -346.7	Yes	****	<0.0001	
S33P vs. T41A	-399.3	-589 to -209.6	Yes	****	<0.0001	
S33P vs. T41S	164.1	-25.65 to 353.8	No	ns		0.1817
S33P vs. T41N	65.72	-124 to 255.4	No	ns		0.9998
S33P vs. T41P	-10.13	-199.9 to 179.6	No	ns	>0.9999	

S33P vs. E14	164.6	0.2815 to 328.9	Yes	*		0.0491
S33C vs. S33F	-4.546	-194.3 to 185.2	No	ns	>0.9999	
S33C vs. S33L	53.08	-136.6 to 242.8	No	ns	>0.9999	
S33C vs. G34V	-106.5	-296.3 to 83.17	No	ns		0.891
S33C vs. G34E	-68.73	-258.5 to 121	No	ns		0.9995
S33C vs. G34R	-122.9	-312.6 to 66.82	No	ns		0.7031
S33C vs. S37C	-432	-644.1 to -219.9	Yes	****	<0.0001	
S33C vs. S37F	-354.4	-544.1 to -164.6	Yes	****	<0.0001	
S33C vs. S37A	-353.8	-566 to -141.7	Yes	****	<0.0001	
S33C vs. S37Y	-573.3	-785.4 to -361.2	Yes	****	<0.0001	
S33C vs. S45Y	95.22	-94.5 to 284.9	No	ns		0.9619
S33C vs. S45P	71.36	-118.4 to 261.1	No	ns		0.9991
S33C vs. S45F	32.79	-156.9 to 222.5	No	ns	>0.9999	
S33C vs. S45C	121.3	-90.78 to 333.4	No	ns		0.8735
S33C vs. T41I	-558.3	-748 to -368.5	Yes	****	<0.0001	
S33C vs. T41A	-421.2	-610.9 to -231.4	Yes	****	<0.0001	
S33C vs. T41S	142.2	-47.49 to 331.9	No	ns		0.424
S33C vs. T41N	43.88	-145.8 to 233.6	No	ns	>0.9999	
S33C vs. T41P	-31.98	-221.7 to 157.7	No	ns	>0.9999	
S33C vs. E14	142.7	-21.56 to 307	No	ns		0.1757
S33F vs. S33L	57.62	-132.1 to 247.3	No	ns	>0.9999	
S33F vs. G34V	-102	-291.7 to 87.72	No	ns		0.9255
S33F vs. G34E	-64.19	-253.9 to 125.5	No	ns		0.9998
S33F vs. G34R	-118.4	-308.1 to 71.36	No	ns		0.7638
S33F vs. S37C	-427.4	-639.5 to -215.3	Yes	****	<0.0001	
S33F vs. S37F	-349.8	-539.5 to -160.1	Yes	****	<0.0001	
S33F vs. S37A	-349.3	-561.4 to -137.2	Yes	****	<0.0001	

S33F vs. S37Y	-568.7	-780.8 to -356.6	Yes	****	<0.0001	
S33F vs. S45Y	99.77	-89.95 to 289.5	No	ns		0.9394
S33F vs. S45P	75.9	-113.8 to 265.6	No	ns		0.9977
S33F vs. S45F	37.34	-152.4 to 227.1	No	ns	>0.9999	
S33F vs. S45C	125.9	-86.24 to 338	No	ns		0.8324
S33F vs. T41I	-553.7	-743.4 to -364	Yes	****	<0.0001	
S33F vs. T41A	-416.6	-606.3 to -226.9	Yes	****	<0.0001	
S33F vs. T41S	146.8	-42.95 to 336.5	No	ns		0.3635
S33F vs. T41N	48.42	-141.3 to 238.1	No	ns	>0.9999	
S33F vs. T41P	-27.43	-217.1 to 162.3	No	ns	>0.9999	
S33F vs. E14	147.3	-17.02 to 311.6	No	ns		0.1378
S33L vs. G34V	-159.6	-349.3 to 30.1	No	ns		0.2205
S33L vs. G34E	-121.8	-311.5 to 67.91	No	ns		0.7181
S33L vs. G34R	-176	-365.7 to 13.74	No	ns		0.1033
S33L vs. S37C	-485.1	-697.2 to -272.9	Yes	****	<0.0001	
S33L vs. S37F	-407.4	-597.1 to -217.7	Yes	****	<0.0001	
S33L vs. S37A	-406.9	-619 to -194.8	Yes	****	<0.0001	
S33L vs. S37Y	-626.3	-838.5 to -414.2	Yes	****	<0.0001	
S33L vs. S45Y	42.15	-147.6 to 231.9	No	ns	>0.9999	
S33L vs. S45P	18.28	-171.4 to 208	No	ns	>0.9999	
S33L vs. S45F	-20.28	-210 to 169.4	No	ns	>0.9999	
S33L vs. S45C	68.25	-143.9 to 280.4	No	ns	>0.9999	
S33L vs. T41I	-611.3	-801 to -421.6	Yes	****	<0.0001	
S33L vs. T41A	-474.2	-664 to -284.5	Yes	****	<0.0001	
S33L vs. T41S	89.15	-100.6 to 278.9	No	ns		0.9815
S33L vs. T41N	-9.2	-198.9 to 180.5	No	ns	>0.9999	
S33L vs. T41P	-85.05	-274.8 to 104.7	No	ns		0.9895

S33L vs. E14	89.66	-74.64 to 254	No	ns		0.9148
G34V vs. G34E	37.82	-151.9 to 227.5	No	ns	>0.9999	
G34V vs. G34R	-16.35	-206.1 to 173.4	No	ns	>0.9999	
G34V vs. S37C	-325.4	-537.5 to -113.3	Yes	****	<0.0001	
G34V vs. S37F	-247.8	-437.5 to -58.08	Yes	**		0.0013
G34V vs. S37A	-247.3	-459.4 to -35.19	Yes	**		0.0075
G34V vs. S37Y	-466.7	-678.8 to -254.6	Yes	****	<0.0001	
G34V vs. S45Y	201.8	12.05 to 391.5	Yes	*		0.0251
G34V vs. S45P	177.9	-11.81 to 367.6	No	ns		0.0937
G34V vs. S45F	139.3	-50.38 to 329.1	No	ns		0.4643
G34V vs. S45C	227.9	15.77 to 440	Yes	*		0.0222
G34V vs. T41I	-451.7	-641.4 to -262	Yes	****	<0.0001	
G34V vs. T41A	-314.6	-504.3 to -124.9	Yes	****	<0.0001	
G34V vs. T41S	248.8	59.06 to 438.5	Yes	**		0.0012
G34V vs. T41N	150.4	-39.29 to 340.1	No	ns		0.3184
G34V vs. T41P	74.57	-115.1 to 264.3	No	ns		0.9982
G34V vs. E14	249.3	84.99 to 413.6	Yes	****	<0.0001	
G34E vs. G34R	-54.17	-243.9 to 135.6	No	ns	>0.9999	
G34E vs. S37C	-363.2	-575.4 to -151.1	Yes	****	<0.0001	
G34E vs. S37F	-285.6	-475.3 to -95.9	Yes	****	<0.0001	
G34E vs. S37A	-285.1	-497.2 to -73	Yes	***		0.0008
G34E vs. S37Y	-504.5	-716.6 to -292.4	Yes	****	<0.0001	
G34E vs. S45Y	164	-25.76 to 353.7	No	ns		0.1826
G34E vs. S45P	140.1	-49.63 to 329.8	No	ns		0.4537
G34E vs. S45F	101.5	-88.19 to 291.2	No	ns		0.9286
G34E vs. S45C	190.1	-22.05 to 402.2	No	ns		0.1383
G34E vs. T41I	-489.5	-679.2 to -299.8	Yes	****	<0.0001	

G34E vs. T41A	-352.4	-542.1 to -162.7	Yes	****	<0.0001	
G34E vs. T41S	211	21.24 to 400.7	Yes	*		0.0144
G34E vs. T41N	112.6	-77.11 to 302.3	No	ns		0.8321
G34E vs. T41P	36.76	-153 to 226.5	No	ns	>0.9999	
G34E vs. E14	211.5	47.17 to 375.8	Yes	**		0.0016
G34R vs. S37C	-309.1	-521.2 to -96.97	Yes	***		0.0002
G34R vs. S37F	-231.5	-421.2 to -41.73	Yes	**		0.0039
G34R vs. S37A	-230.9	-443.1 to -18.83	Yes	*		0.0188
G34R vs. S37Y	-450.4	-662.5 to -238.3	Yes	****	<0.0001	
G34R vs. S45Y	218.1	28.4 to 407.8	Yes	**		0.0092
G34R vs. S45P	194.3	4.537 to 384	Yes	*		0.0388
G34R vs. S45F	155.7	-34.02 to 345.4	No	ns		0.2594
G34R vs. S45C	244.2	32.12 to 456.3	Yes	**		0.009
G34R vs. T41I	-435.4	-625.1 to -245.6	Yes	****	<0.0001	
G34R vs. T41A	-298.3	-488 to -108.5	Yes	****	<0.0001	
G34R vs. T41S	265.1	75.41 to 454.8	Yes	***		0.0004
G34R vs. T41N	166.8	-22.94 to 356.5	No	ns		0.1607
G34R vs. T41P	90.93	-98.79 to 280.6	No	ns		0.9769
G34R vs. E14	265.6	101.3 to 429.9	Yes	****	<0.0001	
S37C vs. S37F	77.63	-134.5 to 289.7	No	ns		0.9994
S37C vs. S37A	78.13	-154.2 to 310.5	No	ns		0.9998
S37C vs. S37Y	-141.3	-373.6 to 91.07	No	ns		0.8008
S37C vs. S45Y	527.2	315.1 to 739.3	Yes	****	<0.0001	
S37C vs. S45P	503.3	291.2 to 715.4	Yes	****	<0.0001	
S37C vs. S45F	464.8	252.7 to 676.9	Yes	****	<0.0001	
S37C vs. S45C	553.3	321 to 785.7	Yes	****	<0.0001	
S37C vs. T41I	-126.3	-338.4 to 85.84	No	ns		0.8285

S37C vs. T41A	10.82	-201.3 to 222.9	No	ns	>0.9999	
S37C vs. T41S	574.2	362.1 to 786.3	Yes	****	<0.0001	
S37C vs. T41N	475.9	263.7 to 688	Yes	****	<0.0001	
S37C vs. T41P	400	187.9 to 612.1	Yes	****	<0.0001	
S37C vs. E14	574.7	385 to 764.4	Yes	****	<0.0001	
S37F vs. S37A	0.5038	-211.6 to 212.6	No	ns	>0.9999	
S37F vs. S37Y	-218.9	-431 to -6.804	Yes	*		0.0355
S37F vs. S45Y	449.6	259.9 to 639.3	Yes	****	<0.0001	
S37F vs. S45P	425.7	236 to 615.4	Yes	****	<0.0001	
S37F vs. S45F	387.1	197.4 to 576.9	Yes	****	<0.0001	
S37F vs. S45C	475.7	263.6 to 687.8	Yes	****	<0.0001	
S37F vs. T41I	-203.9	-393.6 to -14.18	Yes	*		0.0221
S37F vs. T41A	-66.81	-256.5 to 122.9	No	ns		0.9997
S37F vs. T41S	496.6	306.9 to 686.3	Yes	****	<0.0001	
S37F vs. T41N	398.2	208.5 to 587.9	Yes	****	<0.0001	
S37F vs. T41P	322.4	132.7 to 512.1	Yes	****	<0.0001	
S37F vs. E14	497.1	332.8 to 661.4	Yes	****	<0.0001	
S37A vs. S37Y	-219.4	-451.8 to 12.94	No	ns		0.0879
S37A vs. S45Y	449.1	237 to 661.2	Yes	****	<0.0001	
S37A vs. S45P	425.2	213.1 to 637.3	Yes	****	<0.0001	
S37A vs. S45F	386.6	174.5 to 598.8	Yes	****	<0.0001	
S37A vs. S45C	475.2	242.8 to 707.5	Yes	****	<0.0001	
S37A vs. T41I	-204.4	-416.5 to 7.706	No	ns		0.0726
S37A vs. T41A	-67.32	-279.4 to 144.8	No	ns	>0.9999	
S37A vs. T41S	496.1	284 to 708.2	Yes	****	<0.0001	
S37A vs. T41N	397.7	185.6 to 609.8	Yes	****	<0.0001	
S37A vs. T41P	321.9	109.8 to 534	Yes	****	<0.0001	

S37A vs. E14	496.6	306.9 to 686.3	Yes	****	<0.0001	
S37Y vs. S45Y	668.5	456.4 to 880.6	Yes	****	<0.0001	
S37Y vs. S45P	644.6	432.5 to 856.7	Yes	****	<0.0001	
S37Y vs. S45F	606.1	393.9 to 818.2	Yes	****	<0.0001	
S37Y vs. S45C	694.6	462.2 to 927	Yes	****	<0.0001	
S37Y vs. T41I	15.01	-197.1 to 227.1	No	ns	>0.9999	
S37Y vs. T41A	152.1	-60.01 to 364.2	No	ns		0.5118
S37Y vs. T41S	715.5	503.4 to 927.6	Yes	****	<0.0001	
S37Y vs. T41N	617.1	405 to 829.3	Yes	****	<0.0001	
S37Y vs. T41P	541.3	329.2 to 753.4	Yes	****	<0.0001	
S37Y vs. E14	716	526.3 to 905.7	Yes	****	<0.0001	
S45Y vs. S45P	-23.87	-213.6 to 165.9	No	ns	>0.9999	
S45Y vs. S45F	-62.43	-252.1 to 127.3	No	ns		0.9999
S45Y vs. S45C	26.11	-186 to 238.2	No	ns	>0.9999	
S45Y vs. T41I	-653.5	-843.2 to -463.8	Yes	****	<0.0001	
S45Y vs. T41A	-516.4	-706.1 to -326.7	Yes	****	<0.0001	
S45Y vs. T41S	47	-142.7 to 236.7	No	ns	>0.9999	
S45Y vs. T41N	-51.35	-241.1 to 138.4	No	ns	>0.9999	
S45Y vs. T41P	-127.2	-316.9 to 62.52	No	ns		0.6418
S45Y vs. E14	47.52	-116.8 to 211.8	No	ns	>0.9999	
S45P vs. S45F	-38.56	-228.3 to 151.2	No	ns	>0.9999	
S45P vs. S45C	49.97	-162.1 to 262.1	No	ns	>0.9999	
S45P vs. T41I	-629.6	-819.3 to -439.9	Yes	****	<0.0001	
S45P vs. T41A	-492.5	-682.2 to -302.8	Yes	****	<0.0001	
S45P vs. T41S	70.87	-118.8 to 260.6	No	ns		0.9992
S45P vs. T41N	-27.48	-217.2 to 162.2	No	ns	>0.9999	
S45P vs. T41P	-103.3	-293.1 to 86.39	No	ns		0.9163

S45P vs. E14	71.38	-92.92 to 235.7	No	ns		0.9929
S45F vs. S45C	88.54	-123.6 to 300.6	No	ns		0.9958
S45F vs. T41I	-591	-780.8 to -401.3	Yes	****	<0.0001	
S45F vs. T41A	-454	-643.7 to -264.2	Yes	****	<0.0001	
S45F vs. T41S	109.4	-80.29 to 299.2	No	ns		0.8648
S45F vs. T41N	11.08	-178.6 to 200.8	No	ns	>0.9999	
S45F vs. T41P	-64.77	-254.5 to 125	No	ns		0.9998
S45F vs. E14	109.9	-54.36 to 274.2	No	ns		0.6454
S45C vs. T41I	-679.6	-891.7 to -467.5	Yes	****	<0.0001	
S45C vs. T41A	-542.5	-754.6 to -330.4	Yes	****	<0.0001	
S45C vs. T41S	20.9	-191.2 to 233	No	ns	>0.9999	
S45C vs. T41N	-77.45	-289.6 to 134.7	No	ns		0.9994
S45C vs. T41P	-153.3	-365.4 to 58.81	No	ns		0.4962
S45C vs. E14	21.41	-168.3 to 211.1	No	ns	>0.9999	
T41I vs. T41A	137.1	-52.63 to 326.8	No	ns		0.4966
T41I vs. T41S	700.5	510.8 to 890.2	Yes	****	<0.0001	
T41I vs. T41N	602.1	412.4 to 791.8	Yes	****	<0.0001	
T41I vs. T41P	526.3	336.6 to 716	Yes	****	<0.0001	
T41I vs. E14	701	536.7 to 865.3	Yes	****	<0.0001	
T41A vs. T41S	563.4	373.7 to 753.1	Yes	****	<0.0001	
T41A vs. T41N	465	275.3 to 654.8	Yes	****	<0.0001	
T41A vs. T41P	389.2	199.5 to 578.9	Yes	****	<0.0001	
T41A vs. E14	563.9	399.6 to 728.2	Yes	****	<0.0001	
T41S vs. T41N	-98.35	-288.1 to 91.37	No	ns		0.9472
T41S vs. T41P	-174.2	-363.9 to 15.52	No	ns		0.1128
T41S vs. E14	0.5144	-163.8 to 164.8	No	ns	>0.9999	
T41N vs. T41P	-75.85	-265.6 to 113.9	No	ns		0.9977

T41N vs. E14	98.86	-65.44 to 263.2	No	ns	0.8148
T41P vs. E14	174.7	10.41 to 339	Yes	*	0.0251

References

Aberle, H., Bauer, A., Stappert, J., Kispert, A. and Kemler, R. (1997) 'Beta-Catenin Is a Target for the Ubiquitin Proteasome Pathway', *The EMBO Journal*, 16(13), pp. 3797–3804.

Afouda, B. A., Martin, J., Liu, F., Ciau-Uitz, A., Patient, R. and Hoppler, S. (2008) 'GATA transcription factors integrate Wnt signalling during heart development', *Development*, 135(19), pp. 3185–3190.

Aguilar, F., Hussain, S. P. and Cerutti, P. (1993) 'Aflatoxin B1 induces the transversion of G -- T in codon 249 of the p53 tumor suppressor gene in human hepatocytes', *Proceedings of the National Academy of Sciences*, 90(2), pp. 8586–8590.

Albuquerque, C., Breukel, C., van der Luijt, R., Fidalgo, P., Lage, P., Slors, F. J. M., Leitão, C. N., Fodde, R. and Smits, R. (2002) 'The "just-right" signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade.', *Human molecular genetics*, 11(13), pp. 1549–1560.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A. L. and Boyault, S., (2013) 'Signatures of mutational processes in human cancer', *Nature*, 500(7463), pp. 415–421.

Anderson, E. M., Haupt, A., Schiel, J. a, Chou, E., Machado, H. B., Strezoska, Z., Lenger, S., McClelland, S., Birmingham, A., Vermeulen, A. and Smith, A. V. B. (2015) 'Systematic Analysis of CRISPR-Cas9 Mismatch Tolerance Reveals Low Levels of Off-Target Activity.', *Journal of biotechnology*, 211, pp. 56–65.

Arnold, S. J., Stappert, J., Bauer, A., Kispert, A., Herrmann, B. G. and Kemler, R. (2000) 'Brachyury is a target gene of the Wnt/β-catenin signaling pathway', *Mechanisms of Development*, 91(1–2), pp. 249–258.

Audard, V., Grimber, G., Elie, C., Radenen, B., Audebourg, A., Letourneur, F. and Soubrane, O. (2007) 'Cholestasis is a marker for hepatocellular carcinomas displaying β - catenin mutations', *Journal of Pathology*, 212(3), pp. 345–352.

Austinat, M., Dunsch, R., Wittekind, C., Tannapfel, A., Gebhardt, R. and Gaunitz, F. (2008) 'Correlation between β -catenin mutations and expression of Wnt-signaling target genes in hepatocellular carcinoma', *Molecular Cancer*, 7(1), p. 21.

Baba, Y., Yokota, T., Spits, H., Garrett, K. P., Hayashi, S. and Kincade, P. W. (2006) 'Constitutively Active β -Catenin Promotes Expansion of Multipotent Hematopoietic Progenitors in Culture Yoshihiro', *The Journal of Immunology*, 177(4), pp. 2294–2303.

Baeissa, H., Benstead-Hume, G., Richardson, C. J. and Pearl, F. M. G. (2017) 'Identification and analysis of mutational hotspots in oncogenes and tumour suppressors.', *Oncotarget*, 8(13), pp. 21290–21304.

Bakker, E. R. M., Hoekstra, E., Franken, P. F., Helvensteijn, W., Van Deurzen, C. H. M., Van Veelen, W., Kuipers, E. J. and Smits, R. (2013) 'β-Catenin signaling dosage dictates tissue-specific tumor predisposition in Apc-driven cancer', *Oncogene*. Nature Publishing Group, 32(38), pp. 4579–4585.

Bakre, M. M., Hoi, A., Mong, J. C. Y., Koh, Y. Y., Wong, K. Y. and Stanton, L. W. (2007) 'Generation of multipotential mesendodermal progenitors from mouse embryonic stem cells via sustained Wnt pathway activation', *Journal of Biological Chemistry*, 282(43), pp. 31703–31712.

Barker, N., Hurlstone, A., Musisi, H., Miles, A., Bienz, M. and Clevers, H., (2001) 'The chromatin remodelling factor Brg-1 interacts with beta-catenin to promote target gene activation', *The EMBO Journal*, 20(17), pp. 4935–4943.

Bazan, J. F. and de Sauvage, F. J. (2009) 'Correspondence Structural Ties between Cholesterol Transport and Morphogen Signaling', *Cell*, 138(6), pp. 1055–1056.

Bertocchi, C., Rao, M. V. and Zaidel-bar, R. (2012) 'Regulation of Adherens Junction Dynamics by Phosphorylation Switches', *Journal of Signal Transduction*, 2012.

Boch, J. and Bonas, U. (2010) '*Xanthomonas* AvrBs3 Family-Type III Effectors: Discovery and Function', *Annual Review of Phytopathology*, 48(1), pp. 419–436.

Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S. D. (2005) 'Clustered regularly

interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin', *Microbiology*, 151(8), pp. 2551–2561.

Botezatu, A. (2016) 'Mechanisms of Oncogene Activation', in Bulgin, D. (ed.). Rijeka: InTech, p. Ch. 1.

Bradley, A., Evans, M., Kaufman, M. H. and Robertson, E. (1984) 'Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines', *Nature*, 309(5965), pp. 255–256.

Brash, D. E., Rudolph, J. A., Simon, J. A., Lin, A., McKenna, G. J., Baden, H. P., Halperin, A. J. and Ponten, J. (1991) 'A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma.', *Proceedings of the National Academy of Sciences*, 88(22), pp. 10124–10128.

Brault, V., Moore, R., Kutsch, S., Ishibashi, M., Rowitch, D. H., McMahon, a P., Sommer, L., Boussadia, O. and Kemler, R. (2001) 'Inactivation of the beta-catenin gene by Wnt1-Cre-mediated deletion results in dramatic brain malformation and failure of craniofacial development.', *Development (Cambridge, England)*, 128(8), pp. 1253–1264.

Braun, R. E., Lo, D., Pinkert, C. a, Widera, G., Fiavell, R. a, Palmiter, R. D. and Brinster, R. L. (1990) 'Infertility in Male Transgenic Mice : Disruption Expression in Postmeiotic of Sperm Development Germ Cells1 by HSV-tk', *Biology of Reproduction*, 43(July), pp. 684–693.

Brembeck, F. H. and W. B. and Rosario, M. (2006) 'Balancing cell adhesion and Wnt signaling , the key role of b -catenin ´ rio and Walter Birchmeier', *Current Opinion in Genetics & Development*, 16(1), pp. 51–59.

Cadigan, K. M. and Waterman, M. L. (2012) 'TCF / LEFs and Wnt Signaling in the Nucleus', *Cold Spring Harbor Perspectives in Biology*, 4, p. a007906.

Campo, E., Sanjosé, S. De, Montserrat, E., González-Dýaz, M., Jares, P., Himmelbaue, H., Bea, S. and Cancer Genome Consortium, (2010) 'International network of cancer genome projects', *Nature*, 464(7291), pp. 993–998.

Capecchi, M. R. (1989) ‘The new mouse genetics: Altering the genome by gene targeting’, *Trends in Genetics*, 5(C), pp. 70–76.

Carrera, I., Janody, F., Leeds, N., Duveau, F. and Treisman, J. E. (2008) ‘Pygopus activates Wingless target gene transcription through the mediator complex subunits Med12 and Med13.’, *Proceedings of the National Academy of Sciences of the United States of America*, 105(18), pp. 6644–9.

Çelen, I., Ross, K. E., Arighi, C. N. and Wu, C. H. (2015) ‘Bioinformatics knowledge map for analysis of beta-catenin function in cancer’, *PLoS ONE*, 10(10), pp. 1–19.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Onur, S., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P. and Sander, C. (2014) ‘The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data’, *Cancer Discovery*, 2(5), pp. 401–404..

Chan, E. F., Gat, U., Mcniff, J. M. and Fuchs, E. (1999) ‘A common human skin tumour is caused by activating mutations in β -catenin’, *Nature Genetics*, 21(4), pp. 410–413.

Chen, Y., Stewart, D. B. and Nelson, W. J. (1999) ‘Coupling Assembly of the E-Cadherin/β-Catenin Complex to Efficient Endoplasmic Reticulum Exit and Basal-lateral Membrane Targeting of E-Cadherin in Polarized MDCK Cells’, *The Journal of Cell Biology*, 144(4), pp. 687–699.

Chen, Y. T. and Bradley, A. (2000) ‘A new positive/negative selectable marker, puDeltatk, for use in embryonic stem cells.’, *Genesis (New York, N.Y. : 2000)*, 28(1), pp. 31–36.

Cheung, H.-H., Lee, T., Rennet, O. M. and Chan, W.-Y. (2009) ‘DNA methylation of Cancer Genome’, *Birth Defects REsearch Part C: Embryo Today: Reviews*, 87(4), pp. 335–350.

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. and Kühn, R. (2015) ‘Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells’, *Nature Biotechnology*, 33(5), pp. 543–548.

Clevers, H. (2006) ‘Wnt / β -Catenin Signaling in Development and Disease’, *Cell*, 127(3),

pp. 469–480.

Cobb, R. E., Chao, R. and Zhao, H. (2013) ‘Directed evolution: Past, Present and Future’, *American Institute of Chemical Engineers, Journal*, 59(5), pp. 1432–1440.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Hsu, P. D., Wu, X., Jiang, W. and Marraffini, L. A. (2013) ‘Multiplex Genome Engineering Using CRISPR/Cas Systems’, *Science*, 339(6121), pp. 819–823.

Croce, J. C. D. R. M. (2008) ‘Evolution of the Wnt pathways’, *Methods Molecular Biology*, 469, pp. 3–18.

Cruciat, C. and Niehrs, C. (2013) ‘Secreted and Transmembrane Wnt Inhibitors and Activators’, *Cold Spring Harbor perspectives in biology*, 5, p. a015081.

Daniel C. Koboldt, David E. Larson, Ken Chen, Li Ding, and R. K. W. (2012) ‘Massively parallel sequencing approaches for characterization of structural variation’, *Genomic Structural Variants*, 838(6), pp. 369–384.

Daniels, D. L. and Weis, W. I. (2005) ‘β -catenin directly displaces Groucho / TLE repressors from Tcf / Lef in Wnt-mediated transcription activation’, *Nature Structural and Molecular Biology*, 12(4), pp. 364–371.

Deltcheva, E., Chylinski, K., Sharma, C. M. and Gonzales, K. (2011) ‘CRISPR RNA maturation by trans -encoded small RNA and host factor RNase III’, *Nature*, 471(7340), pp. 602–607.

Drees, F., Pokutta, S., Yamada, S., Nelson, W. J. and Weis, W. I. (2005) ‘α-catenin is a molecular switch that binds E-cadherin-β-catenin and regulates actin-filament assembly’, *Cell*, 123(5), pp. 903–915.

Duval, A., Gayet, J. and Zhou, X. (1999) ‘Frequent Frameshift Mutations of the TCF-4 Gene in Colorectal Cancers with Microsatellite Instability Advances in Brief Frequent Frameshift Mutations of the TCF-4 Gene in Colorectal Cancers with’, *Cancer Research*, 59(17), pp. 4213–4215.

Engler, C., Kandzia, R. and Marillonnet, S. (2008) ‘A one pot, one step, precision cloning method with high throughput capability’, *PLoS ONE*, 3(11), p. e3647.

F. Ann Ran, Le Cong, Winston X. Yan, David A. Scott1, Jonathan S. Gootenberg, Andrea J. Kriz, Bernd Zetsche, Ophir Shalem, XuebingWu, Kira S. Makarova, Eugene V. Koonin, P. A. S. & F. Z. (2015) ‘In vivo genome editing using Staphylococcus aureus Cas9’, *Nature*, 520(7546), p. 186.

Fagotto, F., Glück, U. and Gumbiner, B. M. (1998) ‘Nuclear localization signal-independent and importin / karyopherin-independent nuclear import of b -catenin’, *Current Biology*, 8(4), pp. 181–190.

Ferrer-Vaquer, A., Piliszek, A., Tian, G., Aho, R. J., Dufort, D. and Hadjantonakis, A. K. (2010) ‘A sensitive and bright single-cell resolution live imaging reporter of Wnt/β-catenin signaling in the mouse’, *BMC Developmental Biology*, 10(1), p. 121.

Findlay, G. M., Boyle, E. a., Hause, R. J., Klein, J. C. and Shendure, J. (2014) ‘Saturation editing of genomic regions by multiplex homology-directed repair’, *Nature*. Nature Publishing Group, 513(7516), pp. 1–2.

Forbes, S. A., Bindal, N., Bamford, S., Cole, C., Kok, C. Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., Teague, J. W., Campbell, P. J., Stratton, M. R. and Futreal, P. A. (2010) ‘COSMIC : mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer’, *Nucleic Acids Research*, 39(suppl_1), pp. 945–950.

Forbes, S. A., Tang, G., Bindal, N., Bamford, S., Dawson, E., Cole, C., Kok, C. Y., Jia, M., Ewing, R., Menzies, A., Teague, J. W., Stratton, M. R. and Futreal, P. A. (2009) ‘COSMIC (the Catalogue of Somatic Mutations In Cancer): A resource to investigate acquired mutations in human cancer’, *Nucleic Acids Research*, 38(SUPPL.1), pp. 652–657.

Forbes, S. a, Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J. and Menzies, A. (2008) ‘The Catalogue of Somatic Mutations in Cancer (COSMIC)’, *Current protocols in human genetics*, 57(1), pp. 10–11.

Francisco J.M. Mojica, Ce´ sar Dı´ez-Villasen˜ or, Jesu´ s Garcı´a-Martı´ nez, E. S. and

Divisio´ (2005) ‘Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements’, *journal of molecular evolution*, 60(2), pp. 174–182.

Gamallo, C., Palacios, J., Moreno, G., De Mora, J. C., Suárez, A. and Armas, A. (1999) ‘β-Catenin expression pattern in stage I and II ovarian carcinomas: Relationship with β-catenin gene mutations, clinicopathological features, and clinical outcome’, *American Journal of Pathology*, 155(2), pp. 527–536.

Garriock, R. J., Warkman, A. S., Meadows, S. M., Agostino, S. D. and Krieg, P. A. (2007) ‘Census of Vertebrate Wnt genes : Isolation and Developmental Expression of Xenopus Wnt2 , Wnt3 , Wnt9a , Wnt9b , Wnt10a , and Wnt16’, *Developmental Dynamics: an official publication of the American Association of Anatomists*, 236(5), pp. 1249–1258.

Gaspar, C., Franken, P., Molenaar, L., Breukel, C., van der Valk, M., Smits, R. and Fodde, R. (2009) ‘A Targeted Constitutive Mutation in the Apc Tumor Suppressor Gene Underlies Mammary But Not Intestinal Tumorigenesis’, *PLoS Genetics*, 5(7), p. e1000547.

Gerlach, J. P., Emmink, B. L., Nojima, H., Kranenburg, O. and Maurice, M. M. (2014) ‘Wnt signalling induces accumulation of phosphorylated β-catenin in two distinct cytosolic complexes’, *Open Biology*, 4(11).

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. and Smith, H. O. (2009) ‘Enzymatic assembly of DNA molecules up to several hundred kilobases’, *Nature Methods*, 6(5), pp. 343–345.

Guellec, S. Le, Soubeyran, I., Rochaix, P. and Filleron, T. (2012) ‘CTNNB1 mutation analysis is a useful tool for the diagnosis of desmoid tumors : a study of 260 desmoid tumors and 191 potential morphologic mimics’, *Modern Pathology*, 25(12), pp. 1551–1558.

Haegel, H., Larue, L., Ohsugi, M., Fedorov, L., Herrenknecht, K. and Kemler, R. (1995) ‘Lack of β-catenin affects mouse development at gastrulation’, *Development*, 121(11), pp. 3529–3537.

Hart, M., Concordet, J., Lassot, I., Albert, I., Santos, R. L., Durand, H., Perret, C.,

Rubinfeld, B., Margottin, F., Benarous, R., Polakis, P. and Ad, G. (1999) 'The F-box protein β -TrCP associates with phosphorylated β -catenin and regulates its activity in the cell', *Current Biology*, 9(4), pp. 207–211.

Hébert, J. M., Boyle, M. and Martin, G. R. (1991) 'mRNA localization studies suggest that murine FGF-5 plays a role in gastrulation.', *Development*, 112(2), pp. 407–415.

Henderson, S., Chakravarthy, A., Su, X., Boshoff, C. and Fenton, T. R. (2014) 'APOBEC-Mediated Cytosine Deamination Links PIK3CA Helical Domain Mutations to Human Papillomavirus-Driven Tumor Development', *Cell Reports*. Elsevier, 7(6), pp. 1833–1841.

Hendrix, N. D., Wu, R., Kuick, R., Schwartz, D. R., Fearon, E. R. and Cho, K. R. (2006) 'Fibroblast Growth Factor 9 Has Oncogenic Activity and Is a Downstream Target of Wnt Signaling in Ovarian Endometrioid Adenocarcinomas', *Cancer Research*, 66(3), pp. 1354–1362.

Hinck, L., Ntithke, I. S., Papkoff, J., Physiology, C., Program, C. B. and City, R. (1994) 'Dynamics of Cadherin/Catenin Complex Formation: Novel Protein Interactions and Pathways of Complex Assembly', *The Journal of Cell Biology*, 125(6), pp. 1327–1340.

Hsu, P. D., Scott, D. a, Weinstein, J. a, Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. a, Bao, G. and Zhang, F. (2013) 'DNA targeting specificity of RNA-guided Cas9 nucleases.', *Nature biotechnology*, 31(9), pp. 827–32.

Huangfu, D. and Raya, A. (2017) 'CRISPR / Cas9-Based Engineering of the Epigenome', *Cell Stem Cell*, 21(4), pp. 431–447.

Huber, A. H., Nelson, W. J. and Weis, W. I. (1997) 'Three-Dimensional Structure of the Armadillo Repeat Region of β -Catenin', *Cell*, 90(5), pp. 871–882.

Huelsken, J., Vogel, R., Erdmann, B., Cotsarelis, G. and Birchmeier, W. (2001) 'β-Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin', *Cell*, 105(4), pp. 533–545.

Ikeda, S., Kishida, S., Yamamoto, H., Murai, H., Koyama, S. and Kikuchi, A. (1998) 'Axin

, a negative regulator of the Wnt signaling pathway , forms a complex with GSK-3 β and β -catenin and promotes GSK-3 β -dependent phosphorylation of β -catenin', *The EMBO Journal*, 17(5), pp. 1371–1384.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) 'Nucleotide Sequence of the iap Gene , Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia coli , and Identification of the Gene Product', *Journal of bacteriology*, 169(12), pp. 5429–5433.

Iwao, K., Nakamori, S., Kameyama, M., Imaoka, S., Kinoshita, M., Fukui, T., Ishiguro, S., Nakamura, Y. and Miyoshi, Y. (1998) 'Activation of the β-Catenin Gene by Interstitial Deletions Involving Exon 3 in Primary Colorectal Carcinomas without Adenomatous Polyposis Coli Mutations1 ExonS ExonS Normal Transcript Deleted Transcript', *Cancer Research*, 58(5), pp. 1021–1026.

Iyer, N. G., Özdag, H. and Caldas, C. (2004) 'p300/CBP and cancer', *Oncogene*, 23(24), pp. 4225–4231.

J. Mullen, T. Delaney, A. R. et al. (2013) 'β- Catenin Mutation Status and Outcomes in Sporadic Desmoid Tumors', *The Oncologist*, 18(9), pp. 1043–1049.

Jansen, R., Embden, J. D. A. Van, Gaastra, W. and Schouls, L. M. (2002) 'Identification of genes that are associated with DNA repeats in prokaryotes', *Molecular microbiology*, 43(6), pp. 1565–1575.

Jasin, M. (1996) 'Genetic manipulation of genomes with rare-cutting endonucleases', *Trends in Genetics*, 12(6), pp. 224–228.

Jensen, M. A., Ferretti, V., Grossman, R. L. and Staudt, L. M. (2017) 'The NCI Genomic Data Commons as an engine for precision medicine', *Blood*, 130(4), pp. 453–459..

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012) 'A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337, p. 1225829.

José Javier Otero, Weimin Fu, Lixin Kan, A. E. C. and J. A. K. (2004) 'β-Catenin signaling

is required for neural differentiation of embryonic stem cells', *Development*, 131(15), pp. 3545–3557.

Kadowaki, T., Wilder, E., Klingensmith, J., Zachary, K. and Perrimon, N. (1996) 'The segment polarity gene porcupine encodes a putative multitransmembrane protein involved in Wingless processing', *Genes and Development*, 10(24), pp. 3116–3128.

Karreman, C. (1998) 'A new set of positive/negative selectable markers for mammalian cells', *Gene*, 218(1–2), pp. 57–61.

Kass, E. M., Helgadottir, H. R., Chen, C.-C., Barbera, M., Wang, R., Westermark, U. K., Ludwig, T., Moynahan, M. E. and Jasin, M. (2013) 'Double-strand break repair by homologous recombination in primary mouse somatic cells requires BRCA1 but not the ATM kinase', *Proceedings of the National Academy of Sciences*, 110(14), pp. 5564–5569.

Kawano, Y. (2003) 'Secreted antagonists of the Wnt signalling pathway', *Journal of Cell Science*, 116(13), pp. 2627–2634.

Kielman, M. F., Rindapää, M., Gaspar, C., van Poppel, N., Breukel, C., van Leeuwen, S., Taketo, M. M., Roberts, S., Smits, R. and Fodde, R. (2002) 'Apc modulates embryonic stem-cell differentiation by controlling the dosage of β-catenin signaling', *Nature Genetics*, 32(4), pp. 594–605.

Kim, Y. G., Cha, J. and Chandrasegaran, S. (1996) 'Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain.', *Proceedings of the National Academy of Sciences*, 93(3), pp. 1156–1160.

Kishida, S., Yamamoto, H., Ikeda, S., Kishida, M., Sakamoto, I., Koyama, S. and Kikuchi, A. (1998) 'Axin , a Negative Regulator of the Wnt Signaling Pathway , Directly Interacts with Adenomatous Polyposis Coli and Regulates the Stabilization of β-Catenin *', *The Journal of Biological Chemistry*, 273(18), pp. 10823–10827.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N., Zheng, Z., Joung, J. K., Unit, P., Biology, I., Hospital, M. G. and Kong, H. (2016) 'High-fidelity CRISPR-CAS9 variants with undetectable genome-wide off-targets', *Nature*, 529(7587), pp. 490–

495.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V., Nguyen, T., Zheng, Z., Gonzales, A. P. W., Li, Z., Randall, T., Yeh, J. J., Aryee, M. J., Joung, J. K., Unit, P., Hospital, M. G., Biology, I., Hospital, M. G., Institutet, K. and Hospital, M. G. (2015) 'Engineered CRISPR-Cas9 nucleases with altered PAM specificities', *Nature*, 523(7561), pp. 481–485.

Komiya, Y. and Habas, R. (2008) 'Wnt signal transduction pathways', *Organogenesis*, 4(2), pp. 68–75.

Lazar, A. J. F., Tuvin, D., Bolshakov, S., Mayordomo-aranda, E., Warneke, C. L., Lopez-terrada, D., Pollock, R. E. and Lev, D. (2008) 'Specific Mutations in the β -Catenin Gene (CTNNB1) Correlate with Local Recurrence in Sporadic Desmoid Tumors', *American Journal of Pathology*, 173(5), pp. 1518–1527.

Legoix, P., Bluteau, O., Bayer, J., Perret, C., Balabaud, C., Belghiti, J., Franco, D., Thomas, G., Laurent-puig, P. and Zucman-rossi, J. (1999) 'Beta-catenin mutations in hepatocellular carcinoma correlate with a low rate of loss of heterozygosity', *Oncogene*, 18, pp. 4044–4046.

Lickert, H., Domon, C., Huls, G., Wehrle, C., Duluc, I., Clevers, H., Meyer, B. I., Freund, J. N. and Kemler, R. (2000) 'Wnt/(beta)-catenin signaling regulates the expression of the homeobox gene Cdx1 in embryonic intestine.', *Development (Cambridge, England)*, 127(17), pp. 3805–3813.

Lieber, M. R. (2010) 'The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway', *Annual review of biochemistry*, 79, pp. 181–211.

Liu, C., Li, Y., Semenov, M., Han, C., Baeg, G.-H., Tan, Y., Zhang, Z., Lin, X. and He, X. (2002) 'Control of β-Catenin Phosphorylation/Degradation by a Dual-Kinase Mechanism', *Cell*, 108(6), pp. 837–847.

Liu, G., Bafico, A., Harris, V. K. and Aaronson, S. A. (2003) 'A Novel Mechanism for Wnt Activation of Canonical Signaling through the LRP6 Receptor', *Molecular and cellular*

biology, 23(16), pp. 5825–5835.

Liu, W., Dong, X., Mai, M., Seelan, R. S., Taniguchi, K., Krishnadath, K. K., Halling, K. C., Cunningham, J. M., Qian, C., Christensen, E., Roche, P. C., Smith, D. I. and Thibodeau, S. N. (2000) 'Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating β-catenin/TCF signalling', *Nature Genetics*, 26(2), pp. 146–147.

Lustig, B., Jerchow, B., Sachs, M., Weiler, S., Pietsch, T., Karsten, U., van de Wetering, M., Clevers, H., Schlag, P. M., Birchmeier, W. and Behrens, J. (2002) 'Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors.', *Molecular and cellular biology*, 22(4), pp. 1184–93.

Macdonald, B. T. and He, X. (2012) 'Frizzled and LRP5 / 6 Receptors for Wnt / b -Catenin Signaling', *Cold Spring Harbor Perspectives in Biology*, 4, p. a007880.

Maiti, S., Alam, R., Amos, C. I. and Huff, V. (2000) 'Frequent Association of β -Catenin and WT1 Mutations in Wilms Tumors 1', *Cancer Research*, 60(22), pp. 6288–6292.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. and Church, G. M. (2013) 'RNA-Guided Human Genome Engineering via Cas9 Prashant', *Science*, 339(6121), pp. 823–826.

Mansour, S., Thomas, K. and Capecchi, M. (1988) 'Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes', *Nature*, 336(6197), p.348.

Mao, T., Chu, J., Jeng, Y., Lai, P. and Hsu, H. (2001) 'Expression of mutant nuclear b -catenin correlates with non-invasive hepatocellular carcinoma , absence of portal vein spread , and good prognosis', *Journal of Pathology*, 9896(193), pp. 95–101.

Marraffini, L. A. (2016) *The CRISPR-Cas system of Streptococcus pyogenes : function and applications*, *Streptococcus pyogenes : Basic Biology to Clinical Manifestations [Internet].* Oklahoma City (OK).

Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, a M., Ingram, J. R. and Ploegh,

H. L. (2015) ‘Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining’, *Nat Biotechnol*, 33(5), pp. 538–542.

Maxam, a M. and Gilbert, W. (1977) ‘A new method for sequencing DNA.’, *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), pp. 560–564.

Megy, S., Bertho, G., Gharbi-Benarous, J., Evard-Todeschi, N., Coadou, G., Segeral, E., Ihele, C., Quememeur, E., Benarous, R. and Girault, J. P. (2005) ‘STD and TRNOESY NMR Studies on the Conformation of the Oncogenic Protein -Catenin Containing the Phosphorylated Motif DpSGXXpS Bound to the -TrCP Protein’, *Journal of Biological Chemistry*, 280(32), pp. 29107–29116.

Meldrum, C., Doyle, M. a and Tothill, R. W. (2011) ‘Next-Generation Sequencing for Cancer Diagnostics: a Practical Perspective’, *The Clinical Biochemist Reviews*, 32(4), pp. 177–195.

Merkle, F. T., Neuhausser, W. M., Santos, D., Valen, E., Gagnon, J. A., Maas, K., Sandoe, J., Schier, A. F. and Eggan, K. (2015) ‘Efficient CRISPR-Cas9-Mediated Generation of Knockin Human Pluripotent Stem Cells Lacking Undesired Mutations at the Targeted Locus’, *Cell Reports*. Elsevier, 11(6), pp. 875–883.

Merrill, B. J. (2012) ‘Wnt pathway regulation of embryonic stem cell self-renewal’, *Cold Spring Harbor Perspectives in Biology*, 4(9), pp. 1–17.

Miller, M. L., Reznik, E., Gauthier, N. P., Aksoy, A., Korkut, A., Gao, J., Ciriello, G., Schultz, N., Sander, C. and Gao, J.-J. (2015) ‘Pan-Cancer Analysis of Mutation Hotspots in Protein Domains Supplemental Information for: Pan-Cancer Analysis of Mutation Hotspots in Protein Domains’, *Cell Systems*, 1(3), pp. 197–209.

Miyoshi, Y., Iwao, K., Nagasawa, Y., Aihara, T., Sasaki, Y., Imaoka, S., Murata, M. and Shimano, T. (1998) ‘Activation of the β-Catenin Gene in Primary Hepatocellular Carcinomas by’, *Cancer Research*, 58(12), pp. 2524–2527.

Mojica, F.J., Díez-Villaseñor, C., Soria, E. and Juez, G. (2000) ‘Biological significance of a family of regularly spaced repeats in the genomes of Archea, Bacteria and

mitochondria', *Molecular microbiology*, 36(1), pp. 244–246.

Morgan L Maeder, Samantha J Linder, Vincent M Cascio, Yanfang Fu, Quan H Ho1, and J. K. J. (2014) 'CRISPR RNA-guided activation of endogenous human genes Morgan', *Nature Methods*, 10(10), pp. 977–979.

Mori, Y., Nagse, H., Ando, H., Horii, A., Ichii, S., Nakatsuru, S., Aoki, T., Miki, Y., Mori, T. and Nakamura, Y. (1992) 'Somatic mutations of the APC gene in colorectal tumors: Mutation cluster region in the APC gene', *Human Molecular Genetics*, 1(4), pp. 229–233.

Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. and Kinzlert, K. W. (1997) 'Activation of β-Catenin-Tcf Signaling in Colon Cancer by Mutations in β-Catenin or APC. *Science*, 275(5307), pp. 1787–1790.

Murata, M., Iwao, K., Miyoshi, Y. and Nagasawa, Y. (2000) 'Activation of the b -catenin gene by interstitial deletions involving exon 3 as an early event in colorectal tumorigenesis', *Cancer Letters*, 159(1), pp. 73–78.

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A. and Menzies, A., (2012) 'Mutational processes molding the genomes of 21 breast cancers', *Cell*, 149(5), pp. 979–993.

Nusse, R., Brown, A., Papkoff, J., Scambler, P., Shackeford, G., McMahon, A., Moon, R. and Varmus, H. (1991) 'A new Nomenclature for int-1 and Related Genes: The Wnt Gene Family', *Cell*, 64(2), pp. 231–232.

Nusse, R. and Varmus, H. E. (1982) 'Many Tumors Induced by the Mouse Mammary Tumor Virus Contain a Provirus Integrated in the Same Region of the Host Genome', *Cell*, 31(1), pp. 99–109.

Omagnolo, A. R., Illuart, P. I. B. and Nge, C. L. (1998) 'Somatic mutations of the β -catenin gene are frequent in mouse and human hepatocellular carcinomas', *Proceedings of the National Academy of Sciences of the United States of America*, 95(15), pp. 8847–8851.

Ozawa, M., Baribault, H. and Kemler, R. (1989) 'The cytoplasmic domain of the cell

adhesion molecule uvomorulin associates with three independent proteins structurally related in different species', *The EMBO Journal*, 8(6), pp. 1711–1717.

Paix, A., Folkmann, A., Goldman, D. H., Kulaga, H., Grzelak, M. J., Rasoloson, D., Paidemarry, S., Green, R., Reed, R. R. and Seydoux, G. (2017) 'Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks', *Proceedings of the National Academy of Sciences*, 114(50), pp. E10745–E10754.

Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K. M., Gregg, A., Noggle, S. and Tessier-Lavigne, M. (2016) 'Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9', *Nature*, 533(7601), pp. 125–129.

Parker, D. S., Ni, Y. Y., Chang, J. L., Li, J. and Cadigan, K. M. (2008) 'Wingless Signaling Induces Widespread Chromatin Remodeling of Target Loci', *Molecular and Cellular Biology*, 28(5), pp. 1815–1828.

Pastink, A., Eeken, J. C. J. and Lohman, P. H. M. (2001) 'Genomic integrity and the repair of double-strand DNA breaks', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis.*, 480, pp. 37–50.

Peyrieras, N., Louvard, D. and Jacob, F. (1985) 'Characterization of antigens recognised by monoclonal and polyclonal antibodies directed against uvomorulin', *Proceedings of the National Academy of Sciences*, 82(23), pp. 8067–8071.

Pfeifer, G. P. (2010) 'Environmental exposures and mutational patterns of cancer genomes', *Genome Med*, 2(8), p. 54.

Pfeifer, G. P., Denissenko, M. F., Olivier, M., Tretyakova, N., Hecht, S. S. and Hainaut, P. (2002) 'Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers', *Oncogene*, 21–48(6), pp. 7435–7451.

Piedra, J., Miravet, S., Castan, J., Heisterkamp, N., Herreros, A. G. De and Dun, M. (2003) 'p120 Catenin-Associated Fer and Fyn Tyrosine Kinases Regulate β-catenin Tyr-142 Phosphorylation and β-catenin – α-catenin Interaction, *Molecular and cellular biology*, 23(7), pp. 2287–2297.

Pierotti MA, Sozzi G, C. C. (2003) ‘Mechanisms of oncogene activation.’, in Kufe DW, Pollock RE, Weichselbaum RR, et al. (ed.) *Holland-Frei Cancer Medicine 6th edition*. 6th edn. Hamilton (ON): BC Decker.

Pilon, N., Oh, K., Sylvestre, J.-R., Savory, J. G. A. and Lohnes, D. (2007) ‘Wnt signaling is a key mediator of Cdx1 expression in vivo’, *Development*, 134(12), pp. 2315–2323.

Pokutta, S. and Weis, W. I. (2000) ‘Structure of the dimerization and beta-catenin-binding region of alpha-catenin.’, *Molecular Cell*, 5(3), pp. 533–43.

Polakis, P. (1999) ‘The oncogenic activation of β-catenin’, *Current Opinion in Genetics & Development*, 9(1), pp. 15–21.

Polakis, P. (2000) ‘Wnt signaling and cancer’, *Genes and Development*, 14(15), pp. 1837–1851.

Polakis, P. (2012) ‘Wnt signaling in cancer.’, *Cold Spring Harbor perspectives in biology*, 4(5), pp. 1–13.

Port, F. and Basler, K. (2010) ‘Wnt Trafficking : New Insights into Wnt Maturation , Secretion and Spreading’, *Traffic*, 11(10), pp. 1265–1271.

Provost, E., Yamamoto, Y., Lizardi, I., Stern, J., D’Aquila, T. G., Gaynor, R. B. and Rimm, D. L. (2003) ‘Functional correlates of mutations in β-catenin exon 3 phosphorylation sites’, *Journal of Biological Chemistry*, 278(34), pp. 31781–31789.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. and Lim, W. A. (2013) ‘Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression’, *Cell*, 152(5), pp. 1173–1183.

Raggioli, A., Junghans, D., Rudloff, S. and Kemler, R. (2014) ‘Beta-catenin is vital for the integrity of mouse embryonic stem cells’, *PLoS ONE*, 9(1), pp. 1–9.

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. a., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F. (2013) ‘Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity’, *Cell*. Elsevier, 154(6), pp.

1380–1389.

Ran, F. A., Hsu, P. P. D., Wright, J., Agarwala, V., Scott, D. a and Zhang, F. (2013) ‘Genome engineering using the CRISPR-Cas9 system.’, *Nature protocols*, 8(11), pp. 2281–308.

Ratz, M., Testa, I., Hell, S. W. and Jakobs, S. (2015) ‘CRISPR/Cas9-mediated endogenous protein tagging for RESOLFT super-resolution microscopy of living human cells’, *Scientific Reports*, 5, p. 9592.

Renard, C. A., Labalette, C., Armengol, C., Cougot, D., Wei, Y., Cairo, S., Pineau, P., Neuveut, C., De Reyniès, A., Dejean, A., Perret, C. and Buendia, M. A. (2007) ‘Tbx3 is a downstream target of the Wnt/β-catenin pathway and a critical mediator of β-catenin survival functions in liver cancer’, *Cancer Research*, 67(3), pp. 901–910.

Riggleman, B. (1989) ‘Molecular analysis of the armadillo locus : uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene’, *Genes and Development*, 3(1), pp. 96–113.

Rijsewijk, F. (1987) ‘The Drosophila Homolog of the Mouse Mammary Oncogene int-1 Is Identical to the Segment Polarity Gene wingless’, *Cell*, 50(4), pp. 649–657.

Rimm, D. L., Koslov, E. R., Kebriaei, P., Cianci, C. D. and Morrow, J. S. (1995) ‘Alpha 1(E)-catenin is an actin-binding and -bundling protein mediating the attachment of F-actin to the membrane adhesion complex.’, *Proceedings of the National Academy of Sciences*, 92(19), pp. 8813–8817.

Robbins, P. F., E-gamil, M., Li, Y. F., Kawakami, Y., Loftus, D., Appella, E. and Steven, A. (1996) ‘A Mutated β-Catenin Gene Encodes a Melanoma-specific Antigen Recognized by Tumor Infiltrating Lymphocytes’, *Journal of Experimental Medicine*, 183(3), pp. 1185–1192.

Rodolphe, B., Christophe, F., Deveau, H., Melissa, R., Boyaval, P., Moineau, S., Romero, D. A. and Horvath, P. (2007) ‘CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes’, *Science*, 315(5819), pp. 1709–1712.

Roura, S., Miravet, S., Garcı, A. and Herreros, D. (1999) ‘Regulation of E-cadherin / Catenin Association by Tyrosine Phosphorylation *’, *The Journal of Biological Chemistry*, 274(51), pp. 36734–36740.

Russell, R., Ilg, M., Lin, Q., Wu, G., Lechel, A., Bergmann, W., Eiseler, T., Linta, L., Kumar, P. P., Klingenstein, M., Adachi, K., Hohwieler, M., Sakk, O., Raab, S., Moon, A., Zenke, M., Seufferlein, T., Schöler, H. R., Illing, A., Liebau, S. and Kleger, A. (2015) ‘A Dynamic Role of TBX3 in the Pluripotency Circuitry’, *Stem Cell Reports*, 5(6), pp. 1155–1170.

Ryding, A. D. S., Sharp, M. G. F. and Mullins, J. J. (2001) ‘Conditional transgenic technologies’, *Journal of Endocrinology*, 171(1), pp. 1–14.

Sakanaka, C., Weiss, J. B. and Williams, L. T. (1998) ‘Bridging of beta-catenin and glycogen synthase kinase-3beta by axin and inhibition of beta-catenin-mediated transcription.’, *Proceedings of the National Academy of Sciences of the United States of America*, 95(6), pp. 3020–3.

Saleem, M., Singh, P., Mir, R. A. and Jamal, M. (2017) ‘Beta-catenin N-terminal domain : An enigmatic region prone to cancer causing mutations’, *Mutation Research*. Elsevier B.V., 773, pp. 122–133.

San Fillipo, J., Sung, P. and Klein, H. (2008) ‘Mechanism of Eukaryotic Homologous Recombination’, *Annual review of biochemistry*, 77, pp. 229–257.

Sánchez-Rivera, F. J. and Jacks, T. (2015) ‘Applications of the CRISPR-Cas9 system in cancer biology’, *Nature Reviews Cancer*. Nature Publishing Group, 15(7), pp. 387–395.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) ‘DNA sequencing with chain-terminating inhibitors’, *Proceedings of the National Academy of Sciences*, 74(12), pp. 5463–5467.

Schwartz, F., Maeda, N., Smithies, O., Hickey, R., Edelmann, W., Skoultchi, a and Kucherlapati, R. (1991) ‘A dominant positive and negative selectable gene for use in mammalian cells.’, *Proceedings of the National Academy of Sciences of the United States of America*, 88(23), pp. 10416–20.

Sierra, J., Yoshida, T., Joazeiro, C. a and Jones, K. a (2006) 'The APC tumor suppressor counteracts beta-catenin activation and H 3 K 4 methylation at Wnt target genes', *Genes & Dev.*, 20(5), p. 586.

Silberg, D. G., Swain, G. P., Suh, E. R. and Traber, P. G. (2000) 'Cdx1 and Cdx2 Expression During Intestinal Development', *Gastroenterology*, 119(4), pp. 961–971.

Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. and Kucherlapati, R. S. (1985) 'Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination.', *Nature*, 317(6034), pp. 230–4.

Soldner, F., Laganière, J., Cheng, A. W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L. I., Myers, R. H., Lindquist, S., Zhang, L., Guschin, D., Fong, L. K., Vu, B. J., Meng, X., Urnov, F. D., Rebar, E. J., Gregory, P. D., Zhang, H. S. and Jaenisch, R. (2011) 'Generation of isogenic pluripotent stem cells differing exclusively at two early onset parkinson point mutations', *Cell*, 146(2), pp. 318–331.

Sparks, A. B., Morin, P. J., Vogelstein, B. and Kinzler, K. W. (1998) 'Advances in Brief Mutational Analysis of the APC /β-Catenin / Tcf Pathway in Colorectal Cancer', *Cancer Research*, 58(6), pp. 1130–1134.

Stefanovic, S. and Christoffels, V. M. (2015) 'GATA-dependent transcriptional and epigenetic control of cardiac lineage specification and differentiation', *Cellular and Molecular Life Sciences*. Springer Basel, 72(20), pp. 3871–3881.

Stephens, P. J., Greenman, C. D., Fu, B., Yang, F., Bignell, G. R., Mudie, L. J., Pleasance, E. D., Lau, K. W., Beare, D., Stebbings, L. A., McLaren, S., Lin, M. L., McBride, D. J., Varela, I., Nik-Zainal, S., Leroy, C., Jia, M., Menzies, A., Butler, A. P., Teague, J. W., Quail, M. A., Burton, J., Swerdlow, H., Carter, N. P., Morsberger, L. A., Iacobuzio-Donahue, C., Follows, G. A., Green, A. R., Flanagan, A. M., Stratton, M. R., Futreal, P. A. and Campbell, P. J. (2011) 'Massive genomic rearrangement acquired in a single catastrophic event during cancer development', *Cell*. Elsevier Inc., 144(1), pp. 27–40.

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., Jennifer, A., Biophysics, M.,

Biology, C., Division, B. and Berkeley, L. (2014) ‘DNA interrogation by the CRISPR RNA-guided endonuclease Cas9’, *Nature*, 507(7490), pp. 62–67.

Subramanian, V., Meyer, B. I. and Gruss, P. (1995) ‘Disruption of the murine homeobox gene Cdx1 affects axial skeletal identities by altering the mesodermal expression domains of Hox genes’, *Cell*, 83(4), pp. 641–653.

Taniguchi, K., Roberts, L. R., Aderca, I. N., Dong, X., Qian, C., Murphy, L. M., Nagorney, D. M., Burgart, L. J., Roche, P. C., Smith, D. I., Ross, J. A. and Liu, W. (2002) ‘Mutational spectrum of β-catenin, AXIN1, and AXIN2 in hepatocellular carcinomas and hepatoblastomas’, *Oncogene*, 21(31), pp. 4863–4871.

Thomas, K. R. and Capecchi, M. R. (1987) ‘Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells’, *Cell*, 51(3), pp. 503–512.

Thomas, K. R., Folger, K. R. and Capecchi, M. R. (1986) ‘High frequency targeting of genes to specific sites in the mammalian genome’, *Cell*, 44(3), pp. 419–428.

Toyama, T., Lee, H. C., Koga, H., Wands, J. R. and Kim, M. (2010) ‘Noncanonical Wnt11 Inhibits Hepatocellular Carcinoma Cell Proliferation and Migration’, *Molecular Cancer Research*, pp. 1541–7786.

Turner, D. a., Rue, P., Mackenzie, J. P., Davies, E. and Martinez Arias, A. (2014) ‘Brachyury cooperates with Wnt/ β-Catenin signalling to elicit Primitive Streak like behaviour in differentiating mouse ES cells.’, *BMC Biology*, 12(1), p. 63.

Vanamee, É. S., Santagata, S. and Aggarwal, A. K. (2001) ‘FokI requires two specific DNA sites for cleavage’, *Journal of Molecular Biology*, 309(1), pp. 69–78.

Varmus, H. E. (1988) ‘Expression of the int-1 Gene in Transgenic Mice Is Associated with Mammary G land Hyperpbia and Adenocarcinomas in Male and Female Mice’, *Cell*, 55(4), pp. 619–625.

Veeman, M. T., Axelrod, J. D. and Moon, R. T. (2003) ‘A Second Canon : Functions and Mechanisms of β -catenin-Independent Wnt Signaling’, *Developmental Cell*, 5(3), pp.

367–377.

Vestweber, D. and Kemler, R. (1984) ‘Some structural and functional aspects of the cell adhesion molecule uvomorulin’, *Cell Differentiation*, 15(2–4), pp. 269–273.

Viel, A., Bruselles, A., Meccia, E., Fornasarig, M., Quaia, M., Canzonieri, V., Policicchio, E., Urso, E. D., Agostini, M., Genuardi, M., Lucci-Cordisco, E., Venesio, T., Martayan, A., Diodoro, M. G., Sanchez-Mete, L., Stigliano, V., Mazzei, F., Grasso, F., Giuliani, A., Baiocchi, M., Maestro, R., Giannini, G., Tartaglia, M., Alexandrov, L. B. and Bignami, M. (2017) ‘A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer’, *EBioMedicine*. The Authors, 20, pp. 39–49.

Vleminckx, K., Kemler, R. and Hecht, A. (1999) ‘The C-terminal transactivation domain of β-catenin is necessary and sufficient for signaling by the LEF-1 /β-catenin complex in Xenopus laevis’, *Mechanism of development*, 81(1–2), pp. 65–74.

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. a and Kinzler, K. W. (2013) ‘Cancer Genome Lanscapes’, *Science*, 339(6127), pp. 1546–1558.

Vouillot, L., Thélie, A. and Pollet, N. (2015) ‘Comparison of T7E1 and Surveyor Mismatch Cleavage Assays to Detect Mutations Triggered by Engineered Nucleases’, *G3 Genes,Genomes,Genetics*, 5(3), pp. 407–415.

Wang, Z., Vogelstein, B. and Kinzler, K. W. (2003) ‘Phosphorylation of β-catenin at S33, S37, or T41 can occur in the absence of phosphorylation at T45 in colon cancer cells’, *Cancer Research*, 63(17), pp. 5234–5235.

Wege, H., Heim, D., Lütgehetmann, M., Dierlamm, J., Lohse, A. W. and Brümmendorf, T. H. (2011) ‘Forced Activation of β-Catenin Signaling Supports the Transformation of hTERT-Immortalized Human Fetal Hepatocytes.’, *Molecular cancer research : MCR*, 9(9), pp. 1222–1231.

Weidgang, C. E., Russell, R., Tata, P. R., Kühl, S. J., Illing, A., Müller, M., Lin, Q., Brunner, C., Boeckers, T. M., Bauer, K., Kartikasari, A. E. R., Guo, Y., Radenz, M., Bernemann, C., Weiß, M., Seufferlein, T., Zenke, M., Iacovino, M., Kyba, M., Schöler, H. R., Kühl, M., Liebau, S. and Kleger, A. (2013) ‘TBX3 directs cell-fate decision toward

mesendoderm', *Stem Cell Reports*, 1(3), pp. 248–265.

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J. M and Cancer Genome Atlas Research Network, (2013) 'The Cancer Genome Atlas Pan-Cancer analysis project', *Nature Genetics*. Nature Publishing Group, 45(10), pp. 1113–1120.

Wend, P., Fang, L., Zhu, Q., Schipper, J. H., Loddenkemper, C., Kosel, F., Brinkmann, V., Eckert, K., Hindersin, S., Holland, J. D., Lehr, S., Kahn, M., Ziebold, U. and Birchmeier, W. (2013) 'Wnt/β-catenin signalling induces MLL to create epigenetic changes in salivary gland tumours', *The EMBO Journal*, 32(14), pp. 1977–1989. \

Wetering, M. Van De, Cavallo, R., Dooijes, D., Beest, M. Van, Es, J. Van, Loureiro, J., Ypma, A., Hursh, D., Jones, T., Bejsovec, A., Peifer, M., Mortin, M., Clevers, H., Hill, C. and Carolina, N. (1997) 'Armadillo Coactivates Transcription Driven by the Product of the Drosophila Segment Polarity Gene dTCF', *Cell*, 88(6), pp. 789–799.

Whitehead, I. A. N., Kirk, H. and Kay, R. (1995) 'Expression Cloning of Oncogenes by Retroviral Transfer of cDNA Libraries', *Molecular and cellular biology*, 15(2), pp. 704–710.

Willert, K. and Nusse, R. (1998) 'β-catenin: a key mediator of Wnt signaling', *Current Opinion in Genetics & Development*, pp. 95–102.

Winter, M. C., Shasby, S. and Shasby, D. M. (2008) 'Compromised E-cadherin adhesion and epithelial barrier function with activation of G protein-coupled receptors is rescued by Y-to-F mutations in β-catenin', *American Journal of Physiology-Lung Cell Molecular Physiology*, 294(8), pp. 442–448.

Withers, G. S., Zoback, M. D., Apel, R., Baumgarter, J., Brudy, M., Emmermann, R., Engeser, B., Fuchs, K., Kessels, W., Rischmuller, H., Rummel, F. and Vernlk, L. (1993) 'Instability and decay of the primary structure of DNA', *Nat. Lett.*, 36(6422), p. 709.

Wright, D. A., Thibodeau-Beganny, S., Sander, J. D., Winfrey, R. J., Hirsh, A. S., Eichtinger, M., Fu, F., Porteus, M. H., Dobbs, D., Voytas, D. F. and Joung, J. K. (2006) 'Standardized reagents and protocols for engineering zinc finger nucleases by modular

assembly', *Nature Protocols*, 1(3), pp. 1637–1652.

Wright, K. J. and Tjian, R. (2009) 'Wnt signaling targets ETO coactivation domain of TAF4/TFIID in vivo.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(1), pp. 55–60.

Wu, G., Xu, G., Schulman, B. a., Jeffrey, P. D., Harper, J. W. and Pavletich, N. P. (2003) 'Structure of a β-TrCP1-Skp1-β-catenin complex: Destruction motif binding and lysine specificity of the SCFβ-TrCP1 ubiquitin ligase', *Molecular Cell*, 11(6), pp. 1445–1456.

Wu, G., Xu, G., Schulman, B. A., Jeffrey, P. D., Harper, J. W. and Pavletich, N. P. (2003) 'Structure of a β-TrCP1-Skp1-β-catenin complex: Destruction motif binding and lysine specificity of the SCFβ-TrCP1ubiquitin ligase', *Molecular Cell*, 11(6), pp. 1445–1456.

Xu, W. and Kimelman, D. (2007) 'Mechanistic insights from structural studies of beta-catenin and its binding partners', *Journal of Cell Science*, 120(19), pp. 3337–3344.

Yamada, S., Pokutta, S., Drees, F., Weis, W. I. and Nelson, W. J. (2005) 'Deconstructing the cadherin-catenin-actin complex', *Cell*, 123(5), pp. 889–901.

Yang, L., Guell, M., Byrne, S., Yang, J. L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A. W., Rios, X., Huang, P. Y., Daley, G. and Church, G. (2013) 'Optimization of scarless human stem cell genome editing', *Nucleic Acids Research*, 41(19), pp. 9049–9061.

Yi Xing, Ken-Ichi Takemaru, Jing Liu, Jason D. Berndt, Jie J. Zheng, Randall T. Moon, and W. X. (2014) 'Crystal Structure of a Full-Length β -Catenin', *Structure*, 16(3), pp. 478–487.

Yokoya, F., Imamoto, N., Tachibana, T. and Yoneda, Y. (1999) 'beta-catenin can be transported into the nucleus in a Ran-unassisted manner.', *Molecular biology of the cell*, 10(4), pp. 1119–31.

Yu, C., Liu, Y., Ding, S., Qi, L. S., Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., Russa, M. La and Xie, M. (2015) 'Small Molecules Enhance CRISPR Genome Editing in Small Molecules Enhance CRISPR Genome Editing in Pluripotent Stem Cells', *Cell*

Stem Cell. Elsevier Inc., 16(2), pp. 142–147.

Zeng, X., Huang, H., Tamai, K., Zhang, X., Harada, Y. and Yokota, C. (2017) 'Initiation of Wnt signaling : control of Wnt coreceptor Lrp6 phosphorylation / activation via frizzled , dishevelled and axin functions', *Development*, 135(2), pp. 367–375.

Zhai, Y., Wu, R., Schwartz, D. R., Darrah, D., Reed, H., Kolligs, F. T., Nieman, M. T., Fearon, E. R. and Cho, K. R. (2002) 'Role of β -Catenin/T-Cell Factor-Regulated Genes in Ovarian Endometrioid Adenocarcinomas Yali', *American Journal of Pathology*, 160(4), pp. 1229–1238.