FUNCTIONAL GENETICS OF CANCER AND CONGENITAL DISORDERS

JAROSLAV ZAK

A thesis submitted to the Board of the Faculty of Medical Sciences in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

at the

University of Oxford

Lincoln College

Trinity Term 2016

ABSTRACT

The genetic architectures of cancer and congenital disorders are heterogeneous and incompletely mapped. Rare and low-frequency variants of incomplete penetrance are emerging as an important class of germline and somatic variation, but their contribution to disease remains poorly characterised. This thesis aims to identify and assess pathogenic mutations in the 1q41q42 microdeletion syndrome, neural tube defects, neuropsychiatric disorders and cancer. Rare microdeletions at the 1q41q42 locus cause a clinically heterogeneous syndrome characterized by developmental delay, characteristic dysmorphic features and brain morphological abnormalities. Examining new and published patients with 1q41q42 microdeletions, we found that *TP53BP2*, encoding ASPP2, is a strong candidate for being the gene responsible for brain morphological abnormalities of the syndrome. Mice deficient for *Trp53bp2* show multiple abnormalities overlapping the features of the 1q41q42 microdeletion syndrome such as dysmorphic lateral ventricles, heart and urogenital abnormalities. ASPP2 deficiency also causes neural tube defects, hopping gait, and male-specific motion hyperactivity in mice. We further identify candidate pathogenic *TP53BP2* duplications, implicating *TP53BP2* dosage sensitivity in the ganglionic eminences of the developing brain, manifested by structural abnormalities in the striatum and lateral ventricles of both deletion and duplication patients. ASPP2 controls neuroepithelial cell polarity via Par3 and genetic disruption of aPKC-Par3 interaction by rare missense variants was implicated in human neural tube defects. An integrative analysis of cancer genomic data revealed that *PPP1R13B*, encoding ASPP1, bears many hallmarks of a tumour suppressor gene, despite being mutated at a low absolute frequency. A subset of missense somatic mutations in ASPP genes genetically interact with

TP53 mutations, disrupting an autoinhibitory mechanism to modulate p53-dependent transcription. In summary, this work identified novel candidate pathogenic variants in developmental disorders and cancer, and explored the mechanisms underlying their respective genotype-phenotype links.

DEDICATION

This work is dedicated to my mother, in loving memory.

ACKNOWLEDGMENTS

I would like to thank Xin for her guidance and the scientific freedom she has given me, for which I am very grateful. Xin, thank you for being a great mentor. My thanks also go to John who taught me immunology from scratch and allowed me to dive into a risky project. It has been a great ride! Thanks to Ian for looking after us Skaggs-Oxford folk during our Oxford years. A big thanks goes to Ben SB for 'encouraging' me to learn R, and putting up with all the beginners' questions. Let's also not forget the lab meeting cakes. Thanks also to Richard Bryant, Marc, Celine and Kevin - it's been a pleasure collaborating with you. I'm grateful to Elena, Daniel, Dorota, Indi, Lucy, Liz, Shan, Melissa, Halley, Vincent, Ben Davies, Douglas, Khatoun, James, Carmella, Alex, Gay and Ivy for their hands-on help. To Yihua, Min, Ludo, Gopi and Namir - I have learnt a lot from you, thanks for your patience and all the interesting discussions. I am grateful to Mary for her writing assistance and her attention to detail. To Steffi and Isa, talented Master's students-turned-PhD students, thank you for your good work. A big thanks to Veronica, you are the lifeline of the Ludwig! Thanks to Anna, Mario, Luca, David, Mike, Richard Owen, Xiao, Jingyi and the whole Lu lab for making everyday lab life fun. You will be missed! A shout out to the Skaggs cohort, Anne, Keary, Owen, JL and Dan. Last but not least, I'm grateful to all my collaborators in Oxford, at Scripps, at hospitals and elsewhere; thank you for making the research possible.

CONTENTS

ABBREVIATIONS AND ACRONYMS

ACC	Agenesis of corpus callosum
ACMG	American College of Medical Genetics and Genomics
ALL	Acute lymphoblastic leukemia
ALS	Amyotrophic lateral sclerosis
ALVIN	Automatic Lateral Ventricle delineation
AML	Acute myeloid leukaemia
Ank	Ankyrin
APC	Adenomatous polyposis coli
API	Application programming interface
aPKCBD	Atypical protein kinase C binding domain
APOB	Apoliprotein B
AR	Androgen receptor
Arm	LCMV Armstrong clone
ASD	Autism spectrum disorder
ASPP	Apoptosis-stimulating protein of p53; Ankyrin repeats, SH3 domain, Proline-rich region containing protein
ATM	Ataxia telangiectasia mutated
B6	C57BL/6
BAX	Bcl-2-associated X protein
BCL2	B-cell lymphoma 2
BCR	B-cell receptor
BRAF	v-raf murine sarcoma viral oncogene homolog B1
BRCA1	Breast cancer 1, early onset
BWA	Burrows-Wheeler Aligner
C	Cytosine
Ca	Calcium
CagA	Cytotoxin-associated gene A
CAPN2	Calpain 2
CAR T cell	Chimaeric antigen receptor
CCL	Chemokine ligand
CCLE	Cancer Cell Line Encyclopaedia
CCR	Chemokine receptor
CCW	Craniocortical width
CDK	Cyclin-dependent kinase
Cdkn1a	Cyclin-Dependent Kinase Inhibitor 1A
Cdkn1b	Cyclin-Dependent Kinase Inhibitor 1B
cDNA	Complementary DNA
CELSR1	Cadherin EGF LAG Seven-Pass G-Type Receptor 1
CGDS	Cancer Genomic Data Server
CHARGE	Coloboma, heart defect, atresia choanae, retarded growth and development, genital abnormality, and ear abnormality
CHASM	Cancer-Specific High-Throughput Annotation of Somatic Mutations
cl13	LCMV clone 13

CLASP	Cytoplasmic linker associated protein
CLIP	Class II-associated invariant chain peptide
CML	Chronic myelogenous leukemia
CNA	Copy number alteration
CNS	Central nervous system
CNV	Copy number variant
COSMIC	Catalogue of Somatic Mutations in Cancer
CpG	5'—C—phosphate—G—3'
C-Pro	C-terminal segment of the Pro-rich region
CRC	Colorectal cancer
CTLA4	Cytotoxic T-Lymphocyte Associated Protein 4
CWH	Cardiomyopathy and woolly haircoat syndrome
CYP27B1	Cytochrome P450 Family 27 Subfamily B Member 1
DECIPHER	DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources
DISP1	Dispatched RND Transporter Family Member 1
DNA	Deoxyribonucleic acid
DTPC	2-Diazo-3,3,3-trifluropropionyl chloride
EDTA	Ethylenediaminetetraacetic acid
EEG	Electroencephalography
EGFR	Epidermal growth factor receptor
EMT	Epithelial-to-mesenchymal transition
ER	Estrogen receptor
ES	Embryonic stem
EUCOMM	European Conditional Mouse Mutagenesis Program
EUMMCR	European Mouse Mutant Cell Repository
ExAC	Exome Aggregation Consortium
FA	Folic acid
FANCA	Fanconi Anemia Complementation Group A
FATHMM	Functional Analysis Through Hidden Markov Models
FBS	Fetal bovine serum
FBXO28	F-Box Protein 28
FC	Fold change
FHIT	Fragile histidine triad
G	Guanidine
GABA	Gamma-aminobutyric acid
Gadd45a	Growth Arrest And DNA Damage Inducible Alpha
GATK	Genome Analysis Toolkit
Gbp2b	Guanylate binding protein 2b
Gbp3	Guanylate binding protein 3
GDC	Genomic Data Commons
GLUT1	Glucose transporter 1
GO	Gene Ontology
GP	Globus pallidus
GRHL2	Grainyhead Like Transcription Factor 2
GTEX	Genotype Tissue Expression project

GUI	Graphical user interface
GWAS	Genome-wide association study
Gzma	Granzyme A
Gzmb	Granzyme B
HCV	Hepatitis C virus
HIF1A	Hypoxia-inducible factor 1α
HK1	Hexokinase 1
HMGD	Human Gene Mutation Database
Hmgu	Helmholtz Zentrum München
HREM	High-resolution episcopic microscopy
HSC	Haematopoietic stem cell
IARC	International Agency for Research on Cancer
ICGC	International Cancer Genome Consortium
ID	Intellectual disability
IDH	Isocitrate dehydrogenase 1
IFIH1	Interferon Induced With Helicase C Domain 1
Ifng	Interferon gamma
IFNGR	Interferon gamma receptor
IgG	Immunoglobulin class G
Igh	Immunoglobulin heavy chain
Igtp	Interferon gamma induced GTPase
Iigp1	Interferon inducible GTPase 1
IKMC	International Mouse Knockout Consortium
IMPC	International Mouse Phenotyping Consortium
Indels	Insertions and deletions
Irf4	Interferon regulatory factor 4
JCV	John Cunningham Virus
KEAP1	Kelch-like ECH-associated protein 1
LacZ	β-galactosidase
LATS	Large Tumor Suppressor, Drosophila Homolog
LCMV	Lymphocytic choriomeningitis virus
LD	Linkage disequilibrium
LDLR	Low density lipoprotein receptor
LEP1	Leptin 1
LEPR	Leptin Receptor
LMD	Laser microdissection
LoF	Loss of function
LoH	Loss of heterozygosity
LV	Lateral ventricle
Mbp	Mega base pairs
mCNV	Multi-allelic CNV
MEF2C	Myocyte Enhancer Factor 2C
MEK1	Mitogen-activated protein kinase kinase 1
microCT	Micro-computed tomography
MPS	Mutant phenotype score
MRI	Magnetic resonance imaging

Msh2	MutS protein homolog 2
NAHR	Non-allelic homologous recombination
NCI	National Cancer Institute
NDD	Neurodevelopmental disorder
NF	Neurofibromatosis
NF-kB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NHS	National Health Service
Nkx3.1	NK3 Homeobox 1
NMDA	N-methyl-D-aspartate
NPC	Neural progenitor cell
Npm1	Nucleophosmin 1
N-Pro	N-terminal segment of the Pro-rich region
NRF2	Nuclear factor (erythroid-derived 2)-like 2
NSCLC	Non-Small Cell Lung Cancer
NTC	Neural tube closure
NTD	Neural tube defect
NUMT	Nuclear mitochondrial insertion
PAR3	Partitioning defective 3 homolog
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate-buffered saline
PCP	Planar cell polarity
PD1	Programmed cell death protein 1
PDZ	Post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-1 protein (zo-1) domain
PGK	Phosphoglycerate kinase
Pla2g2a	Phospholipase A2 Group IIA
PML	Progressive multifocal leukoencephalopathy
POT1	Protection Of Telomeres 1
PP1	Protein phosphatase 1
Pro	Proline
Pten	Phosphatase and tensin homolog
PTM	Post-translational modification
RaDAR	RanGDP/Ankyrin-Repeat complex-mediated nuclear import pathway
RAS	Rat sarcoma
RB1	Retinoblastoma 1
RBC	Red blood cell
RET	Ret Proto-Oncogene
REVEALER	Repeated evaluation of variables conditional entropy and redundancy
RNA	Ribonucleic acid
ROCK	Rho-associated protein kinase
SAM (domain)	Sterile alpha motif

SAM (molecule)	S-adenosyl-methionine
SCC	Squamous cell carcinoma
SCRIB	Protein scribble homolog
SD	Standard deviation
SELL	L-selectin
SHANK	SH3 and multiple ankyrin repeat domains
SINE-R	Short interspersed nucleotide elements of retroviral origin
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SOD1	Superoxide dismutase 1
SRO	Smallest region of overlap
STAT1	Signal transducer and activator of transcription 1
SV	Structural variant
SVA	SINE–VNTR–Alu elements
T2D	Type 2 diabetes
TAZ	Tafazzin
TCGA	The Cancer Genome Atlas
TCR	T-cell receptor
TF	Transcription factor
T_{FH}	Follicular helper T cell
T_{FC}	Follicular cytotoxic T cell
Thy1.1	Thy-1 cell surface antigen, variant 1
ToF	Tetralogy of Fallot
TSG	Tumour suppressor gene
TSS	Transcription start site
TTN	Titin
UCSC	University of California at Santa Cruz
UTR	Untranslated region
VANGL1	Vang-Like 1 (Van Gogh, Drosophila)
VHL	Von Hippel-Lindau
VIPER	Virtual inference of aberrant protein activity
VNTR	Variable number tandem repeat
VSD	Ventricular septal defect
wa3	Waved 3
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
woe2	Waved with open eyelids 2
WT1	Wilms tumour 1
YAP	Yes-associated protein 1
Zeb1	Zinc Finger E-Box Binding Homeobox 1
Zhx2	Zinc-fingers and homeoboxes 2

CHAPTER 1: INTRODUCTION

1.1 Genetic architectures of human diseases

Genetic architecture comprises the relative contributions of alleles with varying frequency and effect on disease risk. Genetic architecture is a key part of the aetiology of diseases with a genetic component, and its understanding directly affects clinical management of the disease. For example, monogenic disorders with a dominant pattern of inheritance can be prevented by prenatal screening for the presence of the dominant allele. In contrast, developmental diseases caused by a dominant *de novo* germline mutation have a low chance of recurrence within the same family. In addition to aetiological implications, knowledge of the loci causally linked with the disease can lead to a molecular understanding of mechanisms underlying the disease, and development of targeted treatments.

1.1.1 Pathogenic variants in human disease

The origins of genetics as a scientific discipline are usually traced to Gregor Mendel who discovered the rules of heredity based on his observation of trait characteristics in pea plants. Dominant traits, such as yellow pea colour, can be distinguished from recessive traits, such as green pea colour, based on the predictable distribution of traits among individual plants or organisms within subsequent generations of progeny. These principles, published in 1866, still form the backbone of a conceptual framework for understanding the genotype-phenotype relationship today.

Humans as a diploid organism show multiple traits that conform perfectly or near-perfectly to the dominant and recessive models. Autosomal dominant diseases such as familial hypercholesterolemia, autosomal dominant polycystic

kidney disease or type I neurofibromatosis have a 50% chance of transmitting to offspring if a single parent carries a single copy of the pathogenic allele. Within medical genetics, variants underlying dominant Mendelian disorders are now understood to represent the upper bound of a continuous spectrum of effect sizes of disease-causing variants (**Figure 1.1**).

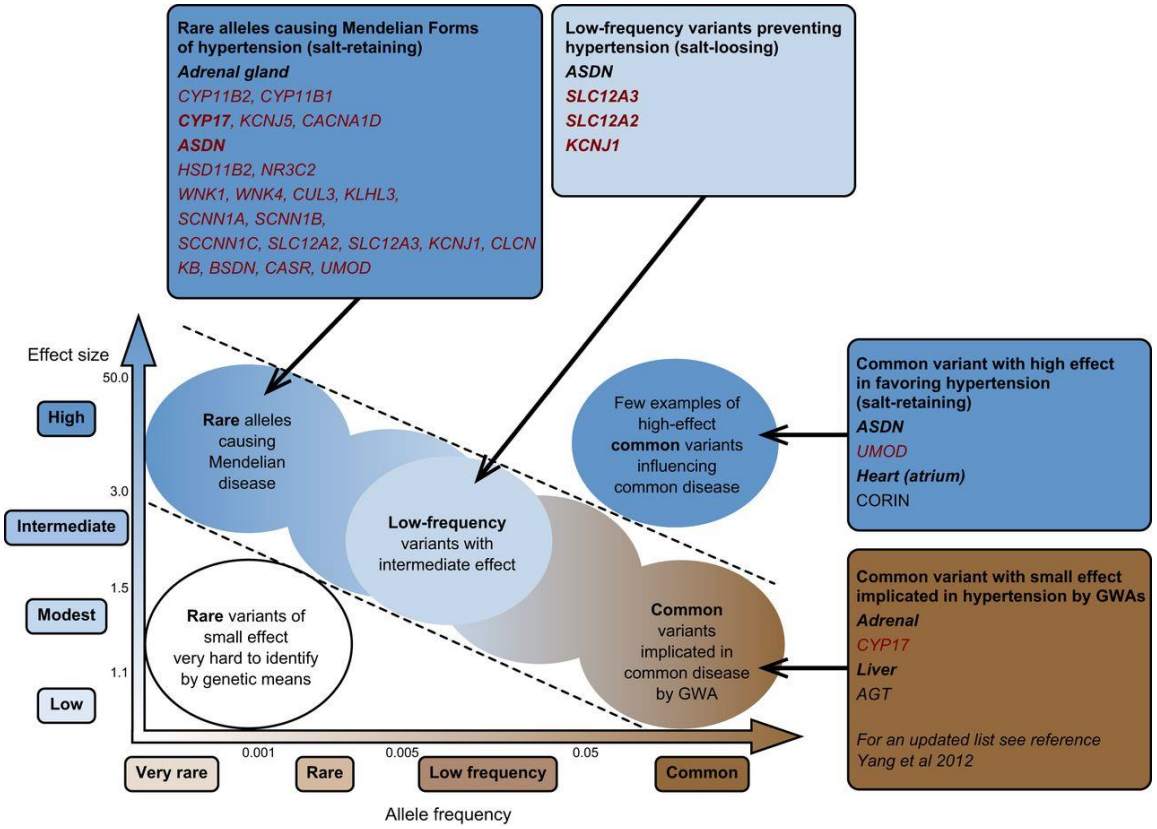


Figure 1.1 Genetic architecture of hypertension. Reproduced from Rossier at al. (1) with permission.

Unlike Mendelian disorders, many prevalent diseases do not show predictable segregation in families. Sporadic diseases and disease-associated traits such as hypertension, coronary artery disease or obesity have been shown to associate with genetic variants which modulate risk by a relatively small amount, typically <20% and often significantly less (2-6). These variants are however highly prevalent in the population, with allele frequencies typically ranging

from 5-90%. Together, dominant Mendelian and common low-risk variants form the opposite bounds of the genetic architectures of human diseases (**Figure 1.1**).

However, the binary classification of pathogenic variants as rare and dominant versus common and low-risk does not translate to a corresponding binary clustering of genetic architectures. Common genetic variation currently explains a small proportion of the heritable disease risk of most human diseases (7). Other diseases manifest as severe developmental phenotypes in the absence of any family history of the disease. Diseases such as type 2 diabetes, schizophrenia, autism spectrum disorder or rheumatoid arthritis affect a large number of individuals, and contain a genetic component in their aetiology, but possess a genetic architecture that has remained largely hidden from genetic studies. These observations demonstrate, firstly, that genetic architectures of most human diseases likely contain contributions from both common and rare variants, and secondly, that variants of other effect size-versus-frequency classes are likely to contribute, beyond Mendelian and common/small-effect variation (8).

1.1.2 Mendelian variants enable functional understanding of disease

Dominant Mendelian diseases have fuelled our understanding of the physiological and molecular aetiology of many disorders, including diseases in which the Mendelian form accounts for a small percentage of cases. This is because dominant alleles allow for a causal link to be established between a mutation in a single gene and the occurrence of the disease phenotype in all mutation carriers. For example, Mendelian forms of hypertension have converged on a narrow set of genes which control salt management in the kidney (**Figure 1.1**) (9, 10). Moreover, most Mendelian variants are nonsynonymous and thus

enable a direct functional analysis of the mutant/truncated protein. In contrast, genome-wide association studies (GWAS) identify disease loci which are incompletely penetrant, usually located in non-coding regions, and might not be directly involved in the pathogenesis of the disease by virtue of linkage disequilibrium.

Because of their defined pattern of inheritance, Mendelian diseases are reliably identified by linkage analysis which remains an important tool in the identification of causal genes in Mendelian disorders (11). An appreciable proportion of severe congenital diseases show a Mendelian pattern of inheritance. Neurofibromatosis type I (NF1), caused by mutations in the *NF1* tumour suppressor gene, is characterised by predisposition to tumours or neural origin, but also by behavioural, cognitive, motor, bone, cardiac and pigmentary abnormalities (12). The prevalence of NF1 is estimated at 1 in 3,000 births worldwide. This disease also represents a genetic overlap between cancer and developmental disorders which occurs with other familial cancer susceptibility syndromes (13). Other examples include Ataxia-telangiectasia (autosomal recessive, mutations in *ATM*), associated with neurological, immune and cancer phenotypes; Fanconi anaemia (typically autosomal recessive; most commonly caused by mutations in *FANCA*; *FANCC* or *FANCG*), characterised by bone marrow failure, irregular skin colouring, infertility and an increased risk of multiple cancer types; or Rothmund-Thomson syndrome (autosomal recessive, mutations in *RECQL4*), associated with skin abnormalities, reduced growth, vision abnormalities and susceptibility to bone and skin cancer (13). The observation that many of these syndromes are caused by mutations in DNA repair genes highlights the importance of DNA repair in both human development and cancer prevention.

1.1.3 The landscape of human functional genomic variation

1.1.3.1 Exonic variation

Until recently, the degree and frequency of variation at different positions in the human genome remained poorly characterised. The release of the Exome Aggregation Consortium (ExAC) catalogue significantly enhanced the degree of characterisation of exonic (and to some extent splice region) variation in the germline of the human population. ExAC pooled a total of 60,706 high-quality whole-exome sequences and analysed these using a common pipeline (14). Interestingly, 99% of the variants observed in ExAC showed a frequency <1% and 54% were singletons, i.e. observed only in a single ExAC individual. Encouragingly, 63.1% of all the possible synonymous CpG transitions within regions well covered by exome sequencing were observed in ExAC, although only 7.5% of synonymous variants overall were seen. This suggests that at least some classes of genomic variation (in this case the frequently mutating CpG nucleotides) are approaching saturation at this sample size. An interesting finding of ExAC was the observation of recurrence of rare variants in independent unrelated individuals. Genetic drift dictates that most rare variants are destined to disappear from the germline, however ExAC data suggest that these variants recur independently; for example, 43% of synonymous *de novo* mutations in an external set of 1,756 trios (15, 16) were also observed in ExAC participants.

Importantly, approximately 3,020 genes were classified as intolerant to monoallelic loss-of-function (LoF) variants, suggestive of haploinsufficiency. This number does not include genes whose haploinsufficiency causes milder and/or adult-onset phenotypes such as *BRCA1*. Therefore, damaging variants in a substantial fraction of human genes are under significant constraint. Conversely,

many genes were tolerant to LoF variants which rose to appreciable frequencies in some populations, for example 1-5% in Finland. Related to this, haploinsufficiency was significantly more prevalent on the X chromosome compared to autosomes, suggesting that wild-type alleles of autosomal genes play an important role in compensating for the effects of monoallelic LoF variants. Provocatively, the authors suggest that given the mutational rates of ca. $1x10^{-9}$, the vast majority of non-lethal single-nucleotide variants probably exist in at least one living human. If so, by sequencing every human, it might be possible to map the complete genomic landscape of variant constraint and thus identify all genes and genomic regions necessary for development. However, given the explosive population rate, many damaging variants have undergone limited purifying selection (17-19). The increasing rate of population growth might also explain the very high percentage of singleton variants which includes nonsynonymous variants that should be subject to minimal purifying selection.

1.1.3.2 Coding versus non-coding variation

Genomic territory outside the protein-coding regions was probed by the 1000 Genomes Project and more recently the UK10K consortium which performed whole-genome sequencing (WGS) of 3,781 healthy individuals and whole-exome sequencing (WES) of approximately 6,000 individuals with rare disease, severe obesity or neurodevelopmental disorders (20). Contrary to expectation, the study showed no appreciable abundance of low-frequency (MAF 1-5%) variants of large effect (β>1), at least in relation with the 31 traits examined. In contrast, focusing on rare variants (MAF <1%), sequence kernel association tests and burden tests

revealed a significant signal in 9, 70 and 196 genes examining LoF, functional (missense and LoF) and naïve (all exonic, untranslated region (UTR) and essential splice site) variants, respectively. Of these, LoF variants in *APOB* achieved experiment-wide significance and other genes were suggestive of real signal. This demonstrates that rare variants are likely to play an important role in modulating phenotypic traits in disease-free individuals, and suggests that purifying selection has reduced most penetrant exonic variants to low frequency. An alternative explanation is that rare variants acquired during recent explosive population growth are enriched in functional variants that have been subject to limited selection but are rare as a result of their recent origin.

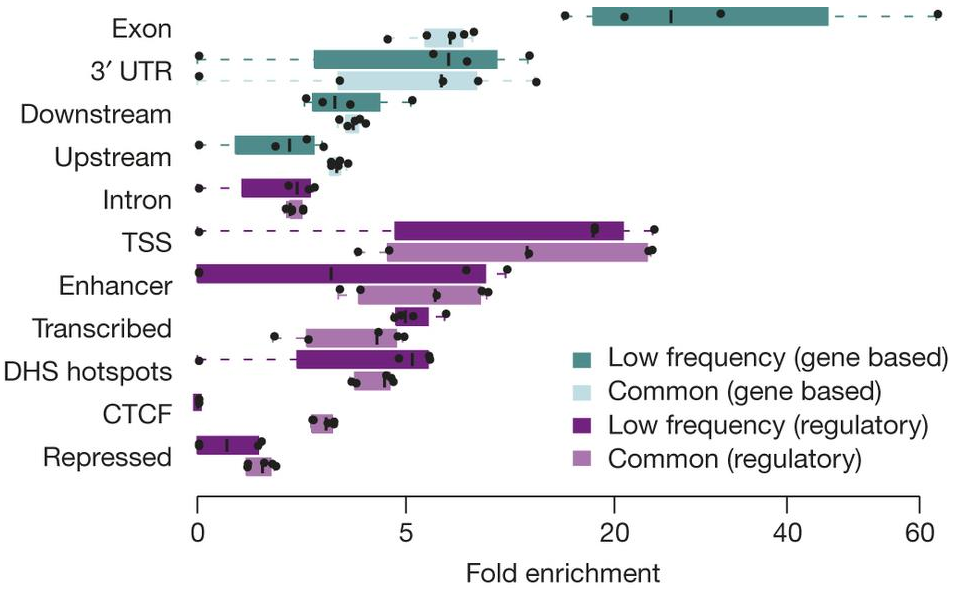


Figure 1.2 Distribution of fold enrichment statistics for single-variant associations with 5 lipid traits. Low frequency (MAF 1-5%) vs common (MAF >5%). Reproduced from (20) under the Creative Commons license.

In a more focused association test with five lipid measures, low-frequency exonic variants showed the strongest enrichment for association (**Figure 1.2**). This confirms the notion that exonic variants, in particular rare and low-frequency variants, are most likely to be phenotypically functional. Interestingly, in contrast to

exonic variants, non-coding variants in the 3' UTR, transcription start site (TSS) and some further non-coding regulatory regions showed a significant enrichment that did not differ between common and low-frequency variants (**Figure 1.2**). Given that these low-frequency variants are present in 1-5% of the population, more frequently than the rare variants discussed above, the observed bias in enrichment of low-frequency variants specific to exonic versus non-coding variants is most likely due to purifying selection of nonsynonymous alleles, independently of their association with lipid measures.

A discovery of 13 associated markers in autism spectrum disorder (ASD) was reported by UK10K; however these only achieved statistical significance after patient samples were pooled with an autism consortium cohort, demonstrating that discovering pathogenic rare variation in some neurodevelopmental diseases (NDDs) requires particularly large sample sizes. Unfortunately, the UK10K consortium have not released data on burden tests in the UK10K-exomes part of the study, therefore the importance of a polygenic burden of rare variants in the diseases examined is uncertain.

1.1.3.3 Structural variation

Structural variation accounts for the majority of varying nucleotides among human genomes (21). Following earlier efforts (22-24), the most comprehensive insight into naturally occurring structural genomic variants in humans to date came from the 1000 Genomes Project (25, 26). The catalogue of structural variants identified in the project comprises 42,279 biallelic deletions, 6,025 biallelic duplications, 2,929 multi-allelic CNVs (mCNVs), 786 inversions, 168 nuclear

mitochondrial insertions (NUMTs), and 16,631 mobile element insertions (MEIs, including 12,748; 3,048 and 835 insertions of Alu, L1 and SVA (SINE-R, VNTR and Alu composite) elements, respectively (26). A combination of long-read Pacific Bioscience sequencing with short-read Illumina sequencing allowed the authors to improve breakpoint resolution with a mean boundary precision of 0-15 bp with the exception of inversions and duplications (32 bp and 683 bp, respectively).

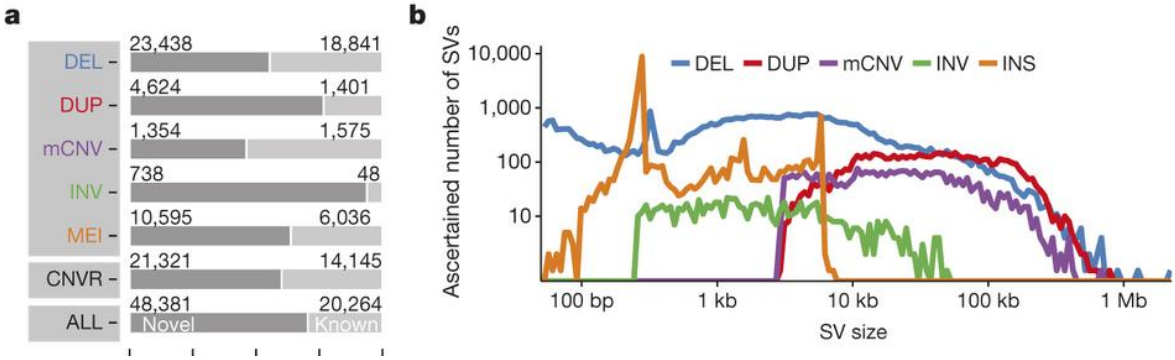


Figure 1.3 Abundance and size distribution of SVs identified by the 1000 Genomes Project. Reproduced from Sudmant et al. (26) under the Creative Commons license.

The majority of SVs occur at a low frequency: 65% SVs have a variant allele frequency <0.2%. Interestingly, biallelic duplications were significantly depleted among common SVs; duplications were instead dominated by multi-allelic CNVs, suggesting duplications are under-ascertained in current tag-SNP-based studies. These SVs could represent an important source of genetic variation underlying the missing heritability in complex diseases.

As expected, deletions are depleted in functional regions such as coding sequences, UTRs or transcription factor binding sites. SVs also seem to have a higher relative impact on gene expression than SNPs, probably owing to their affecting a larger sequence. In fact, a number of previous GWAS associations were shown to be in strong linkage disequilibrium with SVs, suggesting some significant GWAS signal might be attributable to SVs. Complex events

characterised by multiple breakpoints were also observed and ascertained by long-read sequencing. Strikingly, long-read characterisation of inversions revealed that 80% were complex, multi-breakpoint inversions; the majority of which were inverted duplications. This points to the importance of other mutational processes in addition to non-allelic homologous recombination, retrotransposition, and non-homologous end-joining.

Overall, the median SV density is estimated at 18.4 Mbp of SVs per genome, dominated by mCNVs (11.3 Mbp) and biallelic deletions (5.6 Mbp). This estimate excludes low-complexity regions and regions with poor coverage by Illumina reads, which can harbour additional SVs. Interestingly, somatic SVs also occur in humans, for example retrotransposition events in neurons of normal brains which result in mosaicism (27, 28).

1.1.4 Structural variation in disease

Large CNVs were initially discovered by cytogenetics in patients with Down syndrome (formerly called mongolism) and intellectual disability (29-31). Since then, CNVs have been causally implicated in multiple disorders, most notably neuropsychiatric disorders including autism spectrum disorder (32, 33), schizophrenia (34, 35), bipolar disorder (36, 37), and developmental diseases featuring intellectual disability (38-41). Both *de novo* and germline CNVs are overrepresented among neuropsychiatric and developmental disease patients (32, 36). Risk-associated germline CNVs were also found to affect cognition in disease-free carriers (42). Importantly, the penetrance is estimated to range from 10-100% and so disease-associated CNVs do not always reliably co-segregate with the disease (43). Moreover, many CNVs are undetectable by traditional

cytogenetics, necessitating the use of higher-resolution genomic analysis in ascertaining patient CNVs (44).

These studies led to an intimate understanding of the functional genetics of certain forms of NDDs, for example those associated with *SHANK3* CNVs which have been characterised in mouse models as well as patient-derived neuronal cultures (45, 46). Restoration of *Shank3* expression has even been shown to reverse some autistic-like phenotypes, suggesting the identification of pathogenic CNVs can potentially lead to new therapy (47). Moreover, CNV studies firmly implicated certain functional networks of genes in the aetiology of NDDs, notably genes involved in transcription (48), chromatin remodelling (48-50) and synaptic function (16, 48, 51).

A recent analysis of CNVs in ExAC yielded 126,771 CNVs among 59,898 individuals, with 2.1 high-confidence CNV intersecting at least one protein-coding gene observed per individual (52). CNV constraint, estimated by deviation from a linear regression model of CNV rate, significantly correlated with constraint of LoF and missense SNVs. Importantly, highly expressed genes had higher CNV intolerance scores in most tissues, with genes highly expressed in the brain being the most intolerant to CNVs. The 5% most CNV-intolerant genes were significantly enriched in the GO term neuron/axon development, which also comprised the largest enriched gene network. These results provide a rationale for the strong association of CNVs with neurodevelopmental diseases: brain-expressed genes are more dosage-sensitive than genes in other tissues and this effect might be particularly striking during brain development. Accordingly, analysing CNVs within a schizophrenia study showed that schizophrenia-associated CNVs were enriched in genes with higher constraint based on missense and LoF variants.

Given the strong association of CNVs with neuropsychiatric diseases and developmental delay, analysis of chromosomal abnormalities by chromosomal microarray has become the first-tier diagnostic test in patients with developmental disabilities or congenital anomalies (53).

1.1.5 Filling the gaps in the genomic catalogue of human disease

Despite the successes of linkage analysis, genome-wide association studies, next-generation sequencing, haplotype-resolved genome assemblies, and others, the diagnostic yield in many diseases with a strong genetic component remains under 50%. This problem is partly caused by our incomplete knowledge of the genetic architectures of human diseases, exemplified by type 2 diabetes: a recent large sequencing study of a total of 15,597 individuals and an associated genotyping and imputation-powered analysis of a further 111,548 patients and controls identified 126 variants commonly associated with T2D risk and concluded that common variation contributes much more to T2D heritability than rare ones do (54). Unfortunately, currently known common variation accounts for only about 10% of the genetic risk, and the authors concluded that nearly all common-variant associations detectable by whole-genome sequencing were previously found by GWAS. Therefore, it remains unclear which genetic factors determine the remaining and dominant fraction of the heritable risk of T2D.

Other, less robustly characterised diseases, also remain challenging to reliably genetically diagnose. A recent study, published as part of the WGS500 program to sequence 500 patient genomes of patients with a diverse genetic disorders, identified pathogenic variants in 21% of cases, despite taking multiple

steps to improve accuracy of variant calling, filtering and annotation across the pooled patient set (55). Among the challenges highlighted by the authors were the observations that multiple genetic factors likely contribute, and only very large patient cohorts are sufficiently powered for the identification of new variants with modest contribution. In addition, the authors noted that ultimate assignment of pathogenicity relies on the identification of the same variant in multiple patients for whom detailed phenotypic and clinical data are available. Clearly, in rare diseases, rare variants of incomplete penetrance will be very challenging to match, hampering the definitive proof of pathogenicity in many cases. Accordingly, Cooper at al. noted that '…the primary roadblock faced by the field is increasingly one of variant interpretation, rather than data acquisition' (56). Therefore, with the decreasing cost of WGS, the number of patients ascertained and characterised in the public domain will likely become the major limiting factor to improving diagnostic yield.

Neural tube defects (NTDs), the second most frequent group of birth defects, are estimated to contain a strong genetic component to their aetiology, yet the genetic architecture remains incompletely characterised. In particular, there is a vast gap between the list of genes causing NTDs when mutated in mice, and those genes that have been linked with NTDs in humans (57, 58). The genetic architecture of NTDs is poorly understood.

Neurodevelopmental diseases such as autism spectrum disorders (ASD) and schizophrenia are also proving challenging to map genetically, owing to the significant heterogeneity of disease-associated genes. For example, all high-confidence pathogenic variants in ASD identified thus far, including SNVs, CNVs and syndromic causes, explain only a small fraction (<10%) of the genetic

contribution to ASD risk (59). Hence, NDDs are considered polygenic diseases with monogenic form accounting for a small percentage of cases. Mapping the remaining genetic architecture remains an important goal in NDD genetics (60).

Cancer has been the focus of intense genomic investigations, as all or virtually all cases of cancer arise from acquired genomic aberrations. The genetic architecture of cancer comprises germline and somatic components. Germline alleles influence cancer risk with a continuum of effect sizes (**Figure 1.4**) (13, 61), but further somatic mutations are required to develop malignancy. The somatic component is discussed in more detail in section 1.3.

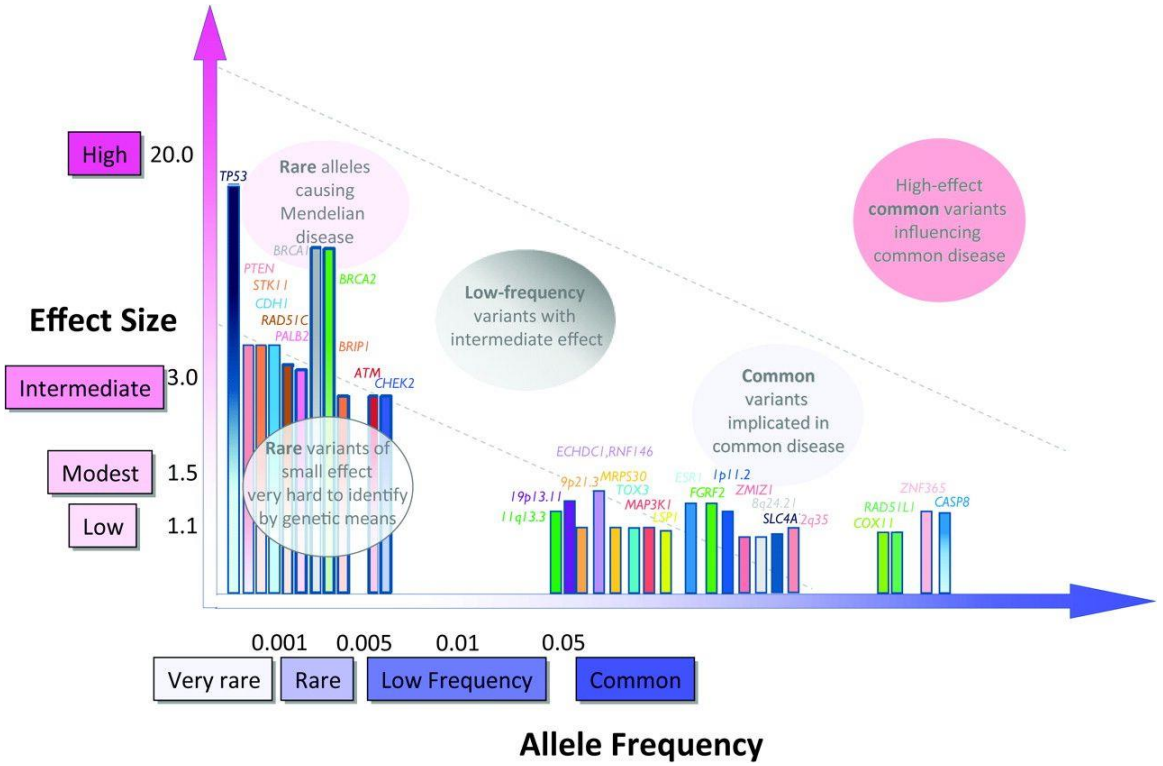


Figure 1.4 Genetic architecture of breast cancer risk. Reproduced from Hindorff et al. (61) with permission.

Overall, future genetic work needs to meet three major challenges: 1) characterise all pathogenic variants of known classes (common low-risk alleles, rare large-risk alleles) amenable to existing genetic methods; 2) improve the

analytical tools for genetic variation outside the scope of current studies (e.g. rare variants of low effect; CNVs in low-complexity regions; mosaic somatic mutations); and 3) characterise their contribution to the genetic architectures of diseases, and identify pathogenic variants of those difficult-to-probe classes.

1.1.6 Discovery and assessment of candidate variants

Given the incomplete knowledge of the genetic architectures of human diseases, the identity of pathogenic variants, and uncertainty surrounding the relevance of many variants identified by WES and WGS, it is necessary to continue the process of discovery and assessment of candidate pathogenic variants across diseases. However, in addition to limited sensitivity, a problem faced by genome-wide variant searches is the rate of false positives. Indeed, The exome aggregation consortium (ExAC) found that on average, an ExAC participant harbours ~54 variants reported as disease-causing in the Human Gene Mutation Database (HMGD; http://www.hgmd.cf.ac.uk/ac/index.php) or ClinVar (14). A manual curation of 192 such variants with an ExAC allele frequency >1% found good evidence of pathogenicity for only 9 variants, and led to the re-classification of some of the remaining variants. This suggest the medical genetic literature is burdened with a relatively high rate of false positives. In order to improve the sensitivity and specificity of identification of pathogenicity, a common assessment framework is required that allows researchers and clinicians to understand the level of confidence surrounding the implication of a specific variant in a specific disease.

A working group of the US National Human Genome Research Institute have published guidelines for investigating causality of sequence variants in

human disease (62). The guidelines highlight a primary role for genetic evidence in supporting causality, noting that many protein-damaging variants are present in human genomes. At the gene level, a burden of rare or *de novo* variants in patients compared to a matched control cohort provides good evidence of gene involvement. Importantly, ethnicity and sequencing depth must be well matched between cases and controls, as rare variants are particularly sensitive to these types of bias.

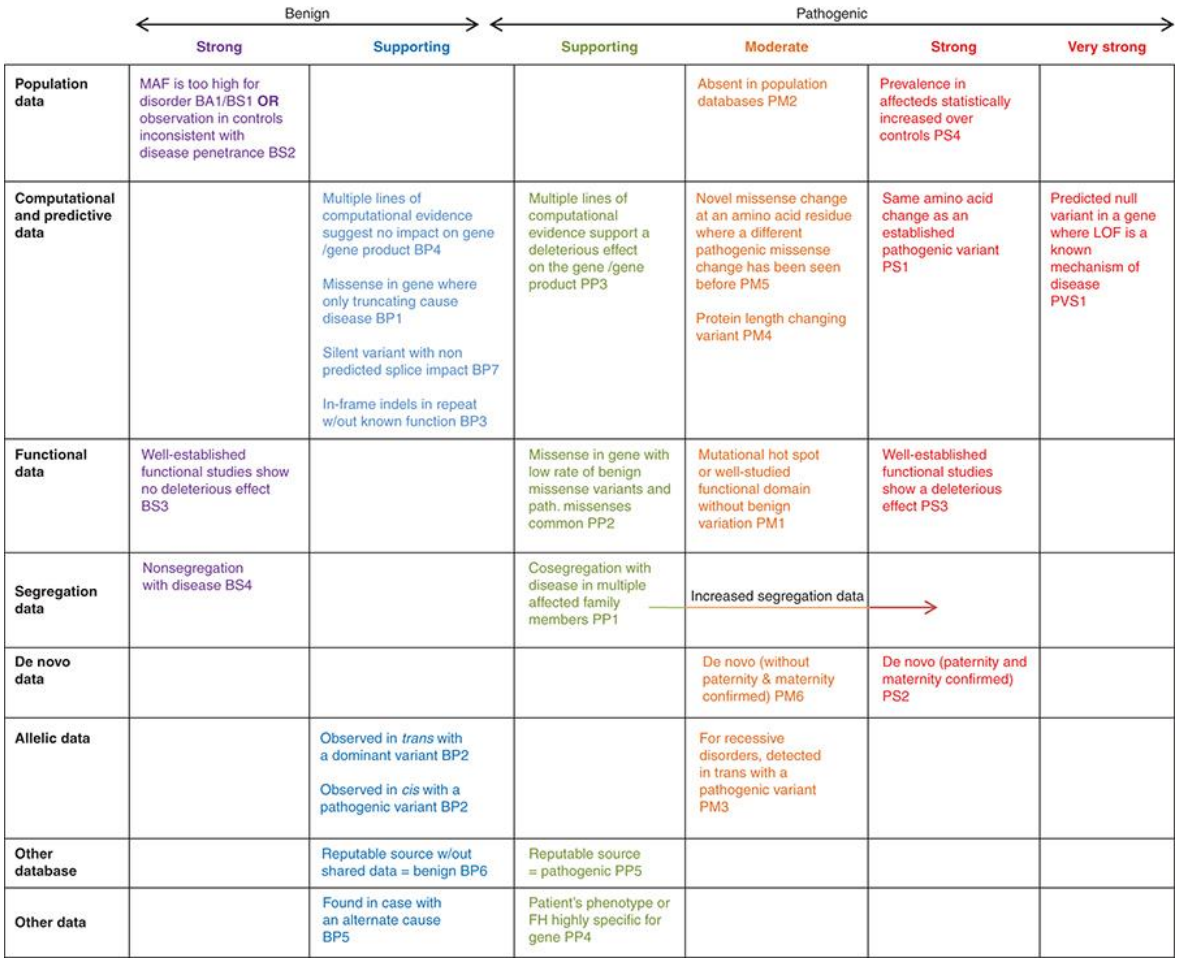
At the variant level, single-variant enrichment in cases over controls provides convincing evidence of disease association, however most disease cohorts are underpowered to detect genome-wide significant associations for rare variants (63). Co-segregation of the variant with disease in multiple family members or in multiple families is also considered strong genetic evidence. However, an important caveat is that variants with moderate penetrance will not always co-segregate with disease. In fact, Bodmer and Bonilla specifically noted that ‘neither common nor rare variants are familial’, demonstrating in an example that for a susceptibility allele *D* with a penetrance of 20% (corresponding to an odds ratio of 3 or higher), heterozygous to wild type matings *Dd* x *dd* will result in only 5.2% families with four offspring having more than one affected offspring (64). A penetrance approaching 50% is necessary to start observing Mendelian-like familial segregation. Therefore, co-segregation evidence in a small number of families will only be available for variants with a relatively high penetrance.

Finally, important evidence is provided by population frequency: zero or low frequency of the variant in large population cohorts with ancestry similar to the disease cohort supports the potential pathogenicity of the variant. However singleton rare variants are observed in both patients and healthy individuals, as

robustly demonstrated by ExAC. Taken together, these observations show that genetic evidence for disease association of single variants is likely to be limited to high-penetrance rare variants in all but very large studies. Low- to moderate-penetrance rare variants can thus be uncovered by gene-level or gene set-level analysis.

Guidelines for variant interpretation have been recently formalised by the American College of Medical Genetics and Genomics (ACMG), and the Association for Molecular Pathology (65). These guidelines provide a useful framework for the classification and quantification of evidence surrounding pathogenicity of sequence variants across diseases (**Figure 1.5** and (**Appendix Tables S1-S3**)). The quantitative scale classifies variants as benign, likely benign, pathogenic, likely pathogenic, or of uncertain significance. In this thesis, the guidelines will be utilised to describe the evidence supporting the pathogenicity of the variants investigated in the following chapters.

ACMG guidelines also exist for the interpretation and reporting of germline CNVs (66). Although lacking a quantitative framework, the guidelines provide a useful reference for weighing the relative strength of different types of evidence in classifying CNVs.



	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Figure 1.5 Evidence framework from ACMG Standards and guidelines on variant interpretation. Reproduced from Richards et al (65) with permission.

1.1.7 The importance of rare variants in disease

Earlier studies of cancer and complex disease susceptibility predicted that rare and low-frequency variants would be important in completing the full picture of the genetic architecture (6, 61, 64). As discussed above, rare variants are now known to correlate with phenotypic traits in healthy individuals, and significantly associate with developmental diseases. Many significant associations between rare variants have been identified to date, including a burden of *SOD1* variants in sporadic amyotrophic lateral sclerosis (67), *CYP27B1* variants in multiple sclerosis (68), variants in 7 neuronal genes in bipolar disorder (69), RAS pathway gene

variant in Noonan syndrome (70), *IFIH1* rare variants in type 1 diabetes (71), polygenic burden of disruptive variants in schizophrenia (72), non-syndromic heart defects (73), colorectal adenomas (74), lung cancer risk (75), but also variants in complex adult diseases such as type 2 diabetes (76), irritable bowel syndrome (77) or hypertriglyceridemia (78). Rare SNVs and CNVs in various genes have also been linked with neural tube defects (79-82).

Based on the numerous studies that succeeded in identifying clinically relevant rare variants, and the extent of missing heritability in complex diseases once common variants have been accounted for, the contribution of rare variants to human disease is likely to be important. These variants represent an important class of genetic variation separating low-risk common variants identified by GWAS, and dominant variants with a Mendelian inheritance pattern. Exonic rare variants are particularly likely to exert significant effects. Approximately 88% of trait-associated variants of weak effect are non-coding (6). The majority of coding variants are rare (<1%) and nonsynonymous SNPs have the highest odds ratios as a class in GWAS (6). Low-frequency (1-5%) exonic variants are significantly more enriched in trait-associated loci than frequent exonic variants, and represent the most highly enriched class of variation in lipid level-associated variants (20). The ratio of non-coding to coding causal variation declines relative to increasing mutational penetrance and disease severity (83). In a recent WGS study of breast cancer, only 2 recurrent somatically mutated non-coding loci with suspected driver function were identified, in contrast to many coding driver mutations (84). In addition, due to the recent explosive human population growth, human coding variants are over-represented in deleterious variants compared with other species due to limited purifying selection (17, 19). Taken together, rare nonsynonymous

SNVs and genic CNVs have a high *a priori* probability of non-neutral effect on disease phenotypes.

Despite the strong rationale for the importance of rare variants in disease, the majority of causal rare variants still elude detection in most scenarios. This is because most genome-wide studies are under-powered to detect their association with disease, and small-scale sequencing studies are usually inconclusive on their importance because of incomplete penetrance and low recurrence rate of the same variant in unrelated cases. Therefore, current methods of uncovering pathogenic rare variants rely on burden tests, which identify gene-level and gene set-level associations, or on targeted hypothesis-driven sequencing (64, 72, 85).

In this thesis, the contribution of rare germline and *de novo* variants (both SNVs and CNVs) to developmental defects, and the contribution of rare somatic mutations to cancer are examined. We focus on the ASPP family of genes as proof of principle and discuss how the discovery methods developed throughout could be generalised genome-wide.

1.2 Congenital disorders of the central nervous system

A large number of congenital disorders affecting the central nervous system exist, ranging from lethal defects such as craniorachischisis, to mild developmental delay. In this thesis, we will focus on abnormalities associated with 1q4 CNVs, neural tube defects, childhood-onset schizophrenia, and autism spectrum disorder.

1.2.1 1q41q42 microdeletion syndrome

Structural variants in the human genome arise from a range of mutational mechanisms, including DNA recombination, replication- and repair-associated processes. CNVs and other structural abnormalities are associated with many developmental abnormalities (86, 87) which led to the adoption of the chromosomal microarray as the first-tier test for individuals with developmental disabilities or congenital anomalies, offering a diagnostic yield over 15% (53). A major source of CNVs is nonallelic homologous recombination (NAHR), which occurs between two paralogous low-copy repeats or segmental duplications (88, 89). Recurrent CNVs at these loci can be pathogenic, especially when spanning gene-rich regions (90).

The composition of non-recurrent rearrangements is more complex and includes compound mechanisms such as duplication-inverted triplication-duplication (87, 91). Breakpoint junctions in non-recurrent rearrangements generally vary in each individual, making the clinical interpretation more difficult (92). The distinct properties of non-recurrent rearrangements suggest they are

formed by a mechanism independent from NAHR, such as non-homologous end joining (NHEJ), and replication-based mechanisms such as break-induced replication, microhomology-mediated break-induced replication, serial replication slippage, and fork stalling and template switching (93, 94). Importantly, some of these repair mechanisms can introduce nucleotide changes such as point mutations and indels, concomitant with the structural variant. In addition, insertion of short <100 bp segments from nearby genomic regions is observed in up to 35% nonrecurrent structural variant junctions (95). Therefore, in addition to breakpoint heterogeneity, changes in genomic sequence associated with CNVs further increase the diversity of these structural variants, which is reflected in the phenotypic heterogeneity of CNV carriers. Moreover, these replication-based mechanisms are thought to occur during mitosis and so can contribute to somatic structural variation and mosaicism.

Chromosomal deletions that cause rare genetic disorders usually affect more than one gene and cause multiple phenotypic features. Deletions in the chromosomal 1q41-q44 region are significantly associated with central nervous system (CNS) defects including neural tube defects (NTDs), agenesis of corpus callosum, microcephaly and hydrocephalus (96, 97). Small interstitial deletions in the 1q41q42 region are implicated in the 1q41q42 microdeletion syndrome with features of severe developmental delay, intellectual disability, and brain morphological abnormalities. This syndrome was discovered by testing over 10,000 patients with developmental disabilities by chromosomal microarray and observing a subset of patients with shared features that harboured small (<10 Mb) deletions in the 1q41q42 region, otherwise undetectable by cytogenetics (44).

A critical region comprising the genes *FBXO28*, *TP53BP2*, *CAPN2* and *CAPN8* has been proposed for the 1q41q42 microdeletion syndrome based on the smallest region of overlap (SRO) (98). Virtually all 1q41q42 deletions causing the 1q41q42 microdeletion syndrome are *de novo*; a patient with a small inherited 1q41q42 microdeletion was diagnosed with an isolated congenital heart abnormality and the deletion does not span the critical region of the syndrome (98). Since then, a small inherited *DISP1* deletion (*DISP1* is located outside the SRO) has been reported in a patient with seizures that lacks most severe features of the syndrome (99). Among the genes within the SRO, *FBXO28* (encoding the F-box only protein 28, a ubiquitin ligase) has been proposed as a candidate gene responsible for seizures and intellectual disability in microdeletion patients (100, 101). However, the genes responsible for other phenotypes, in particular the brain morphological abnormalities, of the syndrome are unknown.

TP53BP2, which is located at the boundary of 1q41 and 1q42, encodes the Apoptosis Stimulating Protein of p53 2 (ASPP2). It is also an ankyrin repeat, SH3 domain and proline-rich containing protein (ASPP), originally identified as a positive regulator of p53-mediated transcription of apoptotic genes (102, 103). In mice, ASPP2, encoded by *Trp53bp2*, is essential for embryonic development (104, 105): an ASPP2-null genotype in mice is lethal prior to embryonic day E7 and mice with homozygous deletion of exon 3 exhibit loss of neuroepithelial cell polarity and hydrocephalus (106). ASPP2 is involved in regulating apoptosis through binding to p53 (and its sibling proteins p63 and p73) via its C-terminus and in regulating cell polarity through binding to the tight junction protein Par3 via its N-terminus (106, 107). Therefore, *TP53BP2* is a potential candidate gene for a causal role in the 1q41q42 microdeletion syndrome. Identification of causative

genes could extend our understanding of molecular pathways that are responsible for common diseases. In Chapter 2, the role of *TP53BP2* in the aetiology of 1q41q42 microdeletion-associated abnormalities is investigated.

1.2.2 Neural tube defects

Neural tube closure (NTC) is a key step of central nervous system (CNS) development. Neural tube defects (NTDs) stem from a failure of the neural tube, the structure that will give rise to the brain and spinal cord, to close during neurulation. NTD often results in death or severe disability and is among the most common birth defects worldwide ranging from 1 to 27 per 2000 pregnancies in the USA and China, respectively (108) (109). During neurulation, the neural plate bends upwards and eventually fuses to form a hollow tube. This process is evolutionarily conserved and occurs following a well-defined sequence of events along the embryonic axis (110). In mouse, initiation of closure starts at the future cervical-hindbrain boundary and goes bi-directionally into the hindbrain and along the spine (closure 1). Two additional closure sites occur at the midbrain-forebrain boundary (closure 2) and at the most rostral extremity of the forebrain (closure 3). NTDs are classified according to the region of the brain and spinal cord that are affected. Complete failure of closure 1 leads to craniorachischisis, the most severe NTD, in which almost the entire brain and spinal cord are open. If closure 1 remains incomplete in the cranial region, anencephaly occurs. In the spinal region, failure of closure at the posterior neuropore leads to spina bifida. The mildest form of spina bifida, myelomeningocele, is the most common NTD compatible with life.

NTDs have a complex aetiology involving both genetic and environmental factors with genetic factors contributing up to 70% (111). Over 250 transgenic

mouse models of NTD have been reported to date, spanning many classes of biological function including planar cell polarity (PCP), apoptosis, epigenetic regulation, metabolism and actin dynamics (58, 112). While homozygous disruption of many genes gives rise to NTD in mice, monogenic NTD in humans is unknown (113). A small number of common loci (114, 115) and rare variants associated with NTDs have been identified (57), limited by the small number of cases and the heterogeneity of the underlying aetiology. Only a very limited number of gene mutations were linked with similar phenotypes in both mouse and human NTDs. PCP genes *SCRIB (116) and CELSR1* (117, 118) are two examples. An important insight came recently from the identification of *de novo SHROOM3* truncating mutations in 2 independent NTD cases, highly significant given the cohort size of 43 trios (119). *Shroom3* mutant mice develop NTDs, confirming these *de novo SHROOM3* truncations as pathogenic in human NTDs. The same study identified truncating *de novo* mutation in *PAX3* in an unrelated case and a missense *de novo* mutation in *GRHL3*; mutations in the orthologous genes cause NTDs in mice.

To date, the gene mutations identified in human NTDs are mainly involved in PCP (e.g. in *VANGL1*, *VANGL2*, *CESLR1* or *SCRIB*) occur at low frequency, and account for a small proportion of cases (79, 117, 118, 120). Interestingly not all PCP genes behave the same. Although mutation in *VANGL1* gene has been detected (79) in human NTDs at a low frequency, *Vangl1*-null mice do not suffer from NTD. It is unknown whether mutations in other mouse NTD-related genes contribute to human NTDs.

Folic acid (FA) supplementation reduced the incidence of NTDs substantially. FA plays a key role in one-carbon metabolism that controls the

production of purine and thymidylate, building blocks of DNA and RNA biosynthesis as well as the production of methyl donor S-adenosyl-methionine (SAM) used in DNA, RNA and protein methylation. Thus FA may prevent NTD by affecting cell proliferation via multiple pathways including gene expression. Mutation status of FA metabolic pathways has been subject to intense study. However transgenic mice deficient in MTHFR and MTHFD1, two key enzymes of FA metabolic pathway, failed to develop NTD phenotype. Similarly *Dact1*-null embryos also do not suffer from NTD (121*).* Few functional mutations have been identified in human NTDs that involve the FA metabolic pathway. Consistent with this, around 30-50% of NTDs are not preventable by FA supplementation (122). Although polymorphisms in some genes (e.g. *GLUT1*, *HK1*, *LEP1* and *LEPR*) have also been associated with NTD risk (57), most mutation screens and case-control studies failed to find significant associations between the known polymorphisms (such as in folate genes *MTR* or *FOLR1,2,3*) and NTDs (57, 123). All these suggest that an unmet clinical need is the identification of pathogenic variants in human NTDs.

A key process in neural tube closure is apical constriction. Following convergent extension of the epithelial sheet, cells reduce the width of their apical surface while allowing the basal surface to enlarge, thus enabling the epithelial sheet to fold. Epithelial apical constriction is regulated in part by Par3 and Willin which prevent apical localisation of ROCK via regulating aPKC (124, 125). Junctional loss of aPKC is required to allow ROCK-mediated cytoskeletal changes leading to apical constriction. Mutations in apical constriction regulators *Shroom3*, *Abl* and *Mena* in mice result in well-defined NTDs comprising exencephaly and

caudally restricted spina bifida (58). Exencephaly is the mouse clinical equivalent of human anencephaly. As described above, truncating *de novo* mutations of *SHROOM3* were found in human NTDs, implicating the genetic disruption of apical constriction in the aetiology of NTDs (119, 126). The interaction of SHROOM3 and ROCK is required for neural tube morphogenesis in mice (127). Interestingly, SNP variants in the gene *PARD3* that encodes Par3, a key component of apicobasal polarity complex, are associated with anencephaly in China (128). The convergence of these risk factors on apical constriction mediated by ROCK implicates the genetic disruption of this process in the aetiology of NTDs.

Recent studies suggested that ASPP2, a binding partner of Par3 and a regulator of apical basal polarity and adherens junctions, is required for CNS development. ASPP2-deficient mice exhibit loss of neuroepithelial cell polarity and hydrocephalus (106). ASPP2 also inhibits the epithelial-to-mesenchymal transition (EMT), a process which has recently been implicated in neural tube closure downstream of GRHL2 (129, 130). Therefore, it would be interesting to investigate whether ASPP2 is disrupted in NTDs.

Interestingly, amongst mouse mutants exhibiting NTDs, a subset of p53-deficient embryos have been shown to develop NTDs, suggesting that p53-dependent transcription programs may be required during neural tube closure (131, 132). Those embryos exhibit exencephaly, consisting of an outgrowth of the neural tissue usually confined to the fore- and midbrain, but sometimes also affecting the hindbrain. Some mice with mutations in p53-target genes, such as *Gadd45a*-null mice, have been shown to develop NTDs, further suggesting that p53 transcriptional activity may be required during neural tube closure (133).

Moreover p73, a p53 sibling, is required for neural stem cell maintenance and its deficiency also causes hydrocephalus suggesting p73 might be involved in NTDs (134, 135).

Chapter 2 of this thesis investigates the role of ASPP2 deficiency in CNS abnormalities including NTDs in mice and humans. In Chapter 3, the role of rare variants in *TP53*, *TP73*, *TP53BP2* and *PARD3* in human NTDs is investigated.

1.2.3 Autism spectrum disorder and schizophrenia

Concordance rates in monozygotic versus dizygotic twins are an important source of evidence of heritability. In psychiatric disorders, twin studies estimate heritability at 0.80 for autism spectrum disorders (ASDs) (136), 0.81 for schizophrenia (137) and 0.75 for bipolar disorder (138) (**Figure 1.6**). These studies indicate a substantial genetic component in the aetiology of ASD and schizophrenia. In addition, *de novo* mutations have been shown to play a role. Taken together, understanding of the genetic basis of ASD and schizophrenia is an essential step towards a mechanistic understanding of these diseases, a pre-requisite for targeted therapy.

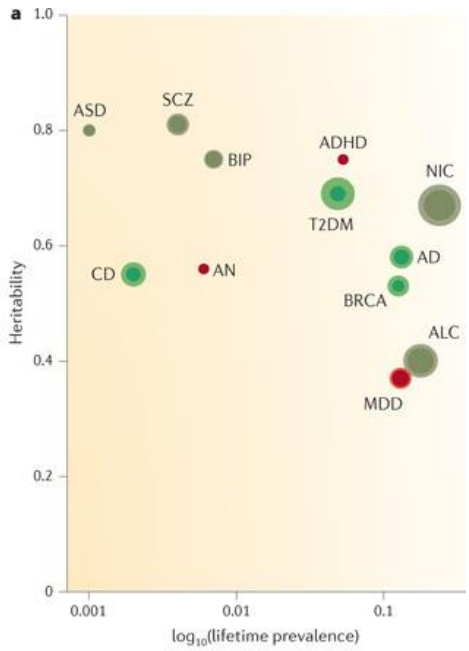


Figure 1.6 Heritability vs lifetime prevalence of psychiatric disorders. Adapted from Sullivan et al. (137) with permission.

No Mendelian forms of schizophrenia seem to be known (139). Therefore, the genetic architecture of schizophrenia comprises incompletely penetrant germline variants and *de novo* mutations. In addition to common loci, an increased burden of structural rare variants in neurodevelopmental pathways, and single-nucleotide variants in synaptic and chromatin remodeling genes have been demonstrated in schizophrenia patients (72, 140, 141). *De novo* mutations are also implicated in schizophrenia (35, 36, 142) and autism (32, 143, 144). These mutations span diverse genes, highlighting the polygenic nature of schizophrenia and ASD.

Estimates of the ASD gene target sizes are in the region of 250-1,800 genes (144). Understanding the function of all these causal genes is a great undertaking. However, some common pathways emerge, which allow us to generalise the likely modes of disruption to classes of genes. For example, synaptic signaling, chromatin remodeling, and protein translation are strongly implicated processes in the

pathophysiology of ASD and other psychiatric disorders (59). Importantly, there is a significant functional overlap between genes associated with multiple psychiatric disorders.

An interesting example of an ASD gene in *SHANK3*, a gene encoding a postsynaptic density scaffolding protein of excitatory synapses. A syndromic form of ASD called Phelan-McDermid syndrome arises from 22q13 deletions that span *SHANK3*. This syndrome represents a possible example of a monogenic form of autism. However, *SHANK3* point mutations have also been observed in both schizophrenia and ASD patients, and even in a Rett syndrome-like patient (145-148). Some of these mutations have been shown to modify dendritic spine morphology by an actin-dependent mechanism (149). SHANK3-deficient mice recapitulate some of the human phenotypes associated with *SHANK3* mutations such as deficits in social interaction (45, 150). This gene exemplifies the partial convergence of psychiatric disorders on common dysregulated gene networks and cellular processes.

1.3 Functional genomics of cancer

1.3.1 Tumour suppressor genetics

Cancer is driven by genetic dysregulation which allows cells to proliferate and migrate outside their physiological spatiotemporal bounds. The genetic architecture of cancer includes a major somatic component but also germline risk alleles of varying penetrance. Traditionally, tumours are thought to be driven by a small number (<8) of 'driver gene' mutations with the remaining somatic mutations considered 'passengers' which confer no selective growth advantage (151). However, defining driver gene mutation as a 'mutation that directly or indirectly confers a selective growth advantage to the cell in which it occurs' allows for a continuum of effect sizes. The traditional view of tumour suppressor genes, based on observations in retinoblastoma, portrays TSGs as recessive: an inactivating mutation requires a second hit in the same gene (such as another inactivating mutation or loss of heterozygosity) in order to cause cancer (152, 153). *RB1* conforms to this model, causing disease in germline mutation carriers who acquire a second, somatic mutation that triggers tumorigenesis.

However, not all TSGs are recessive, as Fearon and Vogelstein noted in 1990 based on the observations of adenomas in patients with familial adenomatous polyposis who did not show allelic losses of the chromosome linked to the syndrome : 'In some cases, mutant tumour suppressor genes appear to exert a phenotypic effect even when present in the heterozygous state; thus, some tumour suppressor genes may not be recessive at all.' (154). Since then, the 'two-hit paradigm' has been revised from a dogma to a model that applies to some but not all TSGs (155, 156). The remaining TSGs show evidence of dosage

sensitivity. For example, not all tumours in Li-Fraumeni syndrome patients, carriers of cancer-predisposing mutations in *TP53*, show loss of *TP53* heterozygosity. Many other haploinsufficient and dosage sensitive TSGs exist and were reviewed by Payne et al.(156). Notable examples include the cell cycle inhibitor *Cdkn1b* (p27); *Nkx3.1; Cdkn1a* (p21) or *Msh2*. *Cdkn1b*$^{+/-}$ mice show an intermediate susceptibility to irradiation-induced tumours between wild type and p27-null mice but do not show deletion of the other *Cdkn1b* allele (157). Similar intermediate susceptibility phenotype in *Cdkn1b*$^{+/-}$ mice was observed with 1,2-dimethylhydrazine induced tumorigenesis (158). Therefore, evidence of biallelic loss is not required for identification of TSGs. Another example of a haploinsufficient TSG is *Trp53bp2*: *Trp53bp2*$^{+/-}$ and *Trp53bp2*$^{\Delta 3/+}$ mice develop spontaneous tumours which include lung adenocarcinomas, squamous cell carcinomas, lymphomas and sarcomas (104, 105, 159). ASPP2 deficiency manifests itself at the cell-autonomous level by increased thresholds for apoptosis, and by defective cell cycle checkpoints in γ-irradiated primary cells (104). Hemizygous deletions occur frequently in human tumours and are thought to act as mini-drivers which optimise the proliferative potential of cancer cells (160).

Berger et al. introduced a concept of ‘obligate haploinsufficiency’, marking TSGs whose optimal inactivation is conditional on one allele being retained (155). For example, in prostate cells, full loss of *Pten* on a wild-type p53 background leads to senescence whereas hemizygous deletion can lead to prostate neoplasia and cancer. Other examples includes *Dicer1* in the lung or *Npm1* in haematologic malignancies.

1.3.2 Tumour mutations are context-dependent

An important concept in understanding tumour suppressor gene function is context dependence. The penetrance of TSG loss is influenced by tissue type and the presence of modifier alleles. For example, germline *BRCA1* mutations increase the lifetime risk of breast cancer and ovarian cancer by up to ~80% and ~50%, respectively, which is thought to be related to increased survival of oestrogen-induced *BRCA1*-deficient mammary epithelial cells (161). *BRCA1* is not somatically mutated at a statistically significant level across pooled cancer types but is significantly mutated in ovarian cancer (162). Another example is *WT1*, the Wilms' tumour gene. Deletions of *WT1* are associated with syndromic cases of Wilms' tumour, a childhood cancer of the kidney whereas frequent somatic mutations drive acute myeloid leukaemia (163, 164).

Modifier alleles can have a decisive effect on the penetrance of TSG mutations. Mice carrying the *Apc*^{min} mutation develop spontaneous polyps in strains positive for a mutation in *Pla2g2a* (originally named *Mom1*), such as C57BL/6, but are resistant to polyps in *Pla2g2a*-wild type strains such as AKR (165). Germline SNPs have been found to modify risk of *BRCA1* and *BRCA2* mutation carriers (166). In human cancers, many mutations are connected by functional networks and co-occur in tumours, for example in the condensin and CLASP and CLIP complexes (167). This suggests some gene networks have built-in redundancy that is broken by multiple mutations within the same network. Copy number variants also show co-occurrence patterns in tumours (168). Conversely, mutations showing mutual exclusivity often mark dominant driver mutations that make another mutation within the same pathway unnecessary or detrimental. It is well known that mutations within druggable pathways can underlie acquisition of

drug resistance. For example, expansion of tumour clones with *MEK1 or MEK2* mutations occurs in some patients with *BRAF* resistance (169, 170). Only a handful of mutations are penetrant enough to be able to transform a wide range of cell types from non-malignant to malignant. The penetrance of most mutations is conditional on the tissue of origin and/or cooperating transforming genetic events. Therefore, mutations in cancer should be considered as being a function of their environment crucially including the tissue of origin and somatic genome of the cell.

A view of cancer as an evolutionary process incorporates much of the complexity discussed above. Cancer shows evidence of complex clonal dynamics including competition, predation, parasitism and mutualism between co-evolving clones (171). Interesting ecological phenomena have also been observed; for example cancer clones injected into opposite flanks of mice and rats can inhibit each other's growth, with only one clone sometimes being specifically affected (172, 173). In ecological terms this is known as an amensal interaction and shows that competition between different cancer clones can directly affect tumour growth. These features along with the empirical parameters that govern and constrain the evolution of each individual cancer determine the clinical course of the disease including its response to therapy. Importantly, understanding the evolutionary and ecological dynamics of cancer can present opportunities for manipulating the disease by applying methods known to affect populations with similar properties (e.g. by introducing a 'predator' such as CAR T cells, or restricting the tumour's resources). A comprehensive understanding of the process of tumorigenesis is needed to identify both stochastic and deterministic parameters that reduce clinical and molecular heterogeneity to predictable phenotypic variability.

A crucial step in building such a framework is linking specific genetic alterations to phenotypic changes and the resulting shift in clonal dynamics. This includes both strong driver mutations and passenger mutations whose collective effect might have a deleterious influence on cell fitness. Indeed, models have shown that the collective burden of passenger mutations is likely to exert deleterious effects (174). One implication of deleterious passengers is an increased requirement of the unfolded protein response which can be exploited therapeutically.

An invaluable insight into the dynamics of cancer has come from the colon in which tumour progression normally follows discrete stages that can be molecularly described and isolated from the preceding and following stages. Thus, genetic events have been reliably linked to the disease progression from normal colon tissue through the hyperproliferative, early to late adenoma, carcinoma, and ultimately metastatic stage (154). *Ras* mutations seem to mark the progression from an early to an intermediate adenoma (greater than 1 cm in size), appearing in over 50% of intermediate and larger adenomas but <10% of benign adenomas below 1 cm. *TP53* loss, for example through 17p deletions, seems to mark the transition from adenoma to carcinoma. However, despite this ‘preferred order’ for a pathway to malignancy, the authors stress that genetic alterations can be acquired in other orders, and it is the progressive accumulation of oncogenic mutations, not their precise order, that is most consistent between distinct colorectal tumours (154). Nonetheless, regularities in the order of acquisition of mutations can be informative with respect to gene function and their relationship to the hallmarks of cancer.

1.3.3 Mini-driver mutations in human cancer

Until recently, the continuous nature of effect sizes among cancer mutations has been underappreciated, with a black-and-white division of drivers and passengers prevailing in cancer research. However, next-generation sequencing of large cohorts uncovered exceptionally diverse sets of mutated genes in many cancer types. These ‘long tail’ distributions argue for the role of many genes in cancer, as they suggest multiple distinct combinations of mutations in multiple different genes can produce the same malignant phenotypes. Moreover, in a recent WGS study of 560 breast cancers, 5% of tumours had no high-confidence driver mutations and many showed only a single high-confidence driver (84). Given that breast cancer is probably the most comprehensively characterised tumour type by genomics to date, this observation underscores a clear gap between existing models of cancer genetics and our knowledge of the causal mutations. Therefore, more attention is being given to rare mutations of potentially large effect as well as small-to-medium-effect ‘mini-drivers’ (175).

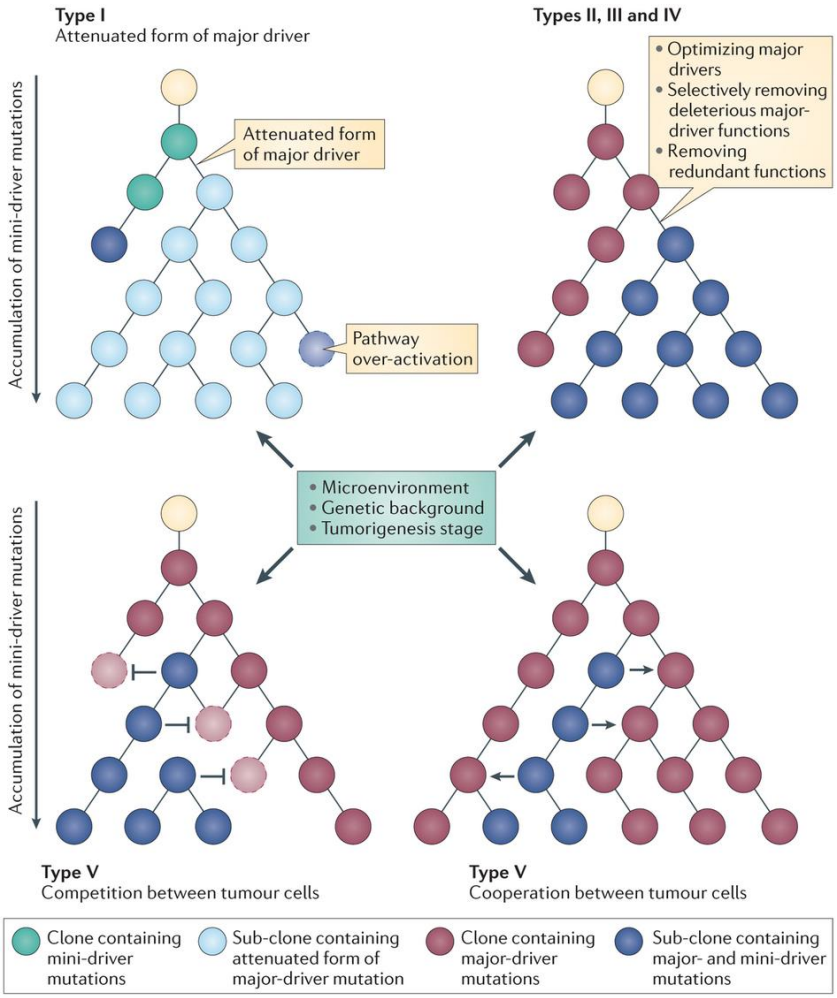


Figure 1.7 Predicted functional classes of mini-driver mutations. Figure reproduced from Castro-Giner et al (175) with permission.

Castro-Giner et al. provide a useful conceptual framework for classifying these mini-driver mutations, and thus help in defining an underappreciated component of the genetic architecture of cancer (**Figure 1.7; Table 1.1**) (175). They propose many cancers might in part follow a polygenic model in which a larger number of mutations with weak effect sizes can drive tumorigenesis. Studies of copy number variations already support this hypothesis (160, 176). Similarly, studies from transposon insertional mutagenesis in mice show that thousands of genes can promote tumorigenesis when disrupted (177, 178). The number of ‘early stage drivers’ is generally much lower than the number of later

stage drivers, indicating that the acquisition of the cancer phenotype requires the convergence on a number of key signalling pathways. Once the cancer has formed, the perturbation of many pathways can influence tumour growth. Hence the heterogeneity of driver mutation is much higher in later stages than in the early stages or *in vitro*. Recently, the *piggyBac* system was used to identify a non-coding regulatory region of *Cdkn2a*, confirming a functional role for non-exonic mutations in cancer, which can also act as mini-drivers (179).

Table 1.1 Summary of mini-driver mutation classes. APC, adenomatous polyposis coli; CNA, copy number alteration; CRC, colorectal cancer; HIF1A, hypoxia-inducible factor 1α; IDH1, isocitrate dehydrogenase 1; VHL, von Hippel–Lindau. Reproduced from Castro-Giner et al. (175) with permission.

Type	Definition	Effect	Possible examples from sporadic human cancers(177)
I	Attenuated form of 'major-driver' mutation	Mutations in the same gene or pathway as the major driver	*KRAS* mutations at codons 146 and 117 (Ref. (180))
II	Optimize major-driver mutations	Keep signalling pathway levels at an optimum	Reduction of the effects of WNT activation, through CNA in *APC* (181) and inactivating mutations in *SOX9* (Ref. (182))
		Modulate the level of genomic instability	None known at this time
		Reverse the effects of deleterious 'hitchhikers'	None known at this time
		Amplify germline variation	Targeting a polymorphism that affected an enhancer that drives an oncogene, such as allelic imbalance at rs6983267 contributing to CRC somatic evolution (183)
III	Selectively remove disadvantageous major-driver functions	Remove growth-suppressing side effects of driver mutations	Renal cancers acquiring inactivating *HIF1A* mutations consistent with selection for removal of that specific consequence of *VHL* mutation (184)
		Remove a major-driver function that has become disadvantageous during somatic evolution	None known at this time
IV	Remove redundant or unnecessary functions	Loss of driver mutations that are no longer needed as cancers progress	Loss of *IDH1* from glioblastoma (although not in therapy-naive tumours) (185)

Type	Definition	Effect	Possible examples from sporadic human cancers(177)
		High frequency of inactivating somatic mutations in some genes with a role in differentiated cells and no plausible role in carcinogenesis	None known at this time
V	Competition or cooperation between tumour cells with minor overall effect	One cell population 'freeloading' or directly harming another	None known at this time
		Promote cooperation among tumour cells	None known at this time

Mini-drivers are predicted to mimic major driver mutations with a smaller effect; to adjust cellular impact of major drivers to optimise fitness; neutralise deleterious passenger mutations; streamline fitness by removing redundant functions; exert inhibitory effects on other clones (for example by secreting a factor the secreting clone is genetically insensitive to); communicate with non-malignant cells to adjust the microenvironment, and play other roles (**Table 1.1**). Identification of these mutations along with a quantitative understanding of their effect sizes will be important in elucidating the somatic genetic architecture of cancer.

1.3.4 Approaches to the discovery and functional analysis of new driver and mini-driver mutations

Given the large number of rare mutations and potential mini-drivers, experimental and computational approaches are needed to aid in identifying these mutations and their effect on cancer growth. For the discovery of traditional Mut-drivers (somatic alleles of large penetrance), most analyses rely on the statistical evaluation of somatic mutation frequency per gene compared to background. As

somatic mutation frequency is not independent of genomic location, tissue type, gene length and other variables, methods such as MutSigCV have been developed to calculate somatic mutation burden corrected for independent confounders (186). Vogelstein et al. proposed a ratiometric method, termed the 20/20 rule, to identify 'Mut-driver genes' that is, cancer driver genes dysregulated through mutations (151). The method classifies oncogenes as those with >20% of mutations at recurrent positions and missense, whereas tumour suppressor genes as having >20% of their mutations inactivating. This approach identified 125 genes based on sequencing data from 3,284 tumours.

In a recent study, Kim et al subjected 474 patient-derived mutant alleles to an *in vivo* tumour formation assay in mice (187). They found that some somatic alleles, which were observed only once in a large pool of patients, and without having known oncogenic function, such as *POT1* G76V, were able to transform cells and cause tumour formation. In this way, mini drivers can be identified experimentally alongside rare major drivers.

Berger et al. took a different approach, screening 194 mutations from lung adenocarcinomas using an expression-based assay which quantified the mutant protein's effect on the transcriptome (188). Lentiviral transduction of mutant genes into lung adenocarcinoma cells followed by transcriptional profiling of 1,000 transcripts yielded transcriptional data that were used to compare the impact of each mutant with wild type and with reference driver alleles. 69% of the variants were characterised as impactful and many of those, including non-canonical mutations in known driver genes such as *KRAS* D33E, promoted tumour growth *in vivo*. Importantly, these newly identified oncogenic alleles were not always predicted to be impactful by *in silico* methods such as SIFT or PolyPhen2.

Although it is difficult to predict from this study what percentage of somatic mutations may be drivers due to the non-random selection of candidate genes, it is important in providing proof that rare somatic mutations can be drivers, irrespective of their statistical enrichment within patient cohorts.

A recently developed program called REVEALER (repeated evaluation of variables conditional entropy and redundancy) demonstrated the feasibility of systematically identifying genetic interactions between driver alterations correlated with functional phenotypes (189). As examples, application of REVEALER across 83 cell lines from the Cancer Cell Line Encyclopaedia (CCLE) found that *APC* mutations of 13q33 amplification can substitute β-catenin mutations. Analysis of 182 lung cancer cell lines from the CCLE revealed that *KEAP1* mutations and 15q22/26 amplification can act as alternative causes of *NRF2* activation. It would be interesting to see how many mutations in unclassified genes would show such correlations with major drivers across human cancers.

Another recently published tool called VIPER (virtual inference of aberrant protein activity) provides the means to estimate protein activity from transcriptional data (190). This can be useful in predicting chemosensitivity, as mutation status alone is a poor predictor of response, which limits precision oncology (191). Interestingly, the authors show that some tumours and cell lines without a mutation in a particular gene show a near-identical reduction in protein activity to samples with a driver mutation in the gene, as measured by VIPER's mutant phenotype score (MPS). This tool might be useful in analysing which rare drivers and mini-drivers substitute for canonical driver mutations – by analysing the genomic makeup of these tumours with wild-type driver genes but mutant-like

signature, it might be possible to discover sets of mutations in a related pathway. However, such application of VIPER remains to be investigated.

Multiple computational predictors of individual driver mutations exist, for example CHASM (192), FATHMM Cancer (193) and ParSNP (194). These methods score and rank somatic mutations based on their functional impact on protein function and likelihood to be involved in cancer. However, the application of their methodology in discovery of mini-drivers presents important caveats. Firstly, with the exception of ParSNP, the algorithms were trained using a set of curated driver mutations. This skews the predictions towards mutations that are most similar to currently known drivers such as *KRAS* G12D. Secondly, the benchmarking of the methods is often limited to the top 1% of candidates out of which only several mutations per tumour are discussed (top 5 in the ParSNP paper) (194). In fact, in developing ParSNP the authors specifically made use of the assumption that the proportion of mutations that are drivers decreases with mutation load. Therefore, while this approach might be successful in identifying rare drivers of large effect, mini-drivers may be less amenable to discovery by these methods because they are less impactful, by definition. Therefore, tools that are optimised for sensitivity in detecting driver mutations might overlook mini-drivers.

In summary, although encouraging progress has been made in the discovery and functional analysis of rare driver and mini-driver mutations in cancer, much more investigation is required before these mutation classes can be reliably and robustly identified. Of particular importance is the synergy between computational and experimental methods, in order to provide causal evidence of function for newly discovered mutations. Similarly, mechanistic studies focused on

individual genes will be important in characterising the landscape of effects of each mutation on gene function. Finally, cancer gene databases have to keep pace with the accelerating rate of cancer gene discovery by curating and synchronising gene lists discovered in multiple studies.

In Chapter 4.2, we discuss how mini-driver mutations can be discovered from cancer genomic data, using the ASPP gene family as an example.

1.3.5 Public databases of cancer driver genes

Thousands of genes have been implicated in the molecular and cellular processes relevant to cancer. Many of those have been shown to drive tumorigenesis in animal models, including unbiased transposon-mediated mutagenesis screens, shRNA screens in mice as well as targeted mutations and knock-in alleles, resulting in wealth of information on genes which have the potential to drive cancer in mice. However, it remains unclear which proportion of these genes drive tumorigenesis in clinically observed human tumours.

As a result, there is a clear gap between our knowledge of the functional genetics of cancer in model systems such as cell lines and mice, and the driver genomic landscape of human cancer. An important component in bridging this gap is a high-quality curated list of human cancer genes. Perhaps surprisingly, in contrast to public portals for cancer genomic data, an up-to-date comprehensive reference of this kind is still lacking. Arguably the most widely used list of cancer genes is the Sanger Cancer Gene Census (http://cancer.sanger.ac.uk/census/).

What are the empirical criteria for inclusion of genes in the Cancer Gene Census? In compiling the Census, Futreal et al. selected genes with ‘at least two independent reports showing mutations in primary patient material’, the most common alteration being chromosomal translocation at the time (195). To filter out likely passenger mutations, they excluded genes with fewer than five unambiguous mutations reported in primary neoplasms, low-penetrance cancer susceptibility genes, genes mutated primarily in mismatch repair-deficient tumours, genes located at fragile genomic sites (e.g. fragile histidine triad gene, *FHIT*) and genes altered by large copy number variants without high confidence of being the causal gene. Applying these criteria yielded 291 human genes in the first draft of the Census, many of these being recognised driver genes such as tyrosine kinases *RET* or *EGFR*.

However, application of these criteria in the era of whole-exome and whole-genome sequencing, when over 10,000 tumours have been sequenced by these techniques, would result in an order-of-magnitude higher number of genes. Because of the ‘long-tails’ distribution of somatic mutation frequency of most tumour types, hundreds of genes are mutated in independent samples (196, 197). This includes large genes such as *TTN* that are statistically most likely to accumulate somatic mutations by random chance due to their genomic size. Additional confounding factors such as genomic locus replication timing, GC-content, transcription-coupled repair also influence mutation rate (186). Therefore, more stringent methods are needed to increase the signal-to-noise ratio in detecting driver mutations. At the same time, as genomic studies continuously produce new evidence of causal role for genes in different cancer types, this stringency should be balanced by periodic updates of the database and the swift

inclusion of newly emerged data. Unfortunately, in the current version of the Census, little to no information is given on the source of current evidence supporting the membership of each gene.

Therefore, other platforms are being developed. One recent alternative is the DriverDBv2 (http://driverdb.tms.cmu.edu.tw/driverdbv2/index.php), a database collating ca. 13,000 exome-sequenced cancer cases and integrating 15 algorithms that analyse the mutation of each gene and call significantly mutated genes per study (198). The portal allows an interactive inspection and filtering of the statistical evidence for study type, for example by selecting the minimum number of algorithms that deem each gene significant. Unfortunately, the current implementation of DriverDBv2 does not allow for an easy access to driver gene lists compiled using specified criteria across all studies. These observations underscore the importance of data curation alongside high-throughput sequencing and data analysis.

As the number of sequenced cases and the quality of algorithms increase, we hope to approach saturation in the discovery of Mut-driver genes based on the over-representation of their somatic mutations. Currently, approximately 100,000 sequenced cancers are estimated to reach saturation, or 2,000 tumours for at least 50 tumour types (**Figure 1.8**) (162).

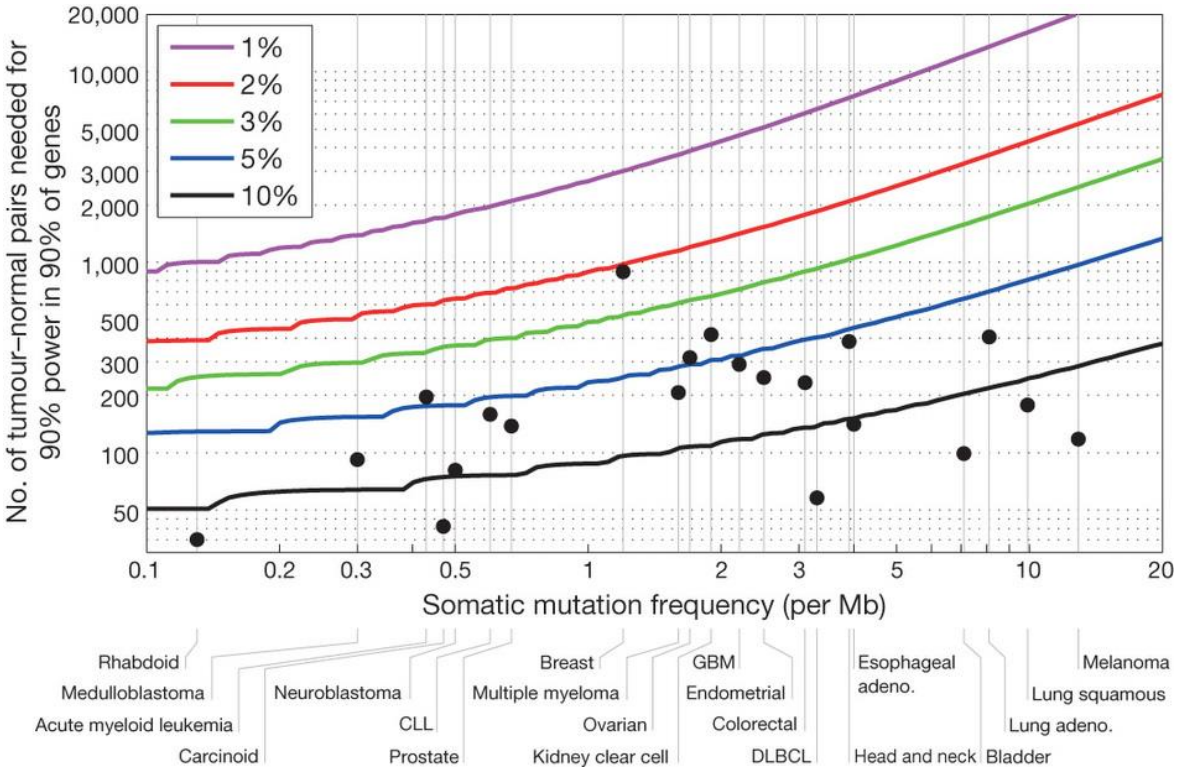


Figure 1.8 Number of tumour-normal pairs required to detect Mut-driver genes as a function of somatic mutation frequency. Reproduced from Lawrence et al (162) with permission.

Excluding mini-drivers, how many Mut-driver genes have been discovered? The Cancer Gene Census currently (as of September 2016) holds 602 genes out of which 264 have mutational evidence other than amplifications, deletions and translocations assigned to them (i.e. putative Mut-drivers). Tamborero et al. reported a list of 291 high-confidence cancer genes from 3,205 tumours of 12 different cancer types (199). Applying a stringent false-discovery threshold, Lawrence et al. reported a list of 219 Mut-driver genes from 4,742 cancer exomes (162). An alternative analysis by Vogelstein et al. using the 20/20 rule yielded 125 Mut-drivers based on data from 3,284 tumours. From DriverDBv2 it is difficult to establish precisely because multiple algorithms are integrated with varying results and no single threshold is used. A resource named Network of Cancer Genes (http://ncg.kcl.ac.uk/index.php) states a considerably higher estimate of the

number of cancer genes: 1,571 in version 5.0 (200). However, this number includes genes dysregulated by amplification, translocation and other means. Taken together, current estimations of Mut-driver genes across human cancers range from 125-291 genes, and estimates of all driver genes from 602-1,571.

However, a closer inspection of the predicted sample sizes required for complete discovery of Mut-driver genes in various cancer types (**Figure 1.8**) reveals that saturation is unlikely to occur soon with rare Mut-drivers; for example in melanoma, lung squamous cell carcinoma and lung adenocarcinoma, over 10,000 tumour-normal pairs would be required to achieve 90% power for 90% genes mutated at >1% frequency. Given the existence of mini-drivers and rare Mut-driver alleles discussed above, significantly more sequencing is required before all relevant pathogenic alleles in the genetic architecture of cancer are uncovered. The need to identify these rare drivers has become a priority in cancer genomics (201). Based on this, approaches alternative to gene-level statistics of somatic mutation burden are necessary to accelerate cancer research and the discovery of cancer genes.

1.3.6 Reactivating CD8+ T cells to target diverse tumour-specific antigens

Cancer is a disease shaped by the somatic evolution of rogue cells at the expense of the host. Adaptive immunity in higher organisms is shaped by controlled somatic evolution of specialised cells that protect the host. Other tissues in the body do not undergo programmed somatic mutation or have no capability of exerting tumour-killing effects. The human repertoire of circulating αβ-T cells is estimated at ca. 10^6 β chains, each pairing on average with ca. 25 different α

chains (202). Similarly, each human contains potentially up to 10^{11} unique antibody clones which can undergo further affinity maturation upon antigen stimulation (203, 204). Thus, antibodies and T cells can recognise a theoretically infinite number of antigens, including epitopes in mutant proteins expressed by cancer cells. Therefore, the interface between cancer and the immune system has an enormous potential to influence cancer progression. Given the high genetic diversity including the presence of mini-drivers, tumour heterogeneity, and the emergence of resistance to treatment observed in most advanced cancers, an exhaustive development of targeted therapies to combat each individual tumour may be beyond reach. A more realistic approach is to enlist the adaptive immune system with its vast clonal repertoire of killer T cells and antibodies to aid by selectively destroying tumour-antigen-expressing cells.

One challenge is to manipulate the immune response to cancer in a way that maximises tumour killing over undesired immunopathology. Although the field of tumour immunology dates back to the early 1900s, immuno-oncology has only reached the clinical mainstream in recent years, mainly due to the success of anti-PD1, anti-CTLA4 antibodies, frequently called immune checkpoint inhibitors, and CAR T cells. Following extraordinary results in clinical trials (205, 206), checkpoint inhibitors have transformed the clinical management of metastatic melanoma. Importantly, a subset of melanoma patients experience a durable response that leads to eradication of metastatic disease, which was unachievable with previous therapies, including BRAF inhibitors. BMS's anti-CTLA4 antibody ipilimumab became the standard of care in advanced melanoma following its FDA approval in 2011, however anti-PD1 antibodies nivolumab and pembrolizumab showed even more promising results and are now approved for melanoma as well (207, 208).

Nivolumab has gained FDA approval for advanced melanoma, advanced lung cancer, metastatic renal cell carcinoma, and most recently Hodgkin lymphoma (www.fda.gov). Merck's pembrolizumab (Keytruda) now holds FDA approval for advanced melanoma, advanced non-small cell lung cancer, and head and neck squamous cell carcinoma. In addition, clinical trials for multiple other cancer types are ongoing.

In order to realise the full potential of T cells in the fight against cancer, a deep understanding of the pathways controlling T cell proliferation, activation and dysfunction is needed. The role of PD1 in regulating the T cell response was originally discovered in the lymphocytic choriomeningitis virus (LCMV) experimental model of viral-induced immune suppression (209). PD1 is an inhibitory receptor which increases the threshold for effector CD8+ T cell activation upon binding of its ligand PD-L1. PD1 blockade enables tumour antigen-specific CD8+ T cells to elicit tumour cell killing by inhibiting PD-L1's binding to its receptor (210, 211). Effective checkpoint blockade typically relies on the presence of neoantigen-specific CD8+ T cells, and studies have indeed shown that neoantigen load correlates with checkpoint inhibitor sensitivity (212, 213). Despite the wealth of knowledge surrounding PD1 and T cell exhaustion, the precise molecular programme which induces PD1 expression and suppresses CD8+ T cell responses is not fully understood. Moreover, recent studies have shown that a durable response to PD1 blockade relies the proliferation of a subset of CD8+ CXCR5+ T cells with a distinct phenotype from exhausted terminally differentiated effector cells (214). Therefore, the molecular basis of T cell exhaustion, and the cell type dynamics of the CD8+ compartment in chronic viral infection are still incompletely understood.

Chapter 5 of this thesis explores a transcriptomic approach to assessing the viral-specific and total CD8+ T cells in chronic vs acute infection.

1.4 ASPP family

1.4.1 Discovery and members of ASPP family

The ASPP family, defined by functional and domain homology, are Apoptosis Stimulating Proteins of p53, and also Ankyrin repeat, SH3 domain and Proline-rich region-containing Proteins (**Table 1.2, Figure 1.9**) (103). The ASPP are not the only protein family characterised by this domain composition; SHANK proteins also contain multiple Ankyrin repeats, an SH3 domain, proline-rich region, in addition to PDZ and SAM domains (**Figure 1.10**) (215). Human ASPP family comprises 3 members: ASPP1, encoded by *PPP1R13B*; ASPP2, encoded by *TP53BP2*; and iASPP, encoded by *PPP1R13L* (**Table 1.2**). These names reflect the proteins' regulatory effects on apoptosis: ASPP1 and ASPP2 are activators or p53-dependent apoptosis whereas iASPP inhibits it.

Table 1.2 ASPP family members and their gene names. Alt, alternative exon.

Protein name	Human gene name	Human exon count	Human protein length	Mouse gene name	Mouse protein length
ASPP1	*PPP1R13B*	17 (+2 alt)	1090	*Ppp1r13b*	1087
ASPP2	*TP53BP2*	18 (+2 alt)	1128	*Trp53bp2*	1128
iASPP	*PPP1R13L*	13 (+1 alt)	828	*Ppp1r13l*	824

The first ASPP member to be discovered was ASPP2. In 1994, using a yeast two-hybrid assay on a B-cell cDNA library, Iwabuchi et al. identified 2 proteins binding wild type but not mutant p53; one of these proteins was named 53BP2, now understood to have been a fragment of full-length ASPP2 (216). Interestingly, 53BP2 specifically bound the DNA-binding domain of p53. This property connects ASPP2 with p53-inactivating viral proteins large T antigen from

the SV40 tumour virus, and E6 from human papilloma virus, which also specifically bind to the non-mutated DNA-binding domain of p53 (217, 218). Iwabuchi found a role for ASPP in stimulating p53-dependent transcription (219). 53BP2 was also shown to bind BCL2, although the function of this interaction is not well understood (220). These initial experiments identified fragments and isoforms of ASPP2 which are shorter than the full-length isoform comprising 18 exons, which is the dominant isoform expressed in human cells. A shorter isoform called Bbp (UCSC transcript name uc001hod) is generated from the *TP53BP2* gene by alternative splicing.

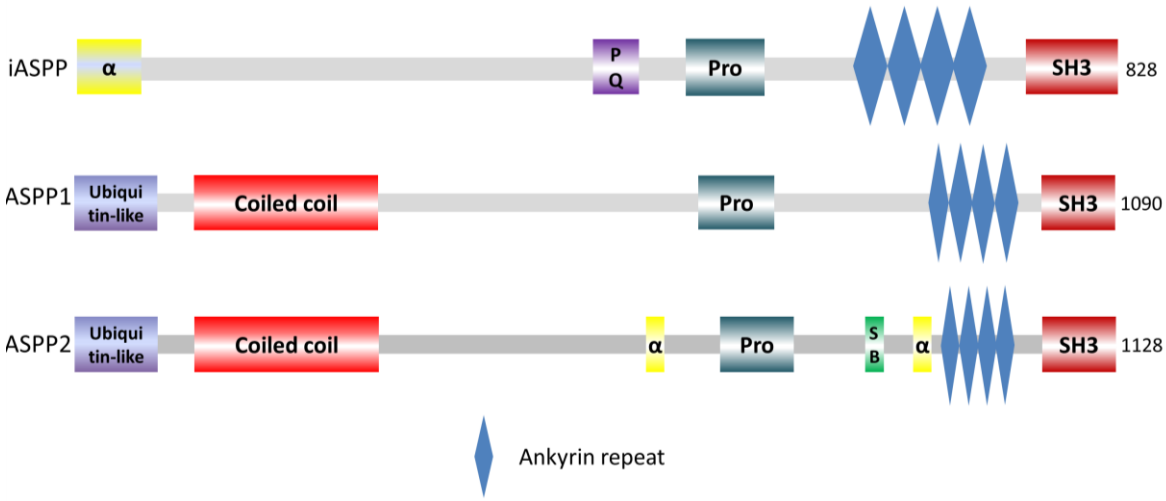


Figure 1.9 Domain maps of ASPPs. Pro, proline-rich region (conserved part); SB, SH3-binding domain; PQ, proline- and glutamine-rich region. Numbers on the right indicate protein length of the full-length isoforms.

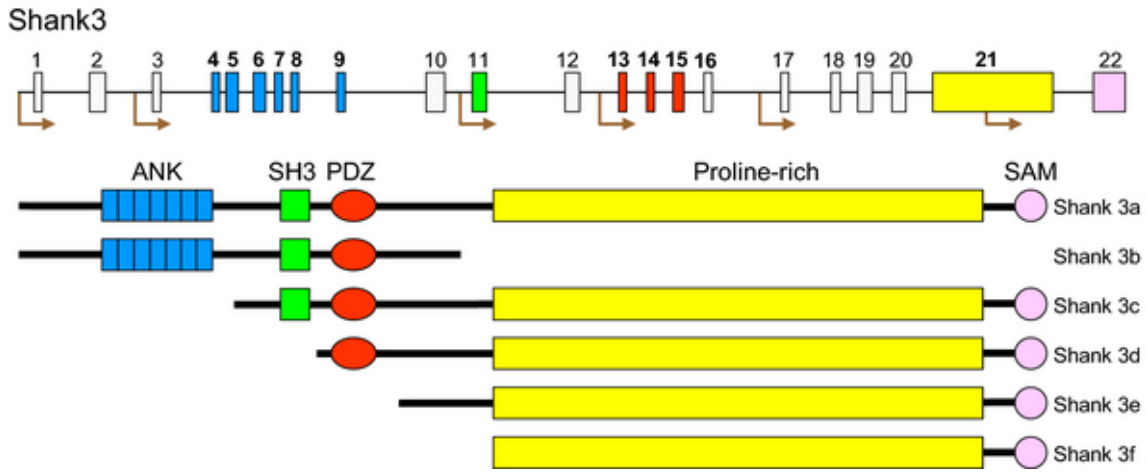


Figure 1.10 SHANK3 domain structure and splice forms. Reproduced from Sala et al. (215) with permission.

A fragment of ASPP1 was first reported in 1998 as a cDNA isolated from the brain with high homology to 53BP2 (221). The full-length and most abundant isoform of ASPP1 in humans is 1090 residues long. ASPP1 and ASPP2 show very high homology, in particular in the C-terminus, and an identical domain structure: both proteins share a C-terminal SH3 domain, 4 ankyrin repeats, a proline-rich region, an N-terminal coiled-coil sequence, and a ubiquitin-like/RAS-association domain at the very N-terminal end (**Figure 1.9**). Samuels-Lev et al. showed that ASPP1 and ASPP2 selectively enhance p53-dependent transcription on pro-apoptotic gene promoters (*BAX*, *BBC3* and *TP53I3*) but not cell cycle arrest gene promoters (such as *CDKN1A*, encoding p21) (103). A crystal structure of ASPP2's C terminus in the complex with p53 has been solved (102).

The third ASPP member, iASPP, was first reported in 1999 as a protein inhibiting the NK-κB transcription factor p65 (222). The full-length isoform of iASPP, 828 residues long, is homologous to ASPP1 and ASPP2 in the C-terminal SH3 domain, Ankyrin repeats, and conserved proline-rich region, but not the preceding sequence which is also proline-rich (**Figure 1.9**). In fact, proline

represents 15.7% of human iASPP's sequence (compared with the total proline content of 6.3% among 18,666 human proteins) placing iASPP in the top 1.5% of human proteins by proline content (calculated using data from Morgan et al (223)). A crystal structure of iASPP's C terminus is available (224). In contrast to ASPP1 and ASPP2, iASPP inhibits p53 function by reducing its transcriptional activity on pro-apoptotic promoters (225). Thus, iASPP is differentiated from ASPP1 and ASPP2 not only by its protein sequence, but also by its opposing effect on p53-dependent apoptosis.

1.4.2 Evolution of the ASPP family

Most vertebrate genomes contain 3 distinct ASPP genes. Single ASPP orthologues have been found in *Drosophila melanogaster* and *C. elegans*. Early studies suggested that iASPP is the ancient member of the family, with ASPP1 and ASPP2 evolving at later stages (225, 226). The *C. elegans* ASPP orthologue inhibited apoptosis in germ cells, consistent with it being an orthologue of iASPP rather than ASPP1 or ASPP2 (225, 227).

However, a recent phylogenetic analysis by Song et al. suggests the issue might be more complicated (228). Song et al. identified an ASPP-like protein in the invertebrate *Branchiostoma belcheri* and showed it to be more homologous to human ASPP1 and ASPP2 than to iASPP. Moreover, a Bayesian analysis of individual ASPP proteins across species produced 4 distinct clusters: a vertebrate ASPP1 group, ASPP2 group, iASPP group, and an invertebrate ASPP-like group which contained proteins with motifs homologous to ASPP1 and ASPP2's ubiquitin-like and coiled-coil domains. These organisms do not seem to contain

another copy of ASPP, suggesting some lower organisms contain an ASPP1/2-like orthologue rather than an iASPP orthologue. The notion that iASPP might not have been the single ancient gene is also supported by the fact that some chordates seem to possess ASPP1 and ASPP2 orthologues but no iASPP orthologues (**Figure 1.11**). In contrast, no chordates seem to possess only iASPP in the absence of ASPP1 and ASPP2. Taken together, these insights present an alternative hypothesis: the ancient ASPP gene contained the N-terminal domains of ASPP1 and ASPP2 which in some invertebrates were retained, diversified, and in others (such as *C. elegans*) were lost. Thus, the ancient ASPP might have been duplicated in an ancestor of all vertebrates to produce ASPP1 and ASPP2, and in some vertebrates including humans, further duplicated to produce iASPP through a similar mechanism that led to the evolution of ASPP in *C. elegans*. The occurrence of an iASPP-like protein in humans and *C. elegans* would thus be an example of convergent evolution, rather than a manifestation of a common ancestor. An alternative hypothesis is that all 3 ASPPs were present in a common vertebrate ancestor, but iASPP was later lost in some vertebrates, for example birds and lamprey (**Figure 1.11**). A more detailed phylogenetic analysis is required to differentiate between these hypotheses.

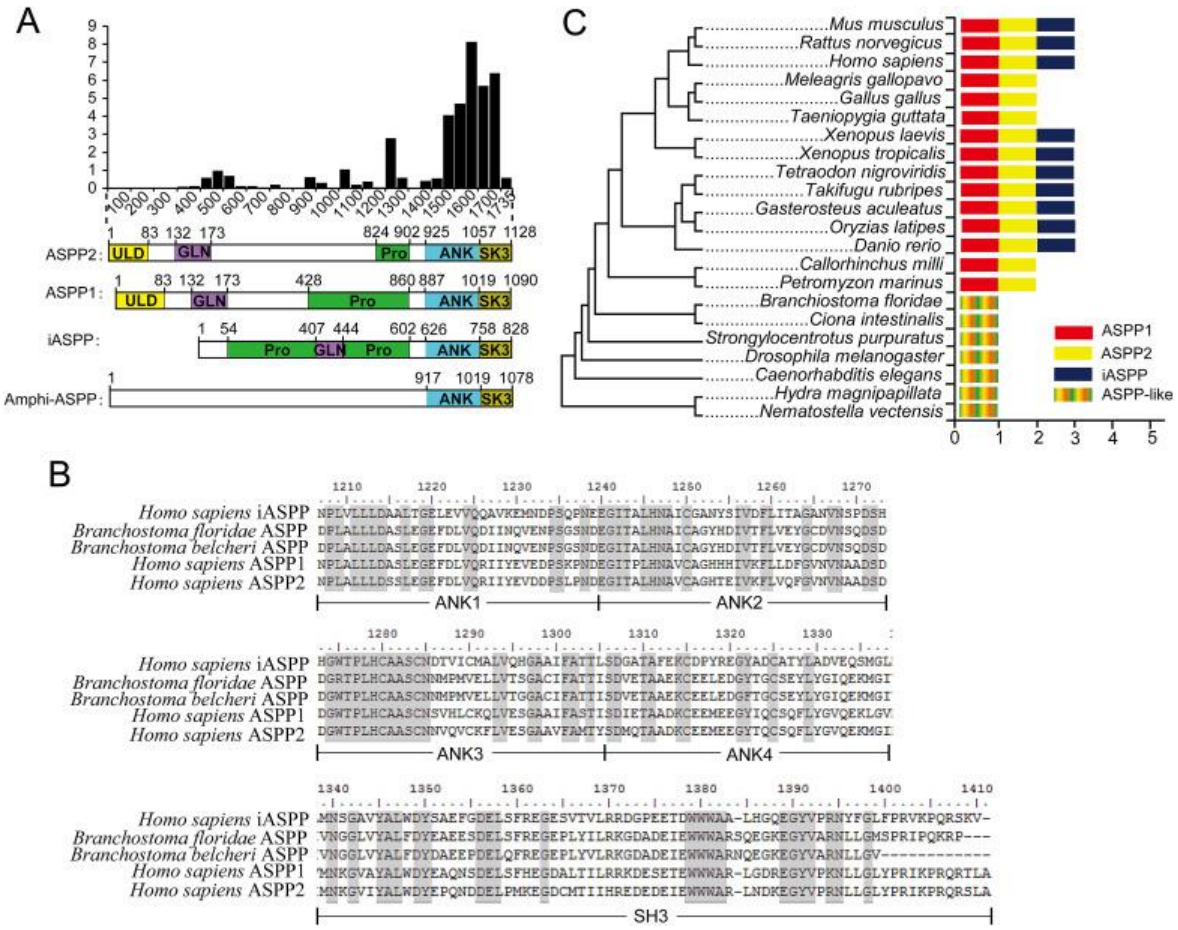


Figure 1.11 A phylogenetic survey of ASPP family genes. Reproduced from Song et al. (228) with permission.

Vertebrate conservation scores support the hypothesis that ASPP2 is the most ancient ASPP protein: exons coding for N-terminal sequence of ASPP2 have much higher conservation scores among vertebrates than exons coding for N-terminal sequence of iASPP (UCSC Genome Browser). Song et al. also showed that among vertebrates, *TP53BP2* has experienced a slower rate of nonsynonymous/synonymous substitutions, suggesting a faster rate of adaptation of *PPP1R13B* and *PPP1R13L*, and their orthologues. This is consistent with a hypothesis that *PPP1R13B* and *PPP1R13L* were generated by duplication of *TP53BP2* during evolution, as the newly created genes would need to undergo a faster rate of evolution in order to diversify from ASPP2's function, especially in

tissues where *TP53BP2* is dosage sensitive. In addition, the proline-rich region conserved between all three human ASPPs is a GC-rich sequence (Pro DNA codons are CCA/T/C/G), which undergoes a higher mutation rate than GC-poor sequence. Altogether, phylogenetic and genomic evidence support a more ancient role for ASPP2 as the common ancestor of the ASPP family, and suggest that vertebrate iASPP and its invertebrate orthologues evolved through loss of ASPP2's N-terminal domains.

1.4.3 Molecular function of ASPPs

As described above, ASPPs modulate p53-dependent apoptosis. In addition, ASPP1 and ASPP2 also directly stimulate pro-apoptotic transcription by the p53-family members p63 and p73 (107). Of note, the relative effect of ASPP1 and ASPP2 on p53/p63/p73 transcription at the *BBC3* (*PUMA*) gene promoter is the inverse of the relative effect on the *BAX* and *TP53I3* promoters: the relative fold enhancement by ASPP1/2 on *BAX* and *TP53I3* is p53>p63>p73, whereas the BBC3 promoter shows p73>p63>p53 (107). The molecular basis of this trend is currently unknown. ASPPs do not affect p53 family transcription factors' activity on the promoter of cell cycle inhibitor *CDKN1A*, suggesting ASPPs play a role in deciding the cellular response to stress by promoting apoptosis at the expense of cell cycle arrest (229).

The molecular interactions of ASPPs stretch far beyond the p53 family and apoptotic regulators. The BioGRID database indexed 110, 38 and 26 unique interactors for ASPP2, ASPP1 and iASPP, respectively (v3.4 checked on 2 October 2016). These figures are also consistent with a more ancient role for ASPP2 as the most highly interconnected protein. These relatively high numbers

of interactions also reflect the domain structure of the ASPPs: Ankyrin repeats, SH3 domain and ubiquitin-like domain all mediate protein-protein interactions. To date, no catalytic, DNA-binding, lipid-binding, membrane-binding, or chaperone activity have been reported for ASPPs, suggesting their primary role is that of protein-protein interaction hubs. This is also consistent with the domain structure being most similar to SHANK proteins, which serve as protein scaffolds.

Among the best characterised functions of ASPPs is modulation of PP1 activity; indeed, *PPP1R13B* and *PPP1R13L* bear their name owing to them being regulatory subunits of PP1, a feature of many Ankyrin-repeat proteins (230). However, all 3 ASPPs regulate PP1; therefore, *TP53BP2* is sometimes called *PPP1R13A*. ASPP2 has been shown to dephosphorylate TAZ to promote TAZ-dependent gene expression, thereby antagonising LATS kinase (231). We showed that another ASPP2-PP1 substrate is YAP which binds to a poly-proline motif in ASPP2 located prior to the Ankyrin repeats (232). Both YAP- and PP1-binding to ASPP2 is required for this process, and YAP's promotion of transcriptional activity is mediated by TEAD. Thus, ASPP2's control of YAP represents modulation of one transcriptional co-factor by another.

ASPP2 has been identified as a YAP-interacting protein in multiple proteomic experiments, suggesting this ASPP2-dependent regulation of YAP is a ubiquitous phenomenon (233, 234). Indeed, Hauri et al. identify the PP1-ASPP module as a key node in the regulation of the Hippo pathway in human cells (235). YAP, in turn, regulates ASPP2 stability by competing with the ubiquitin ligase Itch for the same binding site on ASPP2 (236). YAP is a regulator of cell proliferation and epithelial adhesion, playing a key role in cell contact inhibition (237). ASPP1 also regulates PP1 and the Hippo pathway, for example by binding LATS (238).

ASPP1 has been shown to bind PP1 in the testis and brain, tissues with high ASPP1 levels (239, 240). Interestingly, ASPP2 and SHANK3 were also detected as binding partners of PP1 in the brain (240). ASPP1 and ASPP2 also regulate chromosome segregation by enhancing the PP1-dependent dephosphorylation of kinetochores in mitosis (241).

In addition to cooperating with YAP, ASPP2 regulates apicobasal polarity by binding Par3 (106, 242). These observations suggest ASPP2 stabilises epithelial cell character by maintaining cell-cell junctions through Par3 and YAP. Indeed, we have shown recently that ASPP2 inhibits epithelial-to-mesenchymal transition by repressing ZEB1 expression through β-catenin (130). The integration of cell-cell junction maintenance with regulation of apoptosis also makes ASPP2 a target for pathogens, for example the toxin CagA, injected into epithelial cells by the oncogenic bacterium *Helicobacter Pylori*. CagA has evolved to bind a specific site within the Pro-rich region of ASPP2 and helps remodel tight junctions to allow for bacterial attachment (243, 244). At the same time, CagA hijacks ASPP2 in order to carry out p53 degradation. ASPP2 is also targeted by the core protein of hepatitis C virus to inhibit p53-mediated apoptosis (245).

ASPP2 also regulates the NF-kB pathway by binding IκB (159). Notably, ASPP1 and ASPP2 modulate the dominant oncogene RAS by promoting RAS-dependent senescence and inhibiting autophagy (246-248). In keeping with its role as a protein-protein-interaction hub, ASPP2 has since been shown to act as a pan-RAS nanocluster scaffold which steers cells towards RAS-induced senescence (249). This nanoclustering activity of ASPP2 is conditional on the presence of the coiled-coil N-terminal domain which is absent in iASPP. In liver cells, ASPP2 seems to promote autophagy as a means to trigger apoptosis via

DRAM (250). In oxaliplatin-treated colorectal cancer cells, ASPP2 inhibits autophagy, highlighting the context-dependent nature of ASPP2's functions (251).

1.4.4 Regulation of ASPP activity

Interestingly, we showed recently that all ASPPs, and some additional Ankyrin repeat proteins such as IκB, use an alternative nuclear import pathway we named RanGDP/Ankyrin-Repeat complex-mediated nuclear import pathway (RaDAR) (252). This pathway enables nuclear import of specific Ankyrin repeat motifs into the nucleus without the need for importins by utilising direct binding of the GTPase Ran. Given that ASPPs perform important functions both inside and outside the nucleus, this pathway is important in mediating a nucleocytoplasmic shuttling of ASPPs, for example to integrate signalling at the cell-cell junction with regulation of gene expression.

Given the multitude of ASPPs' binding partners, regulatory mechanisms are needed to make the protein available for a subset of specific interactions. The intrinsically disordered Pro-rich sequence of ASPPs mediates the availability of the C-terminal Ankyrin repeat/SH3 domain (Ank-SH3) by an autoregulatory mechanisms. In ASPP2, the conserved Pro-rich region has been shown to bind to the Ank-SH3 domain which modulates its interaction with p53 (253, 254). A similar mechanism is suspected for ASPP1 although this has not yet been demonstrated experimentally. Since the Ank-SH3 domain is required for nuclear import via Ran, and for the interaction with p53/p63/p73, the Pro <-> Ank-SH3 intramolecular interaction must be disrupted in order for ASPP2 to bind p53/p63/p73. Godin-Heymann et al. showed that phosphorylation of Ser827 of ASPP2 is required for

its pro-apoptotic function (255). S827 is located between the Pro-rich region and ASPP2's YAP-binding site, therefore it is plausible that phosphorylation of this residue promotes the 'open form' of ASPP2 (in the absence of Pro <-> Ank-SH3 interaction) which then translocates to the nucleus. In contrast, ASPP2 S698A was able to enhance p53-*BAX* transcription more significantly than ASPP2 wild type. S698 is located within the 'PRPLS' motif at the core of the conserved Pro-rich region; this motif is conserved between all 3 ASPPs and directly interacts with Ank-SH3 in ASPP2. Therefore, it can be speculated that phosphorylation of S698 might stabilise the 'closed form' of ASPP2 (with the Pro <-> Ank-SH3 interaction in place). This hypothesis is supported by the fact that the orthologous residue in mice is phosphorylated in the brain, kidney, liver, brown fat and spleen tissues (phosphomouse.hms.harvard.edu) which are expected to require the extranuclear functions of ASPP2 such as promoting epithelial cell-cell junctions. Notably, this autoregulatory mechanism is hijacked by CagA. CagA binds the more C-terminal segment of the Pro-rich region, which turns into an α-helix upon encountering CagA, in order to hijack ASPP2's C-terminus and degrade p53 (244, 256).

Although iASPP possesses a homologous Pro-rich region, its vast unique Pro-rich N-terminal domain presents other opportunities for autoregulation and therefore iASPP seems to have evolved a distinct mechanism of self-binding that utilises SPRKA motifs in the N-terminal region (residues 84-88) and in the sequence preceding the Ankyrin repeats (residues 616-620). This peptide is interacting with Ank-SH3 of iASPP in the published crystal structure (224), binding the cleft between the Ankyrin repeats and SH3 domain. Lu et al. showed that the phosphorylation of S120 by CDK1 promotes the dissociation of the N-terminal SPRKA from the Ank-SH3 domain, enabling iASPP to inhibit p53 in the nucleus

(257). This physiological mechanism is dysregulated in some melanomas with high CDK1 activity that keep constitutive iASPP in the nucleus as a means to inactivate wild-type p53. Taken together, studies show that autoregulation via Pro-rich sequences is important for directing ASPP2 and iASPP activity, and that this mechanism is disrupted by a bacterial toxin in the gut, and by aberrant kinase activity in melanoma.

1.4.5 The roles of ASPPs *in vivo*

Given a rich repertoire of molecular functions, it is not surprising that ASPPs are indispensable *in vivo*. Homozygous deletion of *Trp53bp2* in mice results in early embryonic lethality, demonstrated in two independent mouse lines (104, 258). Homozygosity for a hypomorphic allele (exon 3 deletion; *Trp53bp2*$^{\Delta 3/\Delta 3}$) results in prenatal or neonatal death, and is associated with severe brain abnormalities (106). *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos show overgrowth of neuroepithelial cells, loss of cell polarity, abnormal cytoarchitecture of ganglionic eminences, impaired interkinetic nuclear migration and other abnormalities. Ganglionic eminences are a source of a population of neurons that migrate tangentially to the cortex, including GABAergic interneurons (**Figure 1.12**) (259). Most of the GABAergic interneurons are thought to originate in these subpallial structures, migrating tangentially over great distances, in contrast to the default and shorter mode of radial migration used by excitatory neurons (260). Striatal cells, predominantly radially migrating, are also derived from the ganglionic eminences.

Trp53bp2$^{\Delta 3/\Delta 3}$ mice that survive birth have a 100% incidence of hopping gait, and virtually all males and many females develop hydrocephalus. This demonstrates ASPP2 is indispensable for central nervous system development, and points to a possible striatal disruption in the *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice that survive birth. ASPP2 is also a haploinsufficient tumour suppressor in mice (104, 105, 159).

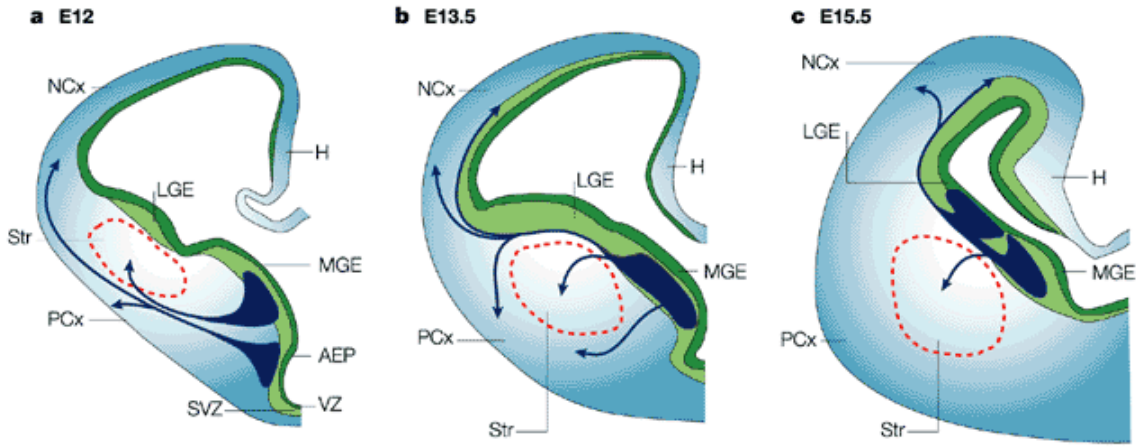


Figure 1.12 Migratory routes of ganglionic eminence-derived cells during mouse brain development. Reproduced from Marin and Rubinstein (259) with permission. AEP, anterior entopenduncular area; Str, striatum; LGE, lateral ganglionic eminence; MGE, medial ganglionic eminence; NCx, neocortex; SVZ, subventricular zone; VZ, ventricular zone; PCx, piriform cortex; H, hippocampus.

ASPP1-deficient mice are born at a reduced rate and show defects in the lymphatic vasculature (261). The reasons for the reduced viability of *Ppp1r13b*$^{-/-}$ mice are unknown.

The role of iASPP in mouse development was first studied in spontaneously arising mutants. Waved 3 mice (wa3) showed skin and heart abnormalities, with a rapidly progressing cardiomyopathy. The mice were born with precocious open eyelids, had disorganised hair follicles and wavy hair (262). The mutation was mapped to a 14 bp deletion in *Ppp1r13l* that removes a splice donor site, resulting in loss of iASPP expression in the homozygous state. Another spontaneous

mutation, deletion of exons 9-11 of *Ppp1r13l*, was reported by Toonen et al. (263). Named waved with open eyelids 2 (woe2), these mice shared the open eyelids at birth phenotype, cardiac abnormalities, and wavy coat, with wa3. In addition, woe2 homozygous mice showed corneal opacities and absence of the Meibomian glands. A frameshift mutation in *Ppp1r13l* is also responsible for an autosomal recessive disorder in cattle called Cardiomyopathy and woolly haircoat syndrome (CWH) (264). This mutation was also associated with cardiac and haircoat abnormalities which usually resulted in death before 12 weeks of age. In addition, CWH homozygotes showed neonatal ocular keratitis. Later, iASPP was shown to be present in cardiomyocytes' intercalated discs, interact with desmoplakin and desmin, and to regulate desmosome integrity in mice (265). Taken together, iASPP is essential for heart, hair and skin development across species. The seemingly high occurrence of spontaneous truncating mutations also suggests that the *PPP1R13L* gene may be a region of genomic instability in some species. It would be interesting to investigate whether this putative instability may be related to *PPP1R13L*'s GC-rich sequence which is linked with polyglutamine expansion diseases such as Huntington's disease and ataxias, for which many causative genes are located on chromosome arm 19q.

However, despite the unequivocal evidence of the non-redundant role of ASPPs in development, the importance of ASPP genes in human disease is unknown. In Chapters 2, 3 and 4, the role of *TP53BP2* and *PPP1R13B* mutations in human disease is investigated.

1.4.6 ASPPs in cancer

As discussed above, ASPPs enhance the pro-apoptotic function of the most prominent cancer gene, *TP53*. ASPP1 and ASPP2 also stimulate the pro-senescence and pro-apoptotic activity of the most frequently mutated proto-oncogene, *KRAS*. Together with the *in vivo* data showing spontaneous tumours in *Trp53bp2*$^{+/-}$ mice at a significantly increased rate over wild type mice (104, 105), *TP53BP2* is a candidate cancer gene. However, *TP53BP2*'s role as a driver gene has not been demonstrated in humans, with the single exception of a statistically significant somatic mutation frequency of *TP53BP2* in a recent genomic study of pancreatic cancer (266).

Other studies were limited by small patient sizes, such as the report of a truncating mutation in *TP53BP2* in gastric cancer (267), or were performed in cell lines or model organisms. Studies in cell lines have mostly reached a consensus of a tumour suppressive function for ASPP2, usually attributed to activation of apoptosis (247, 250, 268-273). Studies with human patient samples observed mostly decreased ASPP2 levels (274-280), but occasionally increased expression, for example in endometrial cancer (281, 282).

PPP1R13B promoter is hypermethylated in acute lymphoblastic leukaemia and hepatocellular carcinoma (277, 283). ASPP1 has also been shown to suppress tumorigenesis in the haematopoietic lineage in mice by restricting the self-renewal capacity of haematopoietic stem cells (HSCs), and inducing apoptosis in DNA-damaged HSCs (284). Concomitant deletion of both *Trp53* and *Ppp1r13b* resulted in significantly more aggressive malignancies than the deletion of either gene alone (284). Interestingly, despite this evidence of ASPP1's tumour

suppressive function *in vivo*, and its role in enhancing p53-dependent apoptosis, ASPP1 was reported to limit apoptosis in human cell lines by inhibiting YAP phosphorylation by LATS1 (238). Activation of YAP by ASPP1 also lowers levels of LATS2, reducing p53's ability to induce p21 expression (285). LATS2, in turn, promotes nuclear localisation of ASPP1 under oncogenic stress, and ASPP1-dependent activation of apoptosis (286). This suggests that under steady-state conditions, ASPP1 promotes cell survival via control of the Hippo pathway, whereas under oncogenic stress, ASPP1's pro-apoptotic nuclear function is activated by LATS2. Taken together, although ASPP1 shows hallmarks of a tumour suppressor in mice and patient samples, it appears to have both pro-survival and pro-apoptotic functions at the molecular level. It is currently unknown which of these is the dominant function of ASPP1 in human cancers.

iASPP, being an inhibitor of p53, is a candidate oncoprotein. A study aimed at refining the association of the 19q13.3 chromosomal band with cancer has linked the *PPP1R13L* locus with breast cancer risk (287). A missense polymorphism in *PPP1R13L* (encoding A67T) has been linked with lung cancer in China (288), and another SNP linked with rectal cancer (289). The lung cancer association of a haplotype containing the iASPP SNP rs1970764(G) has been reproduced in another study (290). Lack of association with *PPP1R13L* SNPs and the associated haplotype have been reported for colorectal cancer (291, 292).

Increased iASPP levels have been observed in colorectal cancer (293), acute leukaemia (294), lobular breast carcinoma (295), head and neck squamous cell carcinoma (296), and cervical cancer (297, 298). Cellular studies support a role for iASPP in inhibiting apoptosis, preventing senescence, and conferring chemoresistance on cancer cell lines (299-307). In melanoma, CDK1-

phosphorylated nuclear iASPP mediates inhibition of wild-type p53 (257). Taken together, iASPP shows hallmarks of an oncogene, and has been linked with human cancer through GWAS. However, it remains unclear whether driver mutations in *PPP1R13L* exist in human cancer, and whether iASPP acts as an oncogene across all tumour types.

Chapter 4 of this thesis employs an unbiased approach to assessing the role of ASPP genes in human cancer, then explores the possible mechanisms of ASPP dysregulation in tumours.

CHAPTER 2: ASPP1 AND ASPP2 VARIANTS IN HUMAN CONGENITAL DISORDERS

RESULTS AND DISCUSSION

Figures 2.1-2.39 and the associated text were part of the following publication:

Zak J., Vives V., Szumska D., Vernet A., Schneider J. E., Miller P., Slee E. A., Joss S., Lacassie Y., Chen E., Escobar L. F., Tucker M., Aylsworth A. S., Dubbs H. A., Collins A. T., Andrieux J., Dieux-Coeslier A., Haberlandt E., Kotzot D., Scott D. A., Parker M. J., Zakaria Z., Choy Y. S., Wieczorek D., Innes A. M., Jun K. R., Zinner S., Prin F., Lygate C. A., Pretorius P., Rosenfeld J. A., Mohun T. J., Lu X. ASPP2 deficiency causes features of 1q41q42 microdeletion syndrome. *Cell Death Differ.* **2016,** 23(12):1973-1984.

2.1 *TP53BP2* deletions are associated with brain abnormalities in 1q41q42 microdeletion syndrome

To identify the gene(s) responsible for structural brain abnormalities in 1q41q42 microdeletion syndrome, we first established the smallest region of overlap (SRO) in 1q41q42 microdeletion cases with structural brain abnormalities. To date there are 31 reported 1q41q42 microdeletion cases and 26 of them have brain MRI/CT recorded. Among the 26 cases with deletions in the 1q41q42 region for which neuroradiology reports and/or MRI scans were available, 19 were from published reports and 7 are newly identified cases (**Tables 2.1, 2.2**) (ref. 44, 98, 100, 308-314). Of these 26 cases, 19 patients showed detectable structural brain abnormalities and their SRO (chr1:223 828 382-224 034 322; hg19) contains the entire coding regions of *CAPN2* and *TP53BP2*, plus exons 1 and 2 of *CAPN8* (**Figure 2.1**) whereas in 4 of the remaining 7 cases without brain abnormalities the *TP53BP2* locus is not affected (**Figure 2.1a and 2.1b**).

CAPN2 and *CAPN8* encode Ca^{2+}-dependent proteases of the calpain family (315). Mice deficient in the orthologous genes *Capn2*, *Capn8* and *Trp53bp2* have been generated previously: *Trp53bp2*-deficient mice exhibit abnormalities in the

central nervous system (106), whereas *Capn2*-deficient and *Capn8*-deficient mice show no abnormalities in the brain or nervous system (316-318). Publicly available data also show that *TP53BP2* is expressed more highly in the brain than all other examined tissues, whereas *CAPN2* is expressed at very low levels in the brain compared to other tissues, and *CAPN8* expression is restricted to the stomach (**Figure 2.2**, www.gtexportal.org, (319)). Additionally, *TP53BP2* is most highly expressed in the ventricular and subventricular zones in the human embryonic brain throughout gestation, consistent with it having a key role in embryonic brain development (**Figure 2.2**) (http://brainspan.org, http://human.brain-map.org) (320, 321). We also found no deletions of *TP53BP2* in publically available copy number data from 2 large cohorts of healthy individuals (19,584 healthy individuals (41, 322); or in an independent pediatric control group of 2,026 (ref. 323)), suggesting that *TP53BP2* deletions may be pathogenic. In contrast, 2 deletions, nsv520609 and nsv521758, spanning multiple exons of *CAPN2*, were detected in the cohort of 2,026 disease-free individuals (323). These observations suggest that *TP53BP2* is the strongest candidate for being the causative gene responsible for brain morphological abnormalities associated with 1q41q42 microdeletions.

Table 2.1 Summary of new and previously published cases of 1q41q42 microdeletions. [a]hypertonia; [b]clinodactyly of 5th digits; ID – intellectual disability; LV – lateral ventricles; N/S – not specified; N/A – not applicable or available.

Published cases

	Shaffer 1-7	Rosenfeld 8-9,12-13	Mazzeu 1, 2	Rice 2006	Slavotinek 2009	Filges 2010	Spreiz 2014	Christensen 2012	Wat 2011	Jun et al, 2013	Au et al, 2014	Rosenfeld 10.11.14
Age at diagnosis	N/A	N/A	N/A	10m	NS	10m	8y	N/S	2y	2m	15y	N/A
Gender	N/A	N/A	N/A	Male	Male	Female	Female	N/S	Male	Female	Female	N/A
Deletion/duplication coordinates (hg19)	N/A	N/A	N/A	1:221,911,605-227,293,265	1:221,288,725-230,748,012	1:223,828,382-229,256,492	1:222,015,102-225,414,828	1:223300702-226350463	1:223,076,895-225,311,293	1:223,104,211-223,287,570	1:224,233,297-224,820,132	N/A
Size of deletion	N/A	N/A	N/A	5.38 Mb	9.46 Mb	5.43 Mb	3.4 Mb	N/S	2.2 Mb	0.18 Mb	0.59 Mb	N/A
***TP53BP2* deleted?**	7/7	YES (4/4)	YES (2/2)	YES	YES	YES	YES	YES	YES	NO	NO	NO
Inheritance	N/A	N/A	*de novo*	N/A	*de novo*	*de novo*	*de novo*	*N/A*	*de novo*	maternal	N/A	N/A
Brain MRI analysed?	2/6	YES	N/A	N/R	N/A	N/A	YES	N/R	YES	YES	YES	N/R
Developmental delay/ID	6/6	4/4	2/2	+	N/S	+	+	N/S	+	N/S	+	2/2
Behaviour problems	2/6	1/4	0/1	-	N/S	-	-	N/S	-	N/S	-	1/2
Hypotonia	1/6	4/4	1/1	+	N/S	-		N/S	+	N/S	-	2/2
Seizures	5/6	3/4	0/2	+	N/S	+	+	N/S	+	+	+	0/2
Brain abnormalities	2/6	4/4	2/2	+	N/S	+	-	+	+	-	-	0/2
Enlarged/dysmorphic LVs	3/6	4/4	1/2	-	N/S	+	-	+	+	-	-	0/1
Corpus callosum abnormalities	1/2	2/4	1/2	-	N/S	+	-	+	-	-	-	0/1
Encephalocele	0/7	0/4	0/2	-	-	-	-	-	-	-	-	0/3
Microcephaly	4/7	2/4	0/2	-	N/S	-	-	N/S	+	N/S	-	1/2
Growth retardation	3/5	1/4	2/2	+	N/S	-	-	N/S	-	N/S	+	1/2
Craniofacial abnormalities	7/7	4/4	2/2	+	+	+	+	+	+	-	+	2/2
Unusual iris	N/A	1/4	0/2	+	N/S	-	N/A	N/S	-	N/S	-	0/2
Strabismus	1/6	1/4	1/1	+	N/S	-	-	N/S	-	N/S	-	1/2
Full or tented lips	4/7	2/4	2/2	+	N/S	+	+	N/S	-	N/S	+	1/2
Congenital diaphragmatic hernia	2/7	0/4	0/2	-	+	-	-	N/S	+	N/S	-	0/3
Congenital heart defects	1/7	2/4	2/2	-	+	+	-	+	-	N/S	-	1/3
Pectus deformity	1/7	0/4	0/2	+	N/S	-	-	N/S	-	N/S	-	0/2
Male genital abnormalities	0/5	1/3	1/2	+	+	N/A	N/S	N/S	+	N/S	N/A	0/1
Limb shortening	2/7	0/4	2/2	-	N/S	-	-	N/S	-	N/S	-	0/2
Clubfoot	3/7	1/4	1/2	-	+	-	-	N/S	-	N/S	-	0/3
Short fingers	2/7	0/4	2/2	-	N/S	-	-	N/S	-	N/S	-	0/2
Nail hypoplasia	3/7	1/4	1/2	-	N/S	+	+	N/S	-	N/S	+	0/2
Pelger-Huet anomaly	1/2	1/4	N/S	N/S	N/S	+	N/S	+	N/S	N/S	N/S	0/1

New cases

	1q41q42 microdeletion postnatal					1q41q42 microdeletion prenatal	
	Case 1 (DECIPHER 263059)	**Case 2 (DECIPHER 279742)**	**Case 3 (DECIPHER 257458)**	**Case 4 (DECIPHER 271297)**	**Case 5 (DECIPHER 266948)**	**Case 6 (DECIPHER 257256)**	**Case 7**
Age at diagnosis	2y	7y	9y	9y	7m	Prenatal	Prenatal
Gender	Male	Male	Male	Female	Female	Male	Female
Deletion/duplication coordinates (hg19)	1:222,075,911-228,894,743	1:222,694,309-224,069,984	1:222,387,076-224,599,205	1:221,654,334-224,660,615	1:222,694,079-227,147,000	1:222,732,707-224,034,322	1:221,484,568-229,116,929
Size of deletion	6.8 Mb	1.38 Mb	2.21 Mb	3.0 Mb	4.45 Mb	1.30 Mb	7.6 Mb
***TP53BP2* deleted?**	YES	YES	YES	YES	YES	YES	YES
Inheritance	*de novo*	paternal	*de novo*	N/A	*de novo*	maternal	N/A
Brain MRI analysed?	N/A	N/R	N/R	YES	YES	N/A	YES
Developmental delay/ID	+	+	+	+	+	N/A	N/S
Behaviour problems	N/S	+	-	+	N/S	N/S	N/S
Hypotonia	a	-	-	-	+	N/S	N/S
Seizures	-	-	-	+	-	N/A	N/A
Brain abnormalities	+	N/A	N/A	+	+	+	+
Enlarged/dysmorphic LVs	+	N/A	N/A	+	+	+	+
Corpus callosum abnormalities	+	N/A	N/A	+	+	N/A	+
Encephalocele	-	-	-	-	-	+	-
Microcephaly	+	-	-	-	-	N/S	N/S
Growth retardation	-	-	-	-	-	N/S	N/S
Craniofacial abnormalities	+	-	+	+	+	N/A	N/A
Unusual iris	-	-	-	-	N/S	N/S	N/S
Strabismus	-	-	-	-	N/S	N/S	N/S
Full or tented lips	+	-	+	+	N/S	N/S	N/S
Congenital diaphragmatic hernia	+	-	-	-	N/S	N/S	N/S
Congenital heart defects	-	-	-	-	N/S	-	N/S
Pectus deformity	-	-	-	-	N/S	N/S	N/S
Male genital abnormalities	-	-	-	-	N/A	+	N/S
Limb shortening	-	-	-	-	N/S	N/S	N/S
Clubfoot	-	-	-	-	N/S	N/S	N/S
Short fingers	-	-	-	-	b	N/S	N/S
Nail hypoplasia	-	-	+	-	N/S	N/S	N/S
Pelger-Huet anomaly		-	-	-	N/S	N/S	N/S

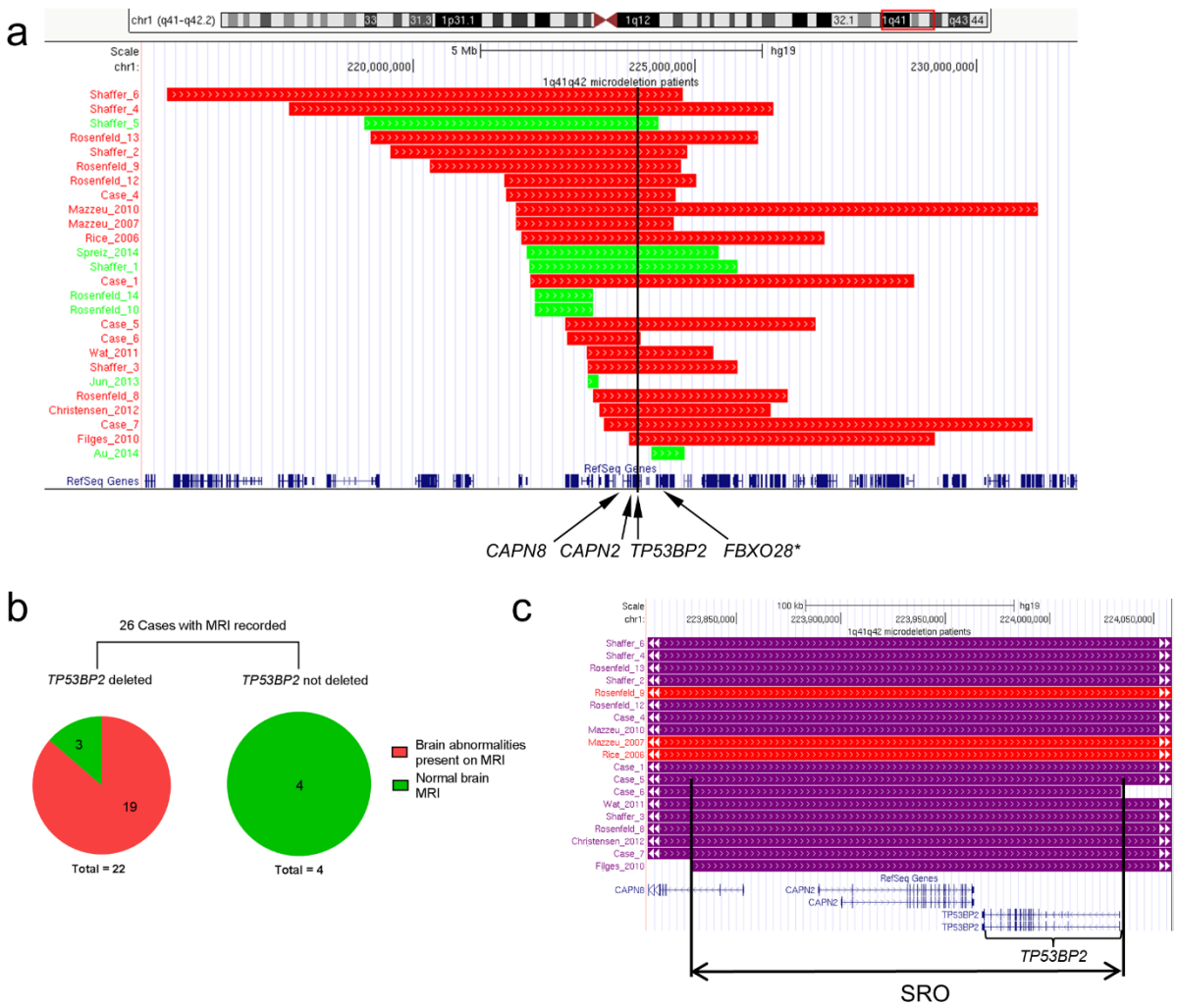
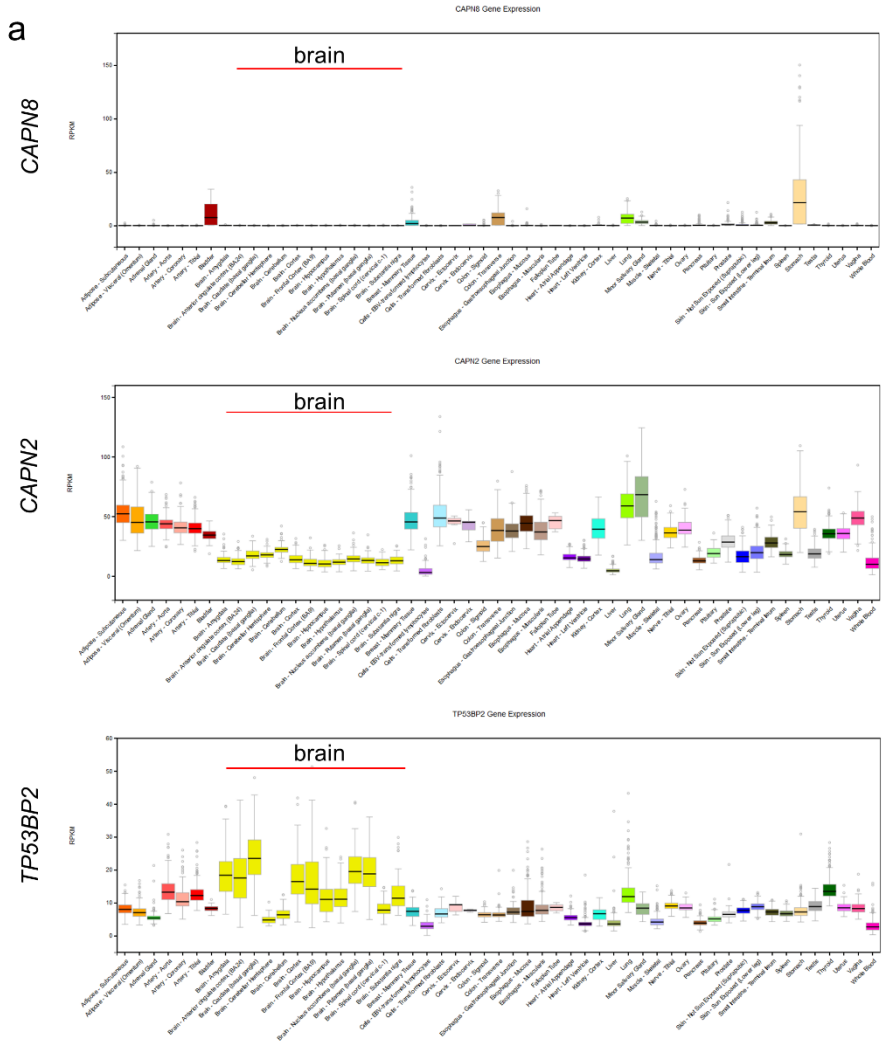


Figure 2.1 *TP53BP2* is deleted in 1q41q42 microdeletion patients with brain abnormalities. (*a*) Genomic locations of 1q41q42 deletions from new and published patients. Red indicates brain morphological abnormalities were reported for the patient, green indicates absence of reported brain abnormalities. The positions of the *CAPN8*, *CAPN2*, *TP53BP2* and *FBXO28* genes are marked with black arrows; black line further marks the position of *TP53BP2*. **FBXO28* is not contained in the smallest region of overlap (SRO) of patient deletions with brain abnormalities. Panel generated using UCSC genome browser (http://genome.ucsc.edu*), hg19 assembly.* (*b*) The proportions of cases with and without *TP53BP2* deletion with and without brain abnormalities. (*c*) Focus on the SRO of patient deletions with brain morphological abnormalities. Patients with abnormal ventricles are marked purple.

a



b

Fetus age	Brain region with highest *TP53BP2* mRNA expression (Allen Brain Atlas)	*TP53BP2* level (z)
15 pcw	ventricular zone in caudal subicular cortex	2.78
16 pcw	ventricular zone in inferolateral temporal cortex	2.99
21 pcw	medial ganglionic eminence	2.21
21 pcw	inner subventricular zone in dorsomedial frontal cortex	3.05

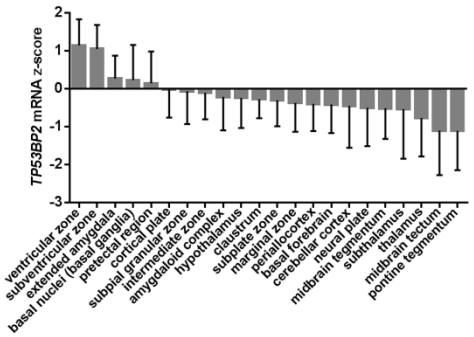


Figure 2.2 Spatial expression patterns of genes in the critical region for brain abnormalities. (a) Gene expression of *CAPN8*, *CAPN2* and *TP53BP2* across human tissues. Data exported from the Genotype Tissue Expression project (GTEX). Brain regions are coloured yellow and highlighted by the red line. RPKM, reads per kilobase of transcript length per million mapped reads. (b) Regions of the prenatal human brain with the highest *TP53BP2* expression. Data mined from the Allen Human Brain Atlas, z-scores shown. **Right**: bar plot of z-scores from regions of the developing brain, combining the data from all sub-regions for each region.

2.2 Brain MRI reveals common ventricular abnormalities among patients with *TP53BP2* deletions

We compared structural brain phenotypes among 1q41q42 microdeletion patients with and without heterozygous *TP53BP2* deletions. Among the 22 *TP53BP2* deletion cases, the most frequently observed brain abnormalities, seen in 16 out of 22 patients (73%) with heterozygous *TP53BP2* deletions, were dilation of ventricles and/or ventricular dysmorphism, frequently associated with hypoplasia or agenesis of corpus callosum (**Table 2.2**). All 4 1q41q42 microdeletion patients without *TP53BP2* deletions showed normal ventricles (**Table 2.2, Figure 2.1b**). Corpus callosum defects were observed in 10/22 (45%) of *TP53BP2*-deletion patients (5 with agenesis and 5 with hypoplasia) but in none of the patients without *TP53BP2* mutations. Cerebellar hypoplasia, encephalocele, hypomyelination, plagiocephaly, holoprosencephaly and malformations of sulcation were also observed in some *TP53BP2* deletion patients at a lower frequency (**Table 2.2**).

For the identified 26 cases of 1q41q42 microdeletion patients, only 14 brain MRI scans were available to us and 12 of them had morphometry-compatible MRI scans (9 with *TP53BP2* deletion and 3 without) (**Figure 2.3a**). We used probabilistic morphometry to quantify the lateral ventricle (LV) volume of the 12 1q41q42 microdeletion patients (**Figure 2.3b**). 28 normal children's brain MRIs from the NIH Paediatric MRI Data Depository were used as controls (**Figure 2.3b**). As expected, LV volume was significantly larger in the *TP53BP2* deletion group (mean 33.4 ml) (**Figure 2.3b**) compared to control group (mean 10.0 ml, SD = 4.5 ml) ($p<0.001$) and the LV volumes of microdeletion patients without *TP53BP2* deletions were within normal range (**Figure 2.3b**).

We observed defects in the ventricles of cases with *TP53BP2* deletions but not in microdeletion cases without *TP53BP2* deletions. For example, in 4 cases with *TP53BP2* deletions the bodies of the lateral ventricles appeared more parallel due to hypoplasia of the corpus callosum (**Figure 2.3c**, Supplementary Table 2: patients Case 4; Shaffer 3**;** Rosenfeld 8; Rosenfeld 13). This is similar but more subtle than the ventricular dysmorphism seen in association with agenesis of corpus callosum. In 3 cases the LVs were asymmetric with one side larger than the other (**Figure 2.3c**: patients Shaffer 3; Rosenfeld 12 and Wat et al 2011). These data show that *TP53BP2* deletions are associated with ventricular abnormalities that are frequently accompanied by hypoplasia of corpus callosum and/or cerebellum.

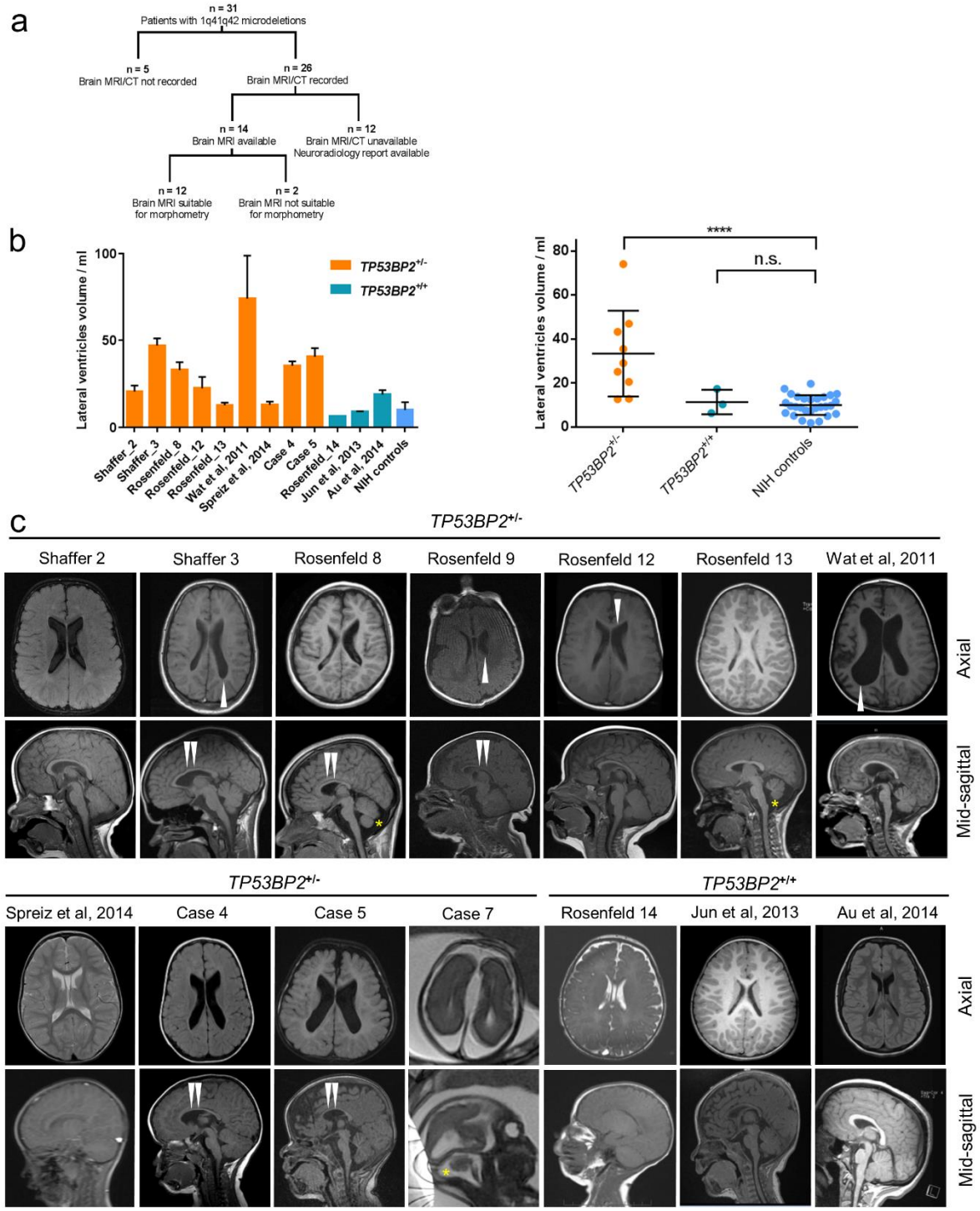


Figure 2.3 Enlarged ventricles and other abnormalities in patients with *TP53BP2* deletions. (*a*) Summary of the patient records analysed in this study. (*b*) Quantification of lateral ventricle volume by probabilistic morphometry. Volumes were computed by ALVIN on all suitable scans in each MRI series, from 1q41q42 microdeletion patients with and without *TP53BP2* deletion. Left panel – T1 and T2 sagittal, axial and coronal scans were used for computations of LV volumes, bars showing mean ± s. d.; right panel – T1 axial scans used only; each data point showing a mean from T1 axial scans for each patient. MRI scans of 12 patients and a group of 28 healthy paediatric individuals (NIH controls) were used. Median age of microdeletion cohort = 19 months, median of NIH controls = 15

months. For statistical comparison of LV volumes in patient groups, values were first tested for normality by the D'Agostino and Pearson omnibus test, then compared to NIH controls using two-tailed student's t-test and corrected for multiple hypothesis testing by multiplying the p-value by the number of tests. Mean ± s. d. shown. (*c*) Axial and sagittal MRI images of brains of 1q41q42 microdeletion patients. T1 non-contrast images are shown, except for cases where only T2 scans were available (Rosenfeld Subject 14, Spreiz et al, 2014 and Case 7). Double white arrowheads: hypoplasia of corpus callosum; single white arrowheads: asymmetry in lateral ventricle size; yellow asterisk: mega cisterna magna.

Table 2.2 Neuroradiological description of patient's brain MRI scans.

Patient	DECIPHER number	*TP53BP2* deleted?	Brain MRI description
Case 1	263059	deleted	Ventricles are obviously dysmorphic with absent corpus callosum. There is a monoventricle, thalamic fusion and holoprosencephaly with a grey matter bridge extending from the frontal lobe towards the occiput. Posteriorly there is some midline differentiation into 2 occipital lobes; microcephaly
Case 2	279742	deleted	N/A
Case 3	257458	deleted	N/A
Case 4	271297	deleted	Lateral ventricles more parallel appearance at their superior aspect; relatively hypoplastic corpus callosum
Case 5	266948	deleted	Generalised cerebral atrophy, especially over the frontotemporal region with large ventricles, thin corpus callosum
Case 6	257256	deleted	Occipital encephalocele
Case 7	-	deleted	Agenesis of corpus callosum, prominent midline interhemispheric fluid, minor cerebellar hypoplasia
Jun 2013	-	not del.	Normal
Au 2014	-	not del.	Normal
Rosenfeld 10	-	not del.	Normal
Rosenfeld 11	-	not del.	N/A
Rosenfeld 14	-	not del.	Normal
Shaffer 1	-	deleted	Normal
Shaffer 2	-	deleted	Mild prominence of lateral ventricles (2y); The ventricular system is top-normal in size, shape and configuration (9y)
Shaffer 3	-	deleted	1992: nonspecific increase in ventricular size; 2006: lateral ventricles somewhat prominent for age; restricted diffusion identified to suggest acute ischemia. Dysmorphic lateral ventricles with left LV being slightly larger than right; The trigones and posterior parts of the bodies of both lateral ventricles are more parallel than usual and slightly more bulbous, this appearance resembles minor colpocephaly which is a configuration associated with agenesis of the corpus callosum. The more lateral aspects of the posterior part of the body of the corpus callosum may be somewhat hypoplastic, leading to the observed ventricular dysmorphism.
Shaffer 4	-	deleted	Malformation of lateral ventricles, pointed frontal horns, fourth ventricle larger than normal; small cerebellum (CT), gyral malformations; malformations of sulcation, particularly around central sulcus
Shaffer 5	-	deleted	Normal
Shaffer 6	-	deleted	Enlarged ventricles prenatally
Shaffer 7	-	deleted	N/A
Rosenfeld 8	-	deleted	Dysmorphism of lateral ventricles, hypoplastic corpus callosum; retrocerebellar and prepontine arachnoid cysts; displacement of the pons away from the clivus by CSF material, mild cortical atrophy, prominent peripapillar atrophy, mega cisterna magna
Rosenfeld 9	-	deleted	Prominent subarachnoid space; Inferior right parietal lobe punctuate focus of abnormal susceptibility; slightly hypoplastic corpus callosum
Rosenfeld 12	-	deleted	Mild thinning of the periventricular white matter, ventricle size mildly prominent; Delayed myelination
Rosenfeld 13	-	deleted	Wide cisterna magna; mild cerebellar atrophy or hypoplasia; dysmorphic appearance of lateral ventricles with a relatively parallel configuration and slight effacement of the supra and lateral aspects of the bodies of the lateral ventricles; somewhat hypoplastic brainstem
Rice	-	deleted	Hypomyelination
Mazzeu 1	-	deleted	Left brain hemiatrophy with cortical dysplasia, possible microgyria at right
Mazzeu 2	-	deleted	Agenesis of corpus callosum, delayed myelination
Slavotinek	-	deleted	N/A
Filges 2010	-	deleted	Agenesis of corpus callosum, colpocephaly
Spreiz 2014	-	deleted	Normal; common incidental anatomical variant of the ventricles in the form a cavum septum pellusidum.
Christensen 2012	-	deleted	Agenesis of corpus callosum, somewhat small cerebellar vermis
Wat 2011	-	deleted	Severe asymmetrical cerebral atrophy with *ex vacuo* dilation of the lateral ventricle

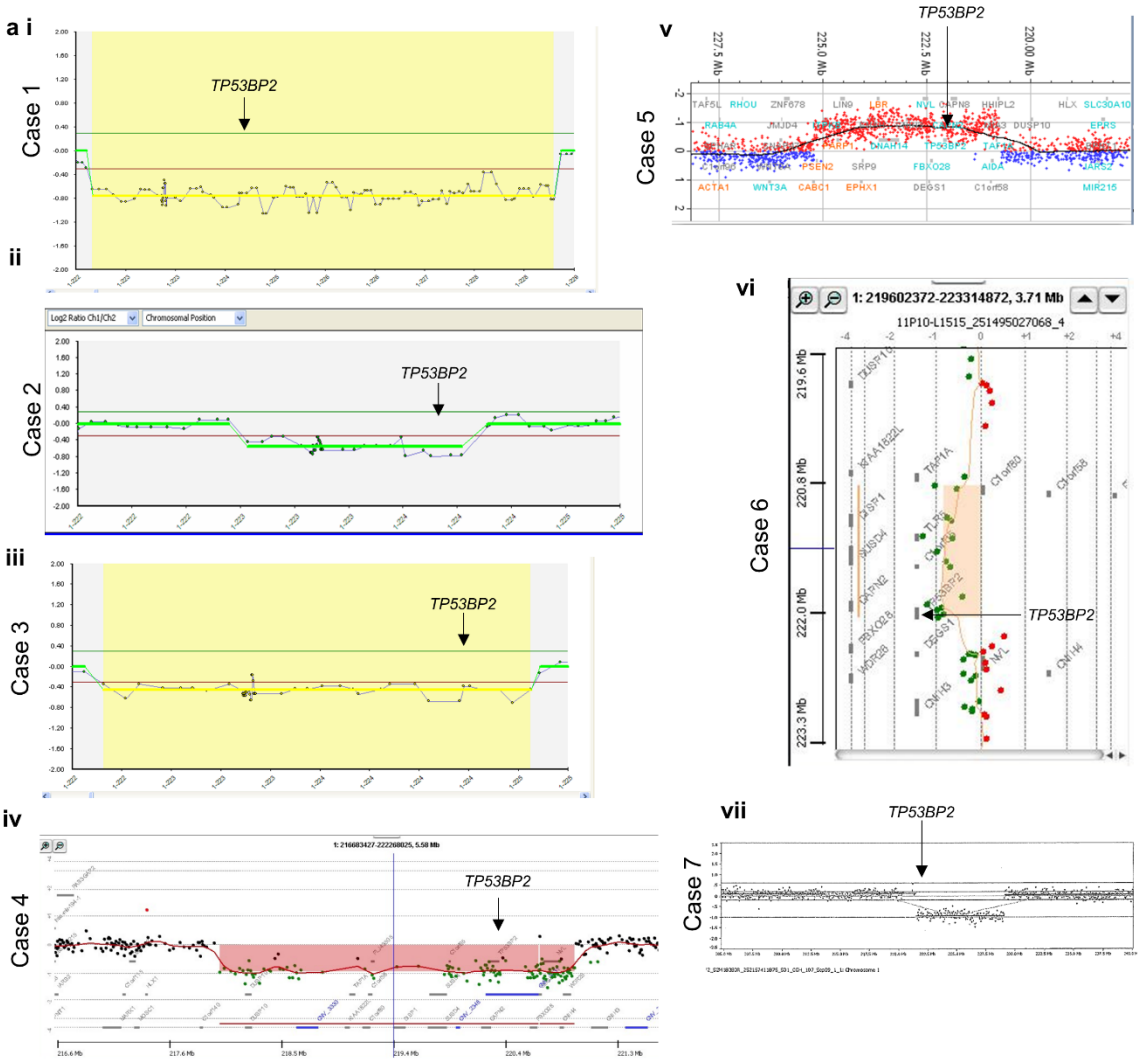


Figure 2.4 Location of chromosomal deletions and duplications of newly described patients. Genomic location of copy number change in patient DNA ascertained by chromosomal microarray. Position of *TP53BP2* is shown with a black arrow. The data are shown are for the following:

- **i.** Case 1 (DECIPHER 263059) – location of deleted region indicated by a yellow line.
- **ii.** Case 2 (DECIPHER 279742) – location of deleted region indicated by probes outside of the marked diploid region.
- **iii.** Case 3 (DECIPHER 257458) – location of deleted region indicated by a yellow line.
- **iv.** Case 4 (DECIPHER 271297) - location of deleted region indicated by red shading.
- **v.** Case 5 (DECIPHER 266948) – location of deleted region indicated by probe positions (red markers unmatched by blue markers).
- **vi.** Case 6 (DECIPHER 257256) - location of deleted region indicated by orange shading.
- **vii.** Case 7 - location of deleted region indicated by probe positions outside the diploidy interval delimited by two lines.

2.3 ASPP2 deficiency in mice causes CNS abnormalities with similar features to 1q41q42 microdeletion patients

We used mice with deletion of *Trp53bp2* exon 3 (*Trp53bp2*$^{\Delta 3}$) to examine how ASPP2 deficiency affects CNS development. We showed previously that the *Trp53bp2*$^{\Delta 3/\Delta 3}$ genotype results in severe developmental defects in a strongly background-dependent manner: BALB/c *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice are born at the expected rate but most develop severe hydrocephalus before weaning, whereas only 30% of mixed C57BL/6-129Sv background (further referred to as 'mixed') *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice survive post-natally (106). On a pure C57BL/6 (referred to as B6) background, which we established by back-crossing over 10 generations, no homozygous *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice survived after birth. The surviving BALB/c *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice have a 100% incidence of hop gait. Most severe cases of neonatal hydrocephalus develop prior to birth and likely result from early defects in brain development.

To provide direct evidence that ASPP2 deficiency causes structural abnormalities in mouse embryonic brain, we used micro-computed tomography (microCT) and high-resolution episcopic microscopy (HREM) to examine *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos. These methods allow reconstruction of 3-dimensional images as well as individual sections, showing the morphology of anatomical defects (324). At embryonic stages E13.5 and E14.5, ~50% of *Trp53bp2*$^{\Delta 3/\Delta 3}$ BALB/c embryos exhibited enlargement of brain ventricles (**Figure 2.5a**). Subcranial and intraventricular haemorrhage were also visible by naked eye and micro-CT in 50% of *Trp53bp2*$^{\Delta 3/\Delta 3}$ BALB/c embryos (**Figure 2.5a**, **Figure 2.6a, b**). Brains of virtually all *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos examined by HREM had absent or

severely hypoplastic pineal gland (**Figure 2.5b**). The spinal canal of all *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos was abnormal with stenosis (**Figure 2.6c**), the presence of blood or small cyst-like structures in the caudal part of the neural tube. All *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos examined had abnormalities in the cerebrospinal fluid space with blockage and/or haemorrhage in at least one part of the ventricular system and/or spinal canal. Heterozygous (*Trp53bp2*$^{\Delta 3/+}$) embryos showed CNS abnormalities, such as spinal canal stenosis or asymmetric lateral ventricles, but with lower penetrance (**Figure 2.5d**). 25% of BALB/c *Trp53bp2*$^{\Delta 3/+}$ embryos showed localised abnormalities in the distribution of neuroepithelial brain tissue associated with small patches of subcranial haemorrhage (**Figure 2.5d**) This shows that ASPP2 is required for early stages of CNS development. Interestingly, surface rendering of lateral ventricles of *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos showed a similar shape to the dysmorphism observed in patients with *TP53BP2* deletion, with bodies of the lateral ventricles being more parallel and bulbous (**Figure 2.5c**). This suggests that ASPP2 deficiency in the *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice might mimic effects of ASPP2 deficiency in humans resulting from the heterozygous deletion of *TP53BP2*.

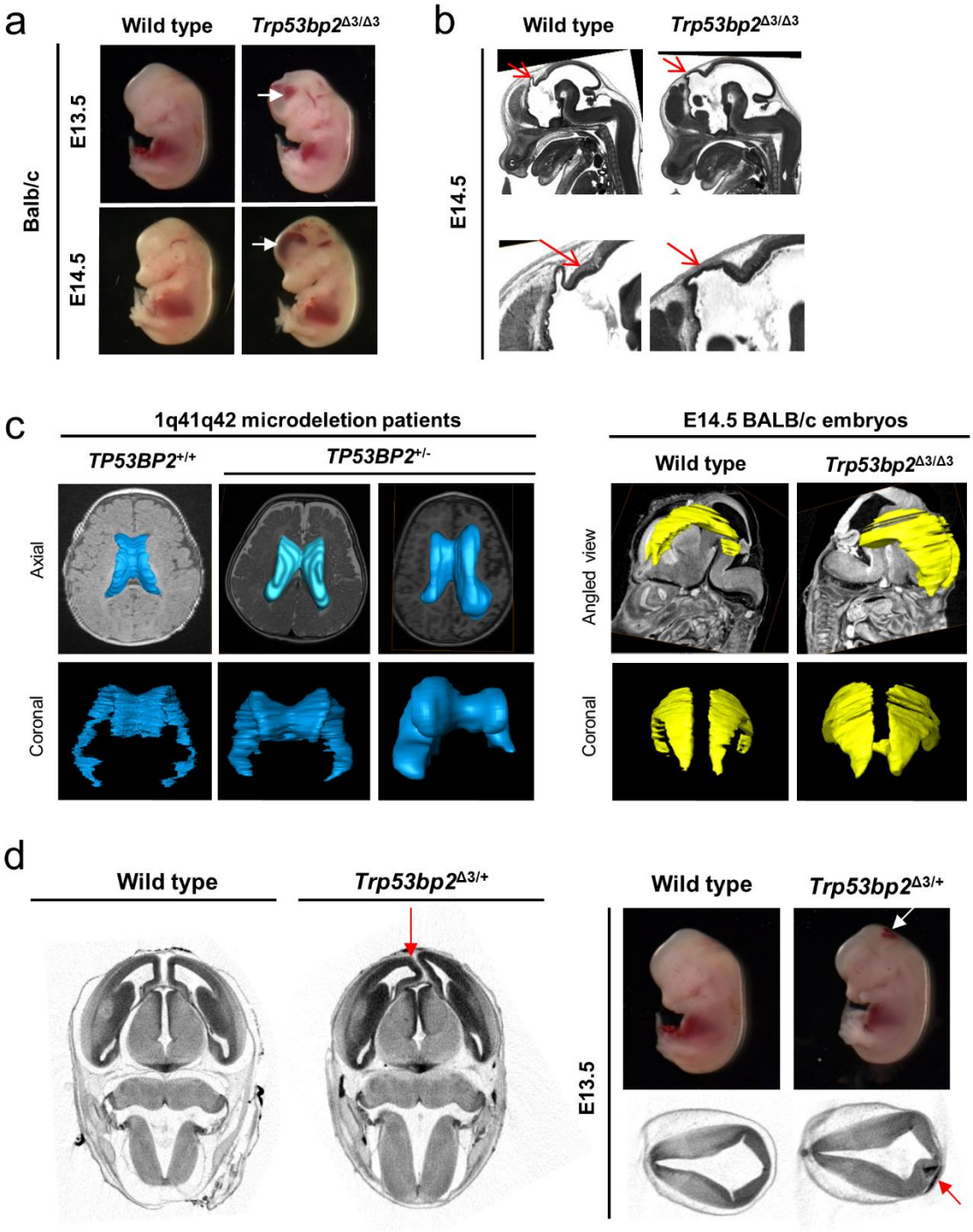


Figure 2.5 *Trp53bp2*$^{\Delta 3/\Delta 3}$ and *Trp53bp2*$^{\Delta 3/+}$ mice have CNS defects. (*a*) Gross dilation of lateral ventricles is visible in *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos, resulting in bulging foreheads. Intraventricular haemorrhage is frequently seen (white arrows). (*b*) Present and absent pineal gland (red arrows) in wild type and *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos, respectively as seen in a single HREM section. (*c*) Surface rendering of lateral ventricles in 1q41q42 microdeletion patients from (left, constructed from MRI scans) and *Trp53bp2*$^{\Delta 3/\Delta 3}$ vs wild-type embryos (right, constructed from microCT scans). (*d*) Brain abnormalities in heterozygous *Trp53bp2*$^{\Delta 3/+}$ embryos. Left: Asymmetry of lateral ventricles in heterozygous *Trp53bp2*$^{\Delta 3/+}$ embryo caused by a structural abnormality (red arrow). MicroCT image shown. Right:

asymmetry in neuroepithelial tissue distribution in the brain (red arrow) and the associated subcranial haemorrhage (white and red arrows) in a BALB/c *Trp53bp2*$^{\Delta 3/+}$ embryo.

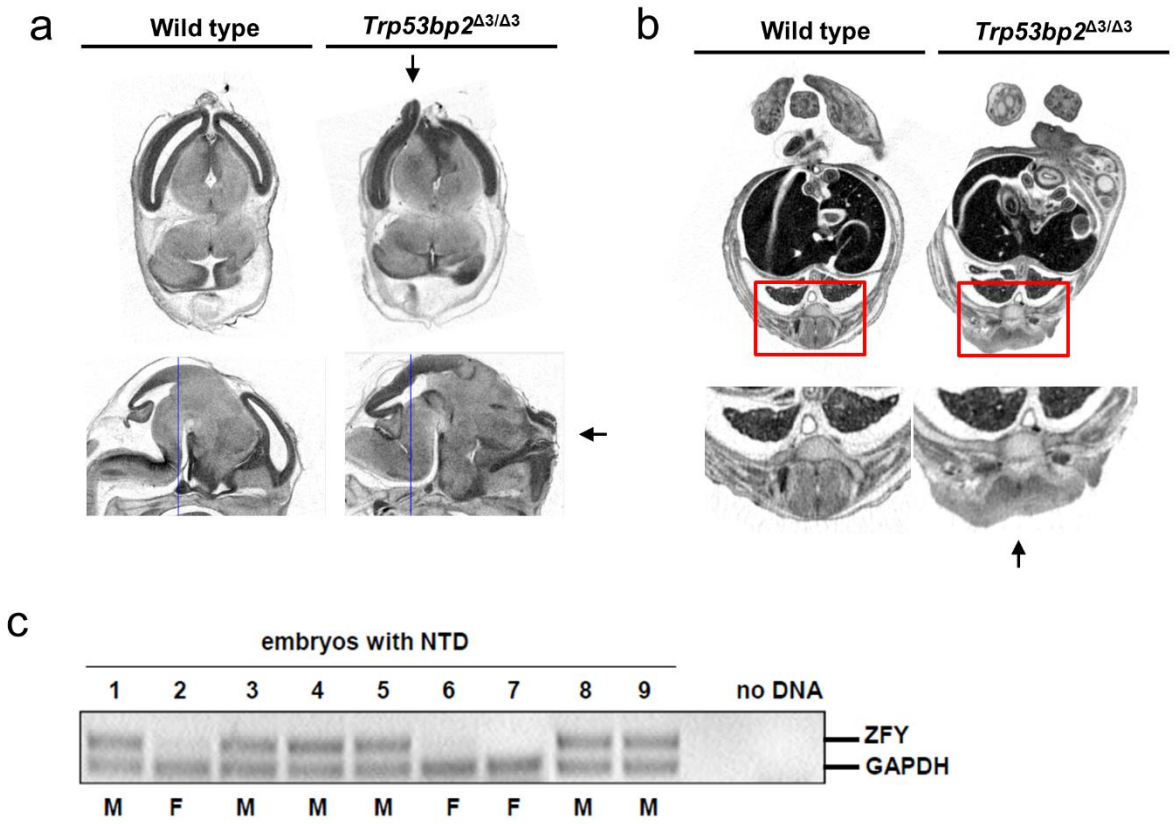


Figure 2.6 Details on CNS phenotypes of *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos. **(a)** Exencephaly in E14.5 *Trp53bp2*$^{\Delta 3/\Delta 3}$ BALB/c embryo: lateral ventricular zone tissue protrudes outside the head, axial (top) and sagittal (bottom) views. MicroCT images shown.
(b) Gross opening of the neural tube as seen in E14.5 *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryo with spina bifida (black arrow). MicroCT images shown.
(c) Gender genotyping of *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos with NTD phenotype. Embryos positive for *ZFY* are male.

2.4 ASPP2 deficiency causes neural tube defects in mice

ASPP2 was reported previously to control the cell polarity and cell proliferation of neural epithelial cells (106). These are two important biological processes involved in neural tube closure, an early event in CNS development the disruption of which causes NTDs. To further investigate the effect of genetic

background on phenotype severity in ASPP2-deficient mice, we compared *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos in mixed, B6 and BALB/c backgrounds. NTDs were detected in 15% (n=127) of embryos on the mixed background, 46% (n=13) on the pure B6 background and 8% (n=12) on the BALB/c background (**Figure 2.7, Table 2.3**). The only NTD observed in BALB/c embryos was exencephaly (**Figure 2.7, 2.6a**). On the mixed background, NTDs were small at the early stages (E9.5-E10.5) but progressed to larger lesions at >E13.5, and featured mainly exencephaly and rostral spina bifida (**Figure 2.7**) but rarely craniorachischisis (**Figure 2.7**-middle panel). In contrast, NTDs in *Trp53bp2*$^{\Delta 3/\Delta 3}$ B6 embryos have a broad spectrum of phenotypes, ranging from spina bifida to craniorachischisis with some spina bifida phenotypes affecting a large section of the neural tube (**Figure 2.7, 2.6b**). MicroCT and HREM inspection of *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos with NTDs revealed occasional overgrowth of neural tissue (**Figure 2.6a**). The spinal column of the neural tube was often wrinkled, even in embryos without NTDs.

Given the association of *TP53BP2* deletions with brain structural abnormalities, we wondered whether *TP53BP2* deletions might also be associated with NTDs in humans. We compared the genomic location of human genes orthologous to genes which have been implicated in NTDs in mice with chromosomal bands reported by Tyshchenko et al. (96) as associated with NTDs in humans. Interestingly, *TP53BP2* is one of only 16 human genes that lie in these significantly associated bands (**Figure 2.8**). Deletions of the 1q41-q44 locus encompassing *TP53BP2* are significantly associated with posterior encephalocele, generally considered a NTD (96, 325, 326). This suggests that *TP53BP2* deficiency might also play a role in a subset of human NTDs.

Notably, we found no bias towards females in *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos presenting NTDs, whereas NTDs in p53-null mice occur preferentially in females (**Figure 2.6c**) (ref. 131). This suggests ASPP2 contributes to NTDs via a p53-independent mechanism**.** In summary, these results demonstrate that 100% of *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice have CNS defects but their phenotypic characteristics vary substantially with the genetic background.

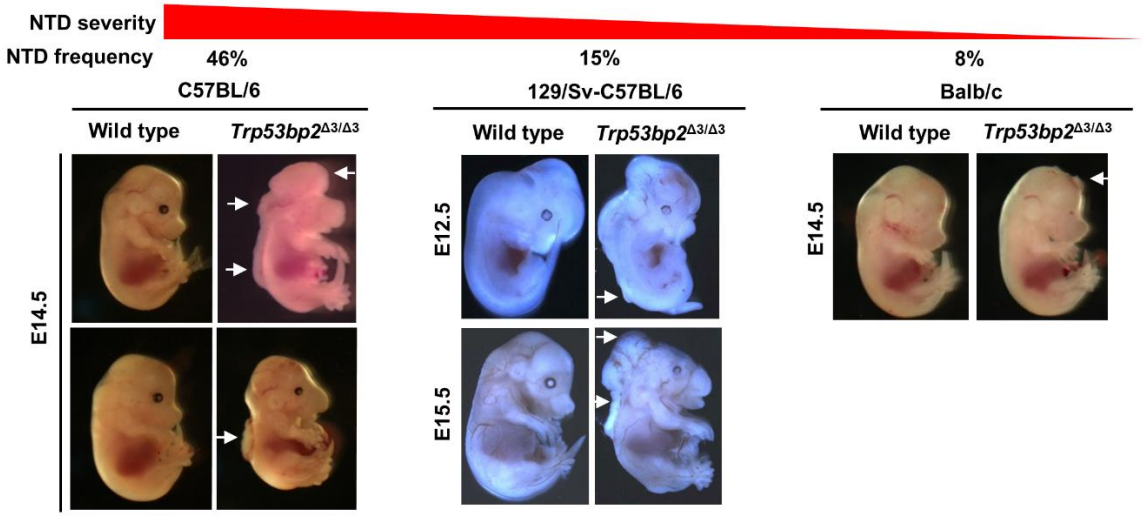


Figure 2.7. *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice have neural tube defects. Strain dependence of NTD penetrance in *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice: Embryo phenotypes in pure C57BL/6 (left), mixed 129Sv - C57BL/6 (middle), and BALB/c (right) background embryos range in severity from small exencephaly (right, white arrow) to severe craniorachischisis (left upper panel, white arrows). Spina bifida is also observed in the mixed (middle panels, white arrows) and B6 (left lower panel, white arrow) backgrounds.

Table 2.3. **Genotype and NTD frequencies of mice derived from *Trp53bp2* heterozygous matings in mixed 129Sv-C57BL/6 ('mixed') and pure C57BL/6 ('B6') backgrounds.** +/+ denotes *Trp53bp2*$^{+/+}$, Δ3/+ denotes *Trp53bp2*$^{\Delta 3/+}$, Δ3/Δ3 denotes *Trp53bp2*$^{\Delta 3/\Delta 3}$. Distribution of +/+, +/Δ3 and Δ3/Δ3 genotypes does not differ significantly from a 1:2:1 ratio among embryos ($\chi^2 = 1.78$, $p > 0.05$, Chi-squared test), whereas there is a significant deviation from this ratio among postnatal mice ($\chi^2 = 198.9$, $p < 0.001$, Chi-squared test).
[a]Percentage of *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos amongst all genotypes.
[b]Percentage of NTDs amongst *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos only.

STAGE	**+/+**	**+/Δ3**	**Δ3/Δ3** [a]	TOTAL	NTDs [b]
Prenatal mixed					
E9.5	6	27	16 (33%)	49	2 (13%)
E10.5	12	28	17 (30%)	57	4 (24%)
E11.5	28	57	22 (21%)	107	2 (9%)
E12.5	5	21	8 (24%)	34	0
E13.5	10	18	10 (26%)	38	2 (20%)
E14.5	4	5	4 (31%)	13	0
E15.5	21	44	18 (22%)	83	5 (28%)
E16.5	11	21	14 (30%)	46	3 (21%)
E17.5	6	20	7 (22%)	33	0
E18.5	5	7	10 (32%)	22	1 (10%)
TOTAL	112	260	127 (26%)	499	19 (15%)
Postnatal mixed					
P5-P30	353	644	68 (6.4%)	1065	0
Prenatal B6					
E14.5	14	34	13 (21%)	61	6 (46%)

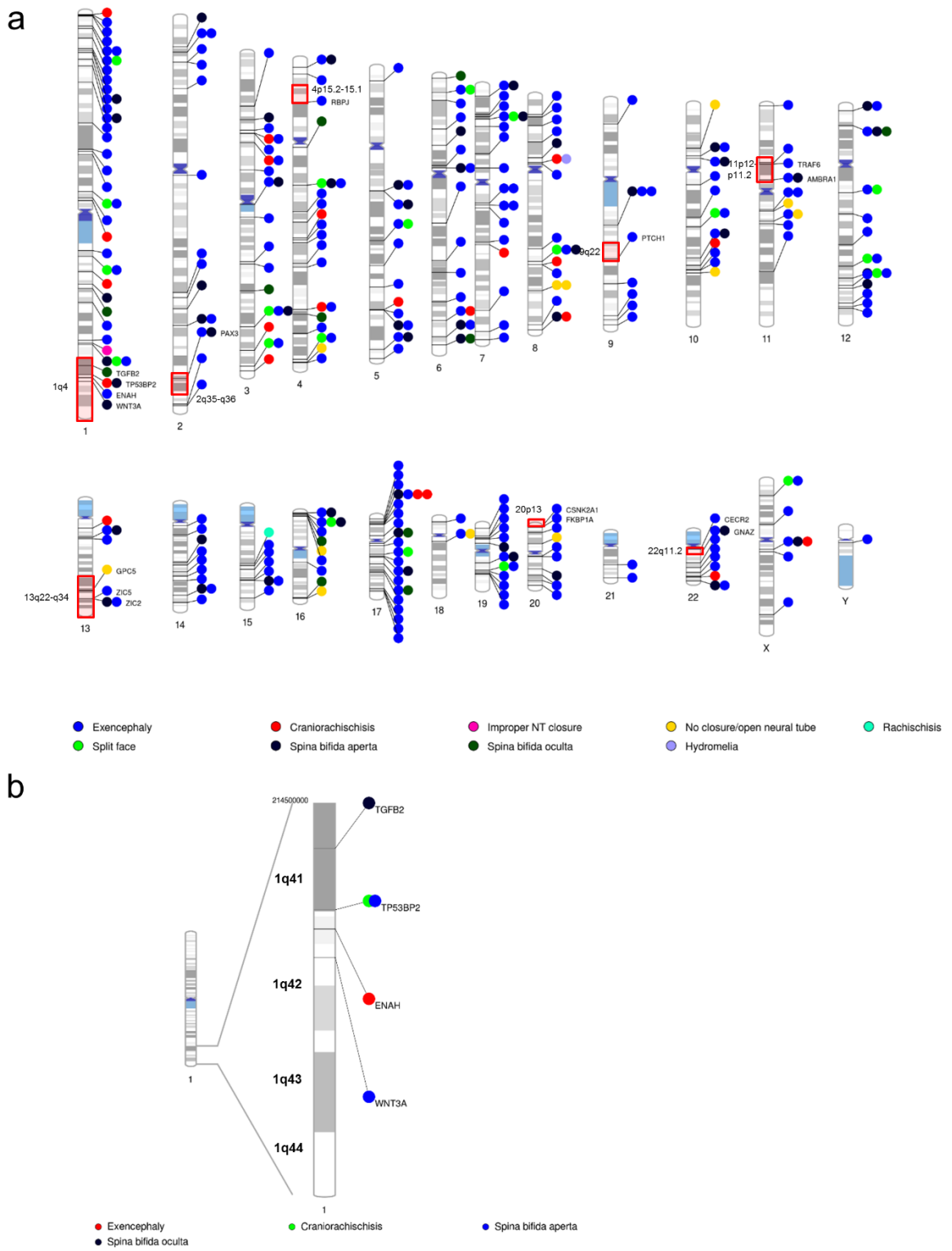


Figure 2.8 Few genes that cause NTDs in mice lie in regions whose deletion is associated with human NTDs. (a) Positions of genes in the human genome that are orthologous to genes that cause NTDs when mutated in mice. Each gene is indicated as a black line; coloured dots mark the associated NTD phenotype in mice. Cytogenetic chromosomal bands are marked in grey and white; regions with heterochromatin in light blue. The location of bands the deletion of which is associated with NTDs in humans marked with red lines and labelled. (b) Focus on the 1q4 region, which is significantly

associated with NTDs in humans. Genes are marked that cause NTDs when deleted in mice.

2.5 Non-CNS phenotypes overlap between *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos and 1q41q42 microdeletion syndrome patients

Visual inspection of *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos also revealed craniofacial and eye abnormalities (**Figure 2.9a, b**). We reported previously that ASPP2-deficient embryos have highly penetrant retinal abnormalities (106). Some B6 *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos lacked one or both eyes (**Figure 2.9a**) while others showed coloboma (incomplete eye tissue) or microphthalmia (abnormally small eye(s)) (**Figure 2.9b**).

We used microCT and HREM to examine *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos comprehensively for developmental abnormalities. At E13.5 and E14.5, 67-100% of B6 and 33-56% of BALB/c *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos had abnormal heart position involving a 30-45^{o} twist along the coronal axis (**Figure 2.9c**) and 67% B6 and 33% BALB/c of *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos had ventricular septal defect (VSD, **Figure 2.9d**). Heart function and chamber volumes of adult BALB/c *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice were normal (**Figure 2.10a**). *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos in B6 and mixed backgrounds showed various craniofacial abnormalities including dolichocephaly (an elongation of the head), micrognathia (smaller lower jaw) (**Figure 2.9a, Figure 2.11**) and cleft palate (**Figure 2.9e**). A high percentage of embryos also showed gonadal abnormalities resulting in unclear gender at E14.5 (**Figure 2.9f**), urethral abnormalities (**Figure 2.9g**), abnormal trigeminal nerve and spinal ganglia (**Figure 2.10b**) and multiple other defects of varying penetrance (Figure 6). Notably, 33% of BALB/c *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos lacked a visible abducens nerve.

Embryos with heterozygous deletion of exon 3 (*Trp53bp2*$^{\Delta 3/+}$) examined by microCT had 30% incidence of cleft palate in the B6 background, and 30% and 13%

incidence of VSD in B6 and BALB/c backgrounds, respectively. Recently, the International Mouse Phenotyping Consortium (IMPC, www.mousephenotype.org) – who are performing broad phenotyping for 20 000 mouse genes (3 154, to date) - released phenotypic data on mice with a *Trp53bp2*-exon4-deletion genotype, termed *Trp53bp2*$^{tm1b(EUCOMM)Hmgu}$ (327). Our analysis of this publicly available data showed that the homozygous deletion of *Trp53bp2* exon 4 (*Trp53bp2*$^{\Delta4/\Delta4}$) is lethal, with 100% penetrance prior to embryonic day 9 (**Figure 2.12a**). *Trp53bp2*$^{\Delta4/+}$ mice were viable and males showed a significant hyperactive phenotype in the open field test (**Figure 2.12b**). The *Trp53bp2*$^{\Delta4/+}$ males showed no difference in the percentage of time spent in the centre of the open field, indicating the phenotype is highly specific to locomotor activation (**Figure 2.12b**). Of note, these mice also showed lower grip strength normalized by body weight, short tibia, higher bone density and an abnormal glucose response (**Figure 2.12c-e**). These data from the heterozygotes demonstrate that ASPP2 is haploinsufficient in mouse development.

There is substantial overlap in the phenotypic spectrum – including brain, neurological, craniofacial, eye, heart and urogenital defects - of ASPP2-deficient mice and patients with deletions involving 1q41q42. 50% of patients with *TP53BP2*-deletion had cleft palate, which was observed in 33% and 67% of BALB/c and B6 *Trp53bp2*$^{\Delta3/\Delta3}$ embryos, respectively. Heart abnormalities, including VSD, were reported for 31% patients with *TP53BP2*-deletion, and 30% B6 *Trp53bp2*$^{\Delta3/+}$ embryos had VSD. 1q41q42 microdeletion patients also frequently present with hypotonia and hypoglycemia with hypoglycemic seizures, representing phenotypic overlap with the reduced grip strength and abnormalities in glucose response in *Trp53bp2*$^{\Delta4/+}$ mice (see above). Overall, the relative penetrance of defects in *Trp53bp2*$^{\Delta3/\Delta3}$ embryos also broadly corresponds to the frequency of their

occurrence in 1q41q42 deletion patients: virtually all diagnosed cases with 1q41q42 deletions have intellectual disability and/or brain structural abnormalities and there is high penetrance of CNS abnormalities in *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice, with lower proportions of 1q41q42 deletion patients and *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice having the other phenotypes. These results suggest that *TP53BP2* haploinsufficiency might be implicated in additional abnormalities associated with the 1q41q42 microdeletion syndrome and 1q41q42 deletions more generally.

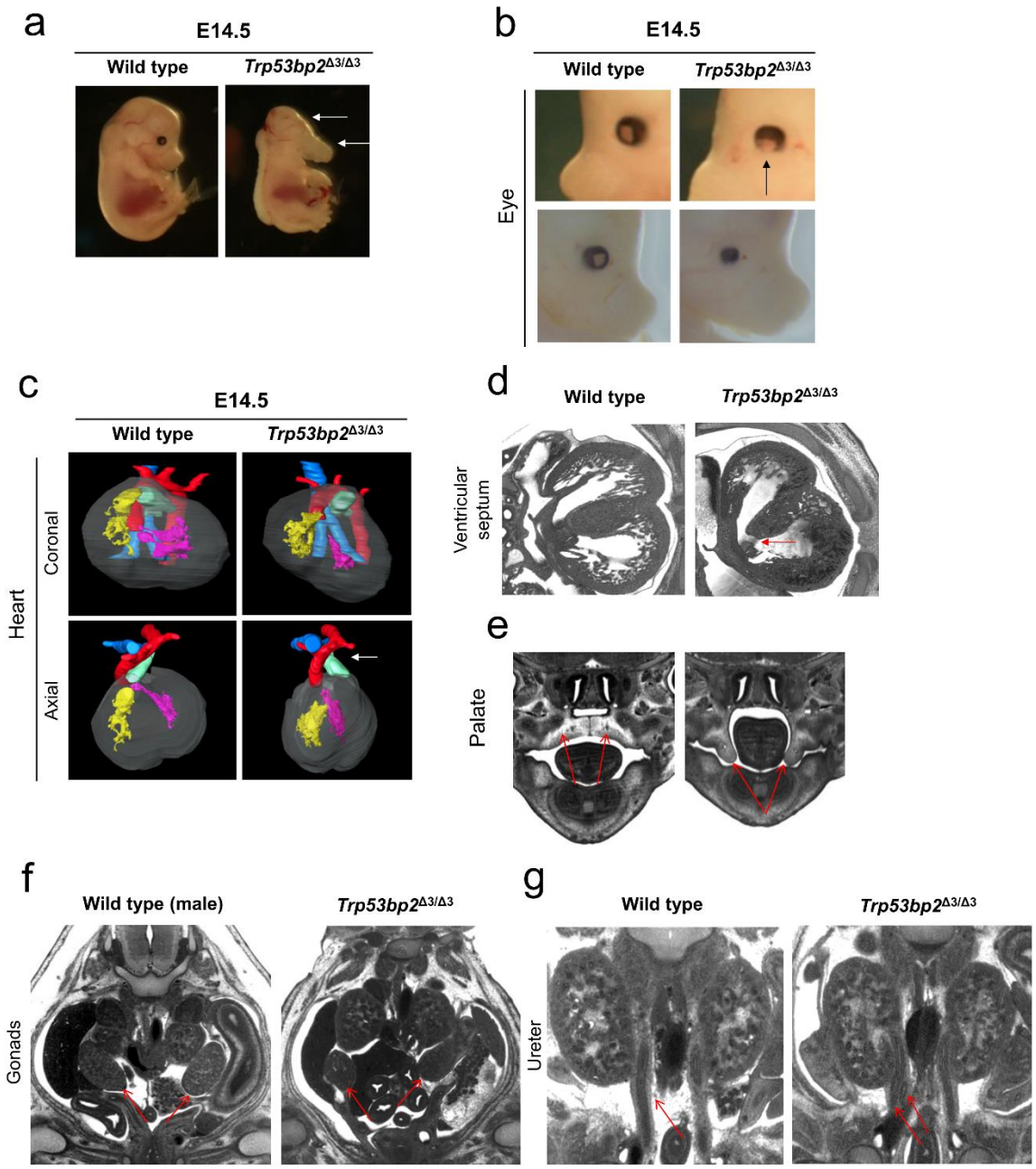


Figure 2.9 ASPP2-deficient mice have non-CNS phenotypes. (a) Craniofacial abnormalities in *Trp53bp2*$^{\Delta 3/\Delta 3}$ E14.5 embryos include dolichocephaly and hypoplasia or agenesis of the lower jaw (white arrows) (*b*) Eye abnormalities observed in *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice feature coloboma (top, missing tissue indicated by red arrow) and microphthalmia (bottom) in E14.5 embryos. (*c*) Abnormal heart position is seen in all B6 embryos examined. The microCT-based 3D reconstruction shows the heart is twisted approximately 45 degrees along the coronal axis (top panel), sometimes resulting in an altered direction of the ductus arteriosus with respect to the aorta (axial view, white arrow). Colour coding: yellow – right ventricle; pink – left ventricle; blue trachea; red – aorta; light green – ductus arteriosus. (*d*) Ventricular septal defect in *Trp53bp2*$^{\Delta 3/\Delta 3}$ E14.5 embryonic heart (red arrow); shown on a representative HREM section. (*e*) Palate (red arrows) showing cleft palate in *Trp53bp2*$^{\Delta 3/\Delta 3}$ E14.5 embryo. (*f*) Largest-area HREM section of gonads (red arrows) showing abnormalities and unclear gender in a *Trp53bp2*$^{\Delta 3/\Delta 3}$ E14.5 embryo vs male E14.5 wild-type embryo. (*g*) Single vs double ureter (red arrows) in wild type vs *Trp53bp2*$^{\Delta 3/\Delta 3}$ E14.5 embryos, respectively as visible by HREM.

a

Measurement	WT (n=7)	Δ3/Δ3 (n=8)	p value
Mean age, weeks	23.2 ± 2.5	22.7 ± 2.6	0.95
Body weight, g	25.7 ± 1.4	23.8 ± 1.19	0.33
Heart rate, BPM	416 ± 18	412 ± 18	0.89
LV end-diastolic volume, µL	51 ± 3	52 ± 2	0.80
LV end-systolic volume, µL	25 ± 3	28 ± 2	0.44
Stroke volume, µL	26 ± 2	24 ± 1	0.36
Ejection fraction, %	52 ± 4	47 ± 3	0.29
Cardiac output, mL/min	11.0 ± 0.9	10.2 ± 0.8	0.49
LV wall thickness, mm	0.72 ± 0.04	0.72 ± 0.01	0.99
Myocardial cross-sectional area, mm^2	9.9 ± 0.5	9.6 ± 0.3	0.58
Pulmonary artery VTI, mm	29 ± 2	31 ± 1	0.48
Pulmonary acceleration time, ms	21 ± 2	18 ± 1	0.23
Pulmonary ejection time, ms	71 ± 4	70 ± 2	0.73

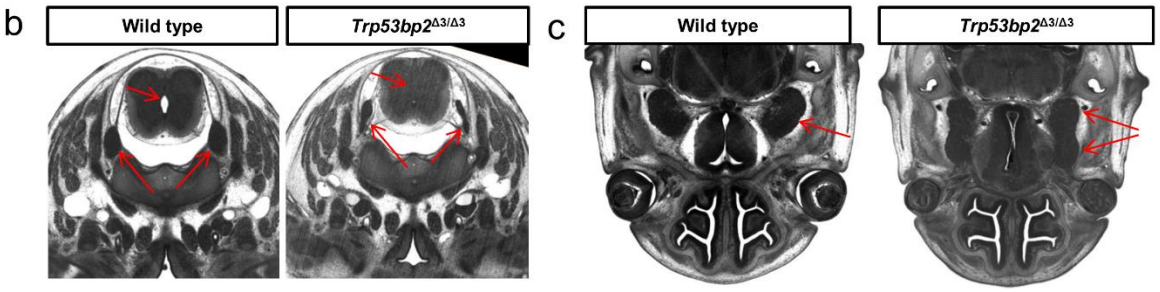


Figure 2.10 Abnormalities in *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos and heart function in *Trp53bp2*$^{\Delta 3/\Delta 3}$ adult mice. (a) Heart function in *Trp53bp2*$^{\Delta 3/\Delta 3}$ vs wild type adult BALB/c mice examined by echocardiography. LV is left ventricular, VTI is velocity time integral. Values are mean ± SEM. (b) Small and elongated 1st pair of spinal ganglia (pair of red arrows) in a *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryo; spinal canal stenosis (single red arrow). HREM images shown. (c) Significantly enlarged trigeminal ganglia (red arrows) in *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryo; HREM images shown.

Visual inspection

	Frequency in *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos		
	C57BL/6	C57BL/6-129Sv	Balb/c
Neural tube defects	6/13 (46%)	19/127 (15%)	1/12 (8.3%)
Abnormal head	12/13 (92%)		6/12 (50%)
Intracranial haemorrhage	4/13 (31%)		6/12 (50%)
Eye abnormalities (macroscopic)	10/13 (77%)		3/3 (100%)
Uni/bi-lateral eye agenesis	5/13 (38%)		0/12 (0%)
Craniofacial abnormalities	9/13 (69%)		0/6 (0%)

microCT

Genotype of embryos (*Trp53bp2*)	Δ3/Δ3		Δ3/+	
	C57BL/6	BALB/c	C57BL/6	BALB/c
CNS abnormalities	3/3 (100%)	9/9 (100%)	2/9 (22%)	4/8 (50%)
Neural tube defects	3/3 (100%)	1/9 (11%)	0/10 (0%)	0/8 (0%)
Enlarged lateral ventricles	0/3 (0%)	6/9 (67%)	0/9 (0%)	0/8 (0%)
Dysmorphic lateral ventricles	3/3 (100%)	8/9 (89%)	1/9 (11%)	2/8 (25%)
Intraventricular haemorrhage	0/3 (0%)	3/9 (33%)	0/10 (0%)	0/8 (0%)
Spinal canal abnormalities	3/3 (100%)	9/9 (100%)	2/10 (20%)	0/8 (0%)
Brain structural abnormalities	3/3 (100%)	9/9 (100%)	2/9 (22%)	4/8 (50%)
Eye abnormalities	3/3 (100%)	9/9 (100%)	3/9 (33%)	1/7 (14%)
Uni/bi-lateral eye agenesis	1/3 (33%)	0/9 (0%)	0/10 (0%)	0/8 (0%)
Retinal abnormalities	3/3 (100%)	9/9 (100%)	3/9 (33%)	1/7 (14%)
Heart abnormalities	3/3 (100%)	6/9 (67%)	3/10 (30%)	1/8 (13%)
Abnormal heart position	3/3 (100%)	5/9 (56%)	0/10 (0%)	0/8 (0%)
Ventricular septal defect	2/3 (67%)	3/9 (33%)	3/10 (30%)	1/8 (13%)
Craniofacial abnormalities	3/3 (100%)	6/9 (67%)	3/10 (30%)	0/8 (0%)
Hypoplastic/absent lower jaw	1/3 (33%)	1/9 (11%)	0/10 (0%)	0/8 (0%)
Cleft/abnormal palate	2/3 (67%)	6/9 (67%)	3/10 (30%)	0/8 (0%)
Absent tongue	1/3 (33%)	0/9 (0%)	0/10 (0%)	0/8 (0%)

HREM

	Frequency in *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos	
	C57BL/6	Balb/c
CNS abnormalities	6/6 (100%)	3/3 (100%)
Neural tube defects	2/6 (33%)	0/3 (0%)
Spinal canal abnormalities	5/6 (83%)	3/3 (33%)
Absent pineal gland	6/6 (100%)	3/3 (100%)
Brain structural abnormalities	6/6 (100%)	3/3 (100%)
Eye abnormalities	6/6 (100%)	3/3 (100%)
Uni/bi-lateral eye agenesis	1/6 (17%)	0/3 (0%)
Retinal abnormalities	6/6 (100%)	3/3 (100%)
Heart abnormalities	5/6 (83%)	1/3 (33%)
Abnormal heart position	4/6 (67%)	1/3 (33%)
Ventricular septal defect	4/6 (67%)	1/3 (33%)
Narrow ductus venosus	3/6 (50%)	2/3 (67%)
Narrow umbilical vein	3/6 (50%)	1/3 (33%)
Craniofacial abnormalities	3/6 (50%)	1/3 (33%)
Abnormal mandible	2/6 (33%)	0/3 (0%)
Cleft/abnormal palate	2/6 (33%)	1/3 (33%)
Absent tongue	1/6 (17%)	0/3 (33%)
Gonadal abnormalities	1/6 (17%)	0/3 (0%)
Urethral abnormalities	4/6 (67%)	3/3 (100%)
Double ureter	1/6 (17%)	1/3 (33%)
Absent ureter	1/6 (17%)	0/3 (0%)
Small spleen	0/6 (0%)	1/3 (33%)
Small/abnormal ganglia	5/6 (83%)	1/3 (33%)
Absent abducens nerve (cranial VI)	0/4 (0%)	2/3 (67%)
Abnormal optic nerve	2/6 (33%)	0/3 (0%)
Olfactory bulbs not pronounced well	0/4 (0%)	2/3 (67%)
Small thymus	1/6 (17%)	2/3 (67%)
Inner ear abnormalities	2/6 (33%)	0/3 (0%)
Narrow lumen of duodenum	1/6 (17%)	1/3 (33%)
Enlarged Rathke's pouch and pituitary	1/6 (17%)	0/3 (0%)
Vertebral and ribs abnormalities	1/6 (17%)	0/3 (0%)
Absent sternum	1/6 (17%)	0/3 (0%)

Figure 2.11 Summary of phenotypes in *Trp53bp2*$^{\Delta 3/\Delta 3}$ and *Trp53bp2*$^{\Delta 3/+}$ embryos

Abnormalities detected by naked eye (visual inspection), microCT and HREM in E13.4 and E14.5 ASPP2-deficient embryos. Frequency and the number of embryos evaluated for each phenotype are given. Wild-type littermate embryos were used as reference controls. In the case of BALB/c embryos, eye abnormalities were difficult to determine visually and therefore quantified for a small number of embryos.

a

	Trp53bp2$^{+/+}$	*Trp53bp2*$^{\Delta4/+}$	*Trp53bp2*$^{\Delta4/\Delta4}$	Total
Pups born	31	65	0	96
Embryos at E12.5	16	26	0	42
Embryos at E9.5	17	20	11 (all dead)	48

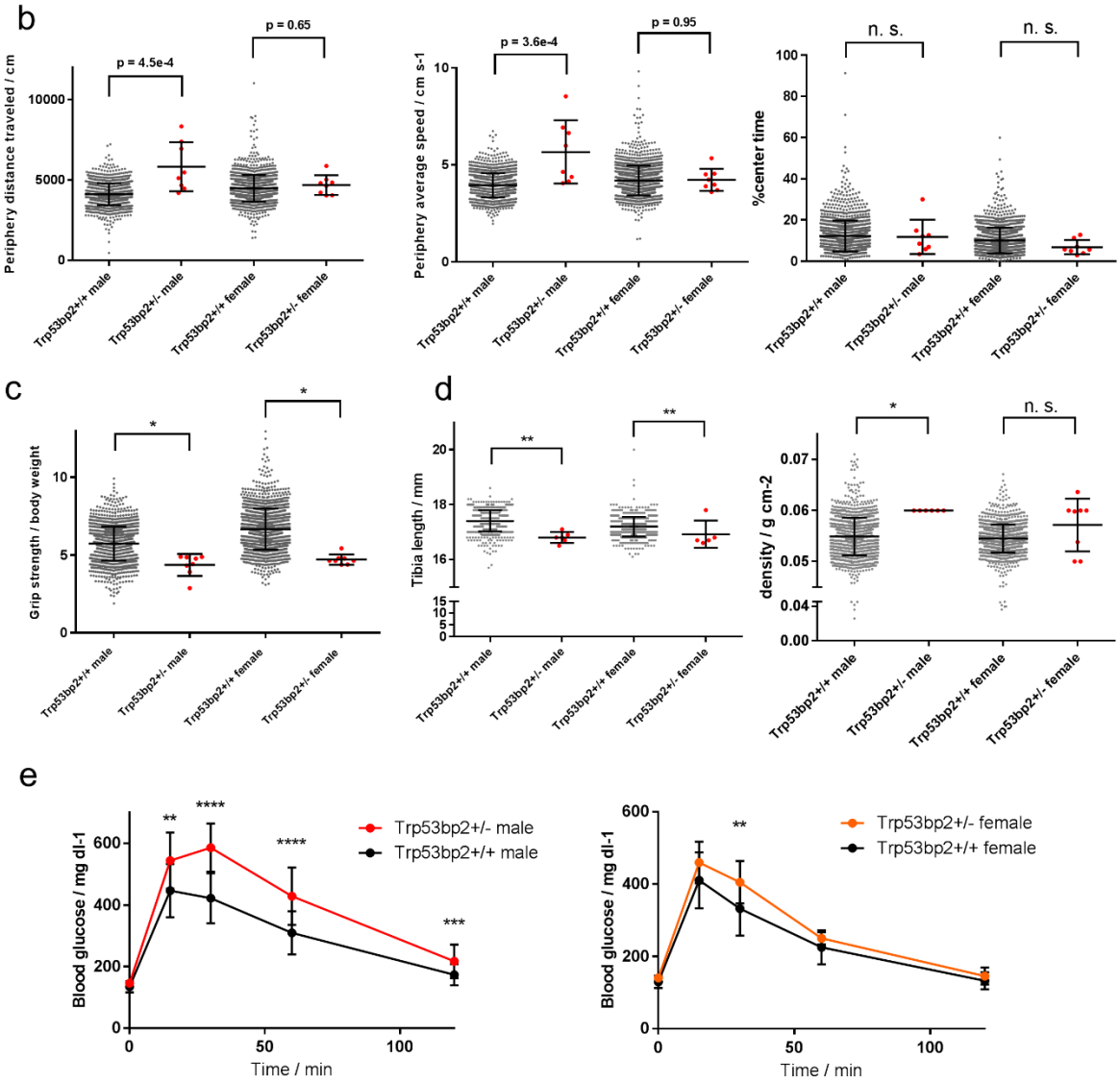


Figure 2.12 Abnormalities shown by adult *Trp53bp2*$^{\Delta4/+}$ mice. (a) Birth rates of wild type, *Trp53bp2*$^{\Delta4/+}$ and *Trp53bp2*$^{\Delta4/\Delta4}$ mice and embryos in heterozygous-heterozygous matings. (b) Results from the Open field test performed on *Trp53bp2*$^{\Delta4/+}$ mice (denoted *Trp53bp2*$^{+/-}$) vs wild type controls (*Trp53bp2*$^{+/+}$): distance travelled in the peripheral part of the arena (left), average speed in the periphery (centre) and % time spent in the centre of the arena. P values were obtained by IMPC's Mixed Model framework, linear-mixed effects model. N. s., not significant. (c) Forelimb and hindlimb grip strength normalised by body weight for *Trp53bp2*$^{\Delta4/+}$ mice (denoted *Trp53bp2*$^{+/-}$) vs wild type controls (*Trp53bp2*$^{+/+}$). P values were obtained by IMPC's Mixed Model framework, linear-mixed effects model. *, $p < 0.05$. (d) Tibia length (left panel) and bone density (right panel) as determined by X-ray and Dual Energy X-ray Absorptiometry, respectively, in *Trp53bp2*$^{\Delta4/+}$ mice (denoted *Trp53bp2*$^{+/-}$) vs wild type controls (*Trp53bp2*$^{+/+}$). P values were obtained by IMPC's Mixed Model framework, linear-mixed effects model. *, $p < 0.05$; **, $p < 0.01$; n. s., not significant. (e) Glucose tolerance

measured by the Intraperitoneal glucose tolerance test in *Trp53bp2*$^{\Delta 4/+}$ mice (denoted *Trp53bp2*$^{+/-}$) vs wild type controls (*Trp53bp2*$^{+/+}$). P values were obtained by IMPC's Mixed Model framework, linear-mixed effects model; **, p < 0.01; ***, p < 0.001; ****, p < 0.0001.

2.6 *TP53BP2* duplications associated with developmental abnormalities have breakpoints overlapping the deletion SRO

Recently, an ExAC study estimated the sensitivity of protein-coding genes to copy number variation based on a linear-regression model of expected CNV frequencies. No high-confidence *TP53BP2* duplications were observed across the 59,898 human exomes examined, which resulted in *TP53BP2* ranking within the top 1% of protein-coding genes based on the likelihood of duplication-sensitivity (z-score = 1.66) (52). Interestingly, this was higher than *TP53BP2*'s and *FBXO28*'s estimated sensitivity to deletions (z = 0.50 and 0.47, respectively), despite the fact that *FBXO28* was called as likely haploinsufficient by ExAC based on its SNV profile. We therefore asked whether *TP53BP2* duplications could be pathogenic.

Table 2.4 CNV frequencies of ASPP genes in selected studies.

Study	No. deletions (patients)	No. deletions in controls	No. duplications (patients)	No. duplications (controls)
Coe 2014: *PPP1R13B*	9 (n=29085)	5 (n=19584)	7 (n=29085)	9 (n=19584)
Coe 2014: *TP53BP2*	6	0	10	0
Coe 2014: *PPP1R13L*	0	0	1	15
Cooper 2011: *PPP1R13B*	0 (n=15767)	4 (n=8329)	2; 2neuro 1craniofacial	9
Cooper 2011: *TP53BP2*	3; 1neuro+1cardio	0	2; 2neuro 1craniofacial	0
Cooper 2011: *PPP1R13L*	0	0	0	1
ALS Wain 2009: *PPP1R13B*	7 (n=575)	0 (n=621)	0	0

Large studies focused on CNVs in various developmental diseases found no *TP53BP2* deletions among controls but found multiple deletions in patients (**Table 2.4**). Interestingly, *TP53BP2* duplications were also found exclusively in patients. This supports the notion that *TP53BP2* is sensitive to both copy number gain and loss in human development.

A DECIPHER search identified several cases with small interstitial duplications that include some *TP53BP2* sequence, the smallest one spanning merely 256 kb (**Figure 2.13a**). This small duplication includes entire coding sequence of *TP53BP2* and *CAPN2* plus two exons of *CAPN8*. Intriguingly, the duplication breakpoints are in the close vicinity of the smallest region of overlap of 1q41q42 deletion patients with brain structural abnormalities (**Figure 2.13c**).

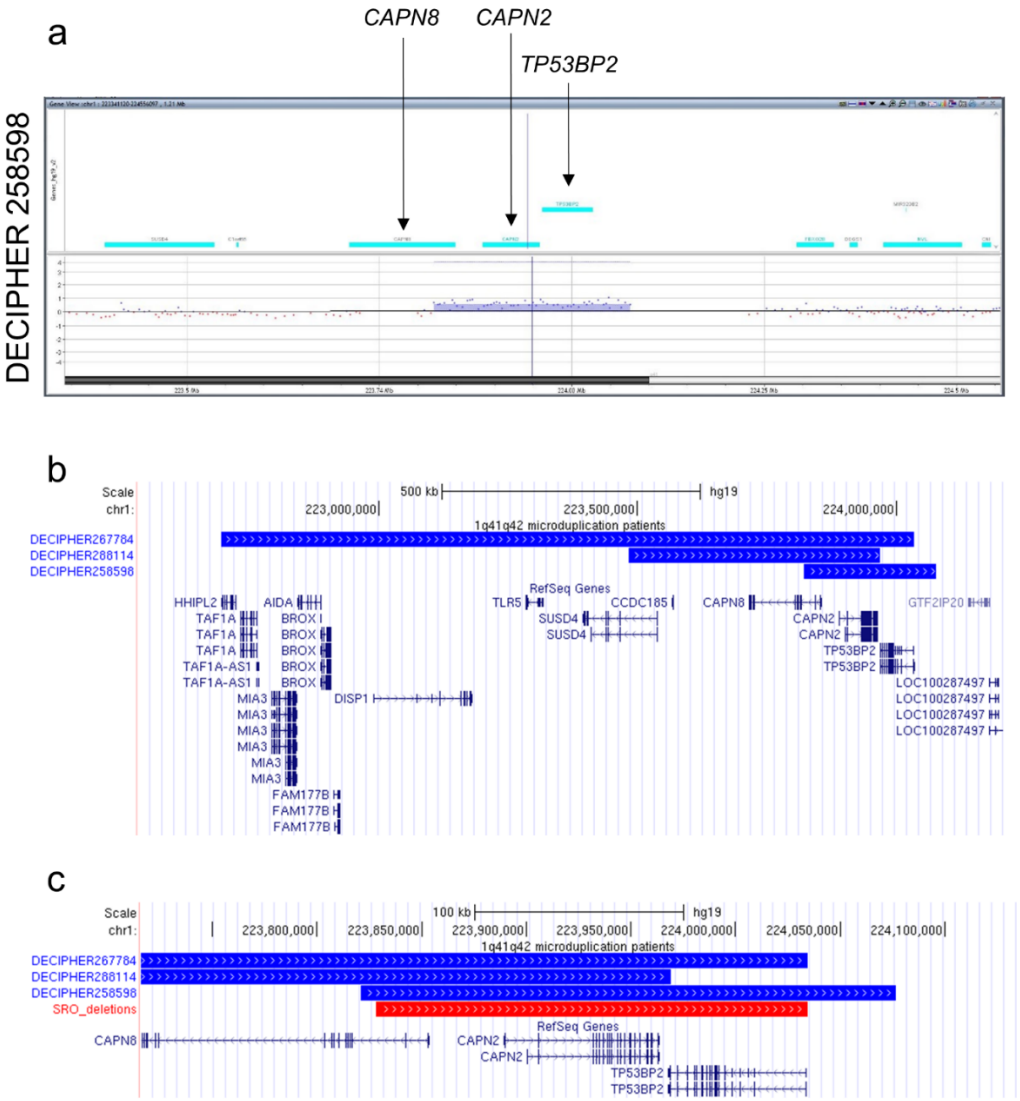


Figure 2.13 Genomic locations of 1q41q42 microduplications in new patient cases. (a) Snapshot of probe-level aCGH data depicting the 255-kb duplicated region in DECIPHER258598. The entire coding sequence of *CAPN2* and *TP53BP2* is included, along with a fragment of *CAPN8* (region highlighted by violet shading). (b) Genomic locations of three cases with <2 Mb duplications involving *TP53BP2*. The distal breakpoint of duplication detected in DECIPHER267784 ends within the coding sequence of *TP53BP2* suggesting the gene might be truncated by the duplication. (c) Comparison of the regions of overlap of duplications (blue) vs deletions (red) with brain abnormalities. Both regions of overlap spans the same 3 genes as shown in the RefSeq track. Images exported from UCSC Genome Browser, hg19 build.

We examined the patient in collaboration with Prof. Charlotte Brasch-Andersen (University of Odense, Denmark). The patient (DECIPHER 258598) is a female with a history of status epilepticus. EEG showed paroxysmal changes in the left temporo-occipital region and right occipital region. She has dysmorphic features, microcephaly (head circumference 2 s. d. below mean), vision

abnormalities and delayed psychomotor development. Initial report of the brain MRI scan at 11 months noted the high-normal depth of 0.6 cm subarachnoid space around the frontal lobes, suggesting frontal atrophy which may be related to the microcephaly. The 95th percentile upper bound in healthy infants was proposed as 0.4 cm for craniocortical width (CCW) (328). However, a later ultrasonographic study estimated a higher 95th percentile value of ca. 8 mm for CCW 11 months (329).

Reassessment of the brain scan revealed a slightly unusual configuration of the bodies of the lateral ventricles in the form of effacement of the supero-lateral aspect of the ventricular lumens, limited to the bodies of the lateral ventricles (**Figure 2.14** – coronal scan, red arrowheads). This abnormality is likely congenital. Ventricular volume analysed by ALVIN was not significantly increased compared to age-matched controls.

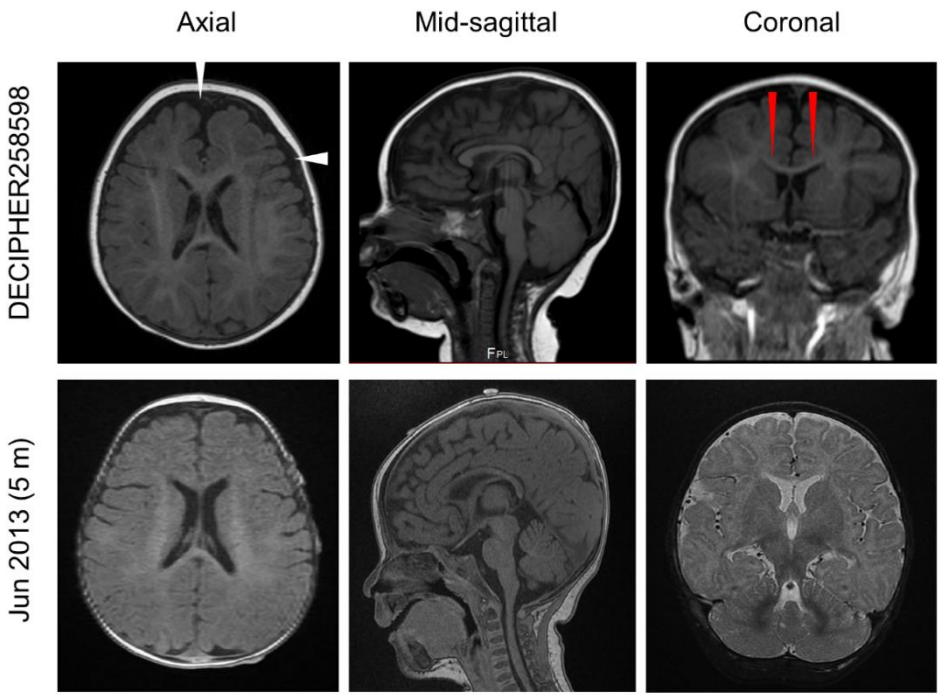


Figure 2.14 Brain MRI images of patients with triploid (top) vs diploid (bottom) *TP53BP2* in the germline. MRI of DECIPHER258598 taken at 11 months and Jun et al at 5 months of age as a normal control. White arrowheads – prominent subarachnoid space; red arrowheads – effacement of superolateral aspects of ventricular lumens.

Patient DECIPHER 267784 was examined in collaboration with Dr Kristina Aubell (Medical University of Graz). The patient is female and presented with postnatal microcephaly (at 6 years 11 months: head circumference 48.4 cm, <3rd percentile; weight 26.8 kg, 90th percentile; height 123.3 cm, 50-75th percentile). Head MRI was not performed; EEG was normal. She shows mild intellectual disability and requires special schooling.

Patient DECIPHER 288114 was unavailable for contact at the time of submission. However, a history of seizures was reported for this patient on DECIPHER.

Taken together, these data suggest that duplications in the 1q41q42 region involving *TP53BP2* (with or without truncating potential) are associated with developmental delay, neurological abnormalities and microcephaly. Given that 1q41q42 deletions associated with structural abnormalities of the brain converge on the same critical region, it is possible that *TP53BP2* and/or *CAPN2* are sensitive to both deletion and duplication. However, we cannot exclude the possibility that disruption of a regulatory element within the intergenic region that separates *TP53BP2* and *FBXO28* also plays a role. Additionally, we cannot rule out that other pathogenic variants, such as single-nucleotide variants, are responsible for the phenotype. Therefore, DNA samples collected from the patients and parents will be subjected to whole-genome sequencing in order to scan for known pathogenic SNVs including *de novo* SNVs.

2.7 Dosage sensitivity of *TP53BP2* in the ganglionic eminences and basal ganglia correlates with structural abnormalities of lateral ventricles of CNV patients

As a regulator of neural tube closure and proliferation of neuroepithelial cells, ASPP2 controls the early stages of CNS development. What role ASPP2 plays in the developed brain is unknown. We observed that *TP53BP2* is highly expressed in the cerebral nuclei in mice and humans (**Figure 2.15, Figure 2.16**).

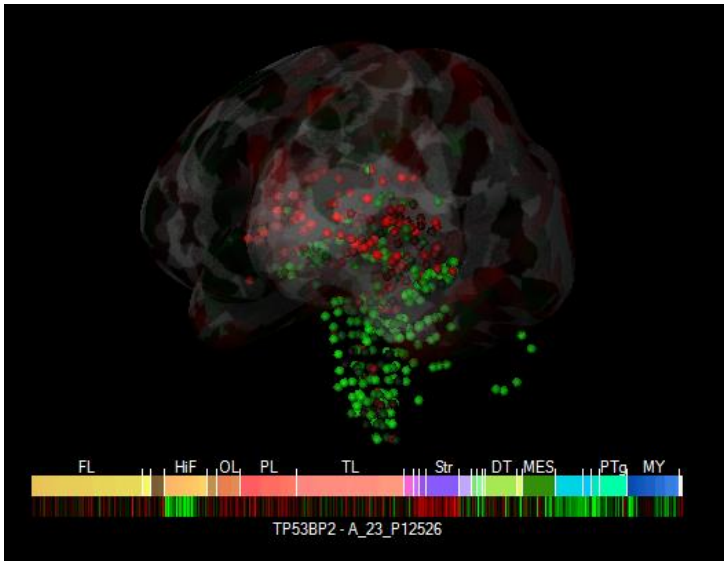


Figure 2.15 In the adult human brain, *TP53BP2* expression dominates within the cerebral nuclei. Expression data for a Representative adult brain; data retrieved from the Allen Institute's Brain Atlas.

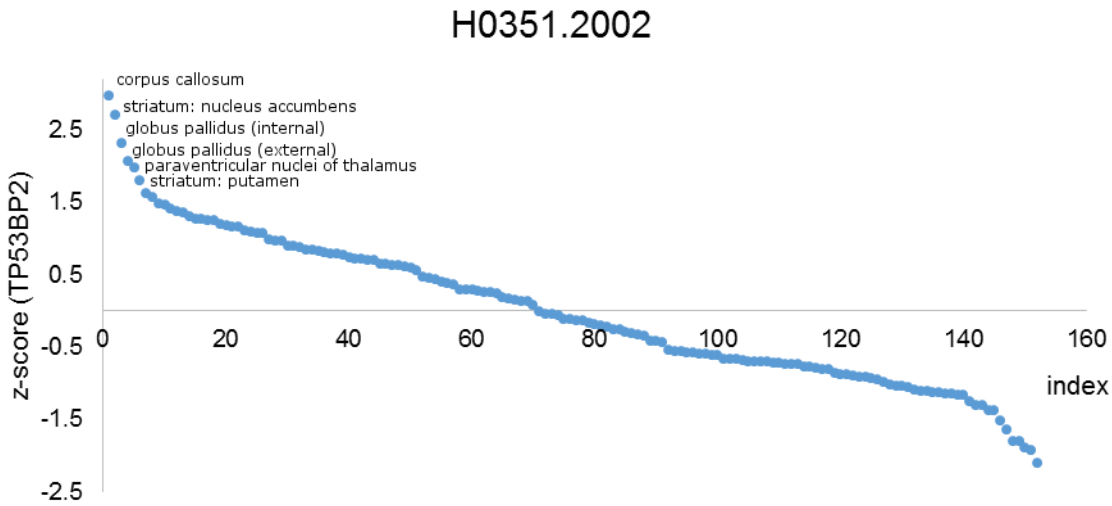


Figure 2.16 *TP53BP2* is highly expressed in the corpus callosum and basal ganglia in the developed brain. Data from donor H0351.2002; top 6 regions annotated. Data retrieved from Allen Institute's Brain Atlas.

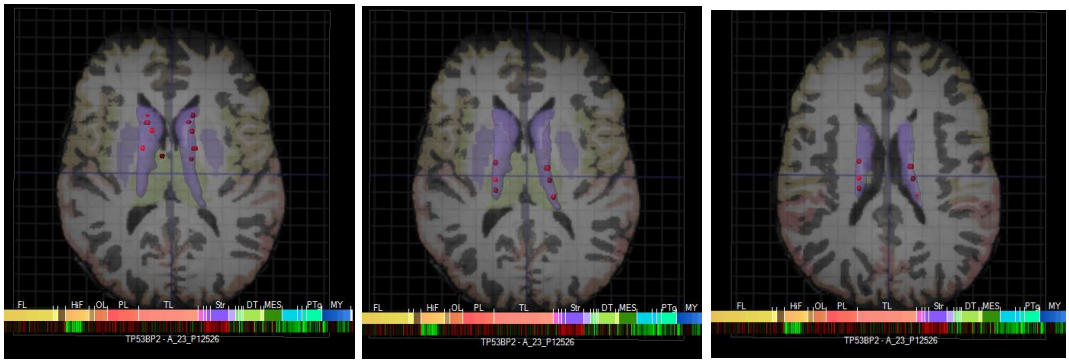


Figure 2.17 Position of the head (left) vs body (centre and right) of the caudate nucleus with respect to the lateral ventricles. Violet surface marks the entire caudate nucleus. Dots signify tissue samples analysed by microarray; colour coding indicates mRNA z-score of *TP53BP2*, ranging from green (negative) to red (positive).

Because copy number changes affect gene expression levels, dosage sensitivity is most likely to affect areas in which *TP53BP2* levels are tightly regulated. As shown in **Figure 2.20**, globus pallidus shows very consistent and high *TP53BP2* levels. Within the basal ganglia, *TP53BP2* is most consistently expressed in the head of the caudate nucleus between individuals (**Figure 2.17**). The relative level of expression with respect to other regions (z-score) is the same for male and female brains. *TP53BP2* is still very consistently expressed within the tail of the caudate nucleus, however at lower levels than the head and body of the caudate. This spatial gradient of *TP53BP2* might play a role in regulating caudate development and/or function. Therefore, among the striatal regions, the caudate is most likely to exhibit *TP53BP2* dosage sensitivity.

Inspecting the position of the caudate nucleus with respect to the lateral ventricles (**Figure 2.17**) revealed a striking correlation with the dysmorphic regions in patients with *TP53BP2* deletions and duplications. The head of the caudate nucleus lies adjacent to the frontal horns of the lateral ventricles; this region is hypoplastic in 1q41q42 microdeletion patients who show wider frontal horns of the lateral ventricles. In contrast, patient DECIPHER258598 who carries a *TP53BP2*

duplication, showed an effacement of the superolateral aspects of the lateral ventricles, suggesting the head of the caudate nucleus might be increased in size (**Figure 2.18**). The body of the caudate nucleus runs parallel to the bodies of the lateral ventricles. *TP53BP2*-deletion patients frequently showed more parallel and bulbous bodies of the lateral ventricles (**Figure 2.18, Figure 2.3**), suggesting a hypoplastic body of the caudate nucleus might have led to the expansion of LVs *in vacuo*. Taken together, these results suggest that *TP53BP2* gene dosage controls the size of the caudate nucleus in humans, and *TP53BP2* copy number variants lead indirectly to lateral ventricle abnormalities by reducing (deletions) or increasing (duplications) the size of the caudate nucleus.

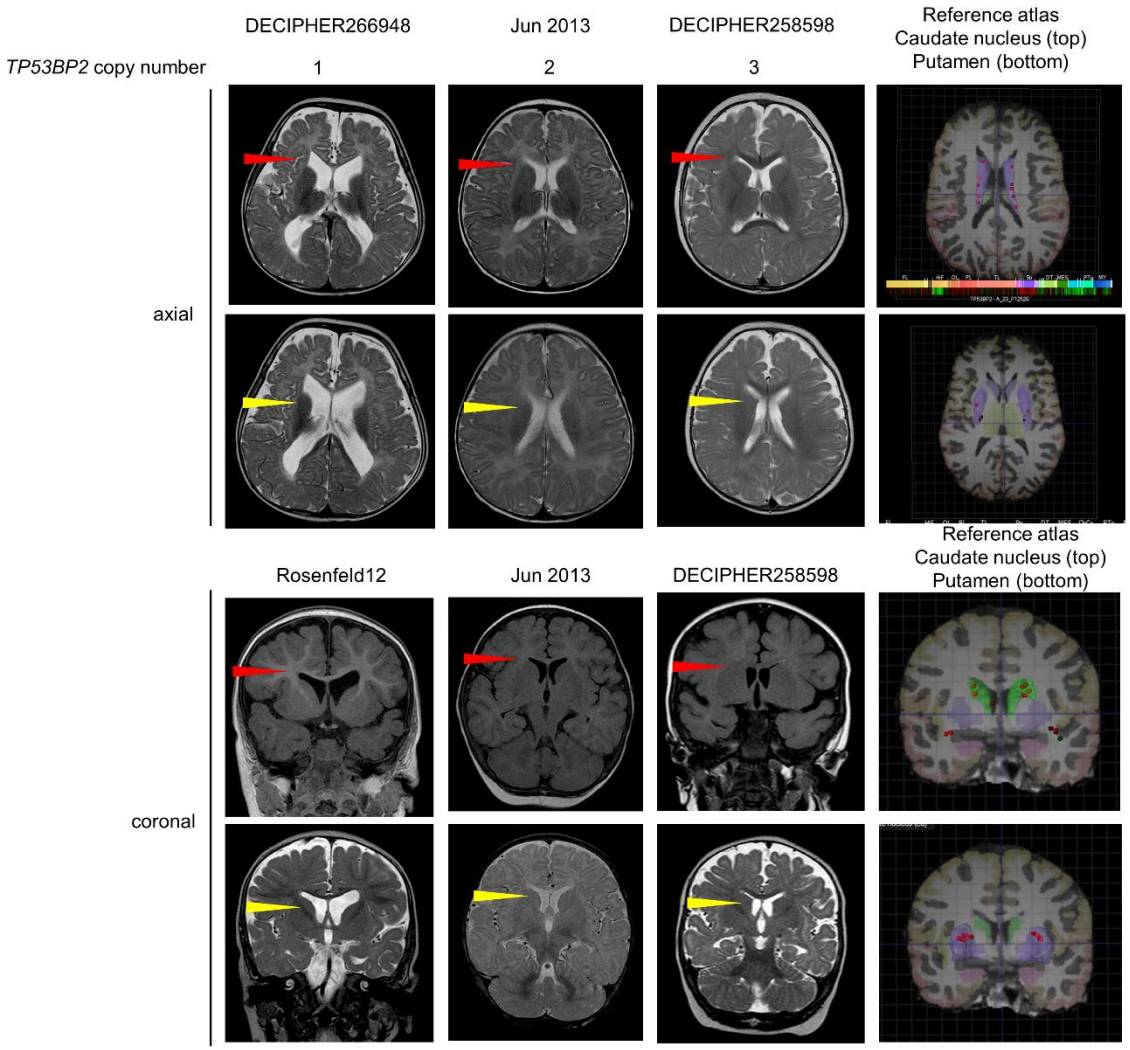
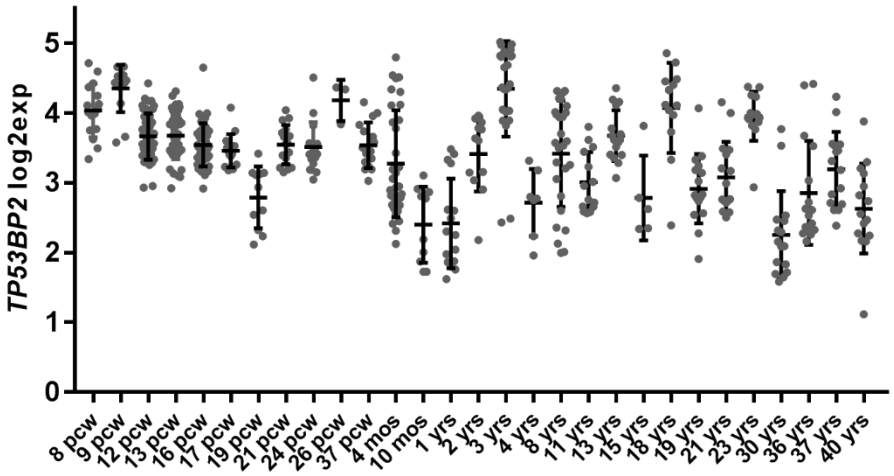


Figure 2.18 *TP53BP2* germline copy number correlates with caudate nucleus size, and lateral ventricle size and morphology in patients. Red arrowheads = anterior horns of lateral ventricles (caudate nucleus adjacent); yellow arrowheads = bodies of lateral ventricles



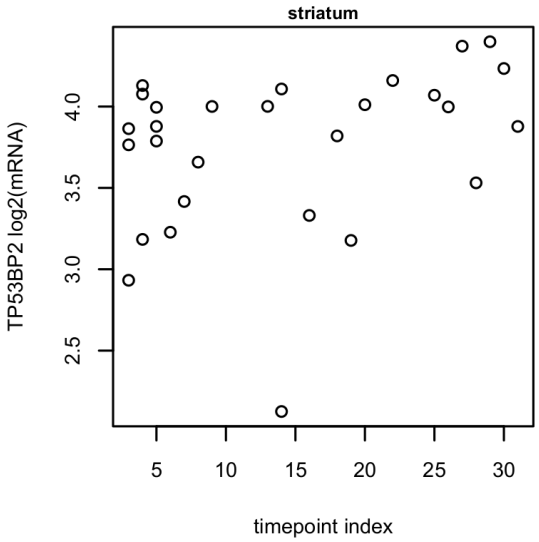


Figure 2.19 Opposing temporal gradients of *TP53BP2* levels in the developing and aging brain: global decrease vs increase of *TP53BP2* in the striatum.

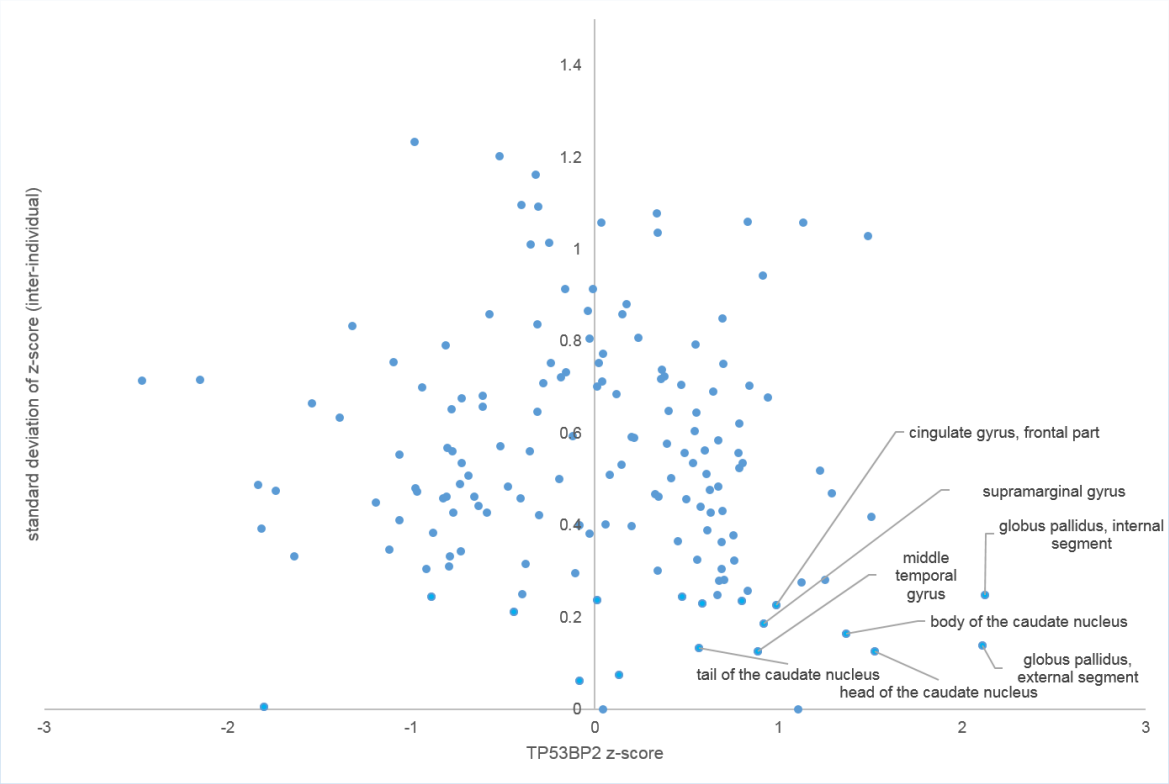


Figure 2.20 Standard deviation vs mean *TP53BP2* expression per brain region across multiple male adult brain specimens. Regions with high TP53BP2 expression and low s. d. labelled.

The correlation between consistent *TP53BP2* levels in the caudate nucleus with structural abnormalities associated with *TP53BP2* dosage in humans prompted us to investigate the spatiotemporal pattern of *TP53BP2* expression in the human brain (**Figure 2.19**). Interestingly, whereas *TP53BP2* levels show an overall decreasing trend in most brain regions, consistent with a strong foetal developmental role of ASPP2, *TP53BP2* levels increase in the striatum (which includes the caudate nucleus) during development and throughout adulthood (**Figure 2.19**-bottom panel). This opposing temporal gradient suggests ASPP2 might play a role in controlling striatal development.

Analysing the standard deviation of *TP53BP2* levels in brain regions between different donors versus mean *TP53BP2* z-scores (**Figure 2.20, Figure 2.21**) revealed that in the adult brains, globus pallidus and caudate nucleus are

among the brain regions with the highest and most consistent *TP53BP2* levels (**Figure 2.20**). In the developing brain, the medial and caudal ganglionic eminences strikingly show the highest *TP53BP2* mean level and the lowest standard deviation between donors (**Figure 2.21**). These observations were validated in an independent dataset of high-resolution microarray analysis of 4 foetal brains (**Figure 2.22**, LMD microarray). Given that ganglionic eminences give rise to striatal cell types and cortical interneurons, these data suggest ASPP2 promotes the differentiation of glial cells from the ganglionic eminence into the striatum. This model is consistent with the observation of hypoplastic caudates among *TP53BP2*-deletion patients and a hyperplastic caudate nucleus in a *TP53BP2*-triploid patient. In addition, a premature differentiation of ganglionic eminence progenitors can lead to decreased populations of other ganglionic eminence-derived cell types including cortical interneurons, the loss of which may underlie the observed microcephaly and prominent subarachnoid space in the *TP53BP2*-triploid patient (**Figure 2.18**). Consistently, defects in GABAergic interneurons have been linked with epilepsy.

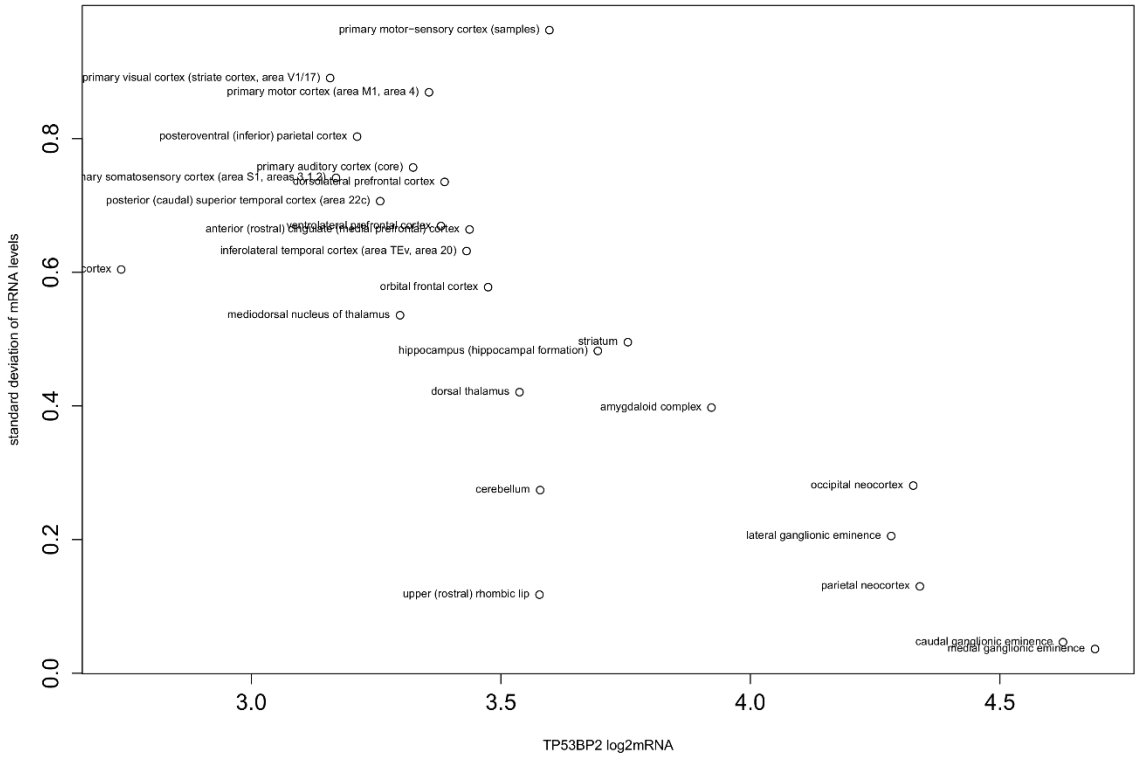


Figure 2.21 *TP53BP2* is highly and consistently expressed in the medial and caudal ganglionic eminence in the developing human brains. Span of brain ages: 8 pcw – 40 years. NB: low sample numbers will influence this but the absolute expression levels not influenced, neither is the overall shape of regression line. Prominent outliers: upper (rostral) rhombic lip; cerebellum.

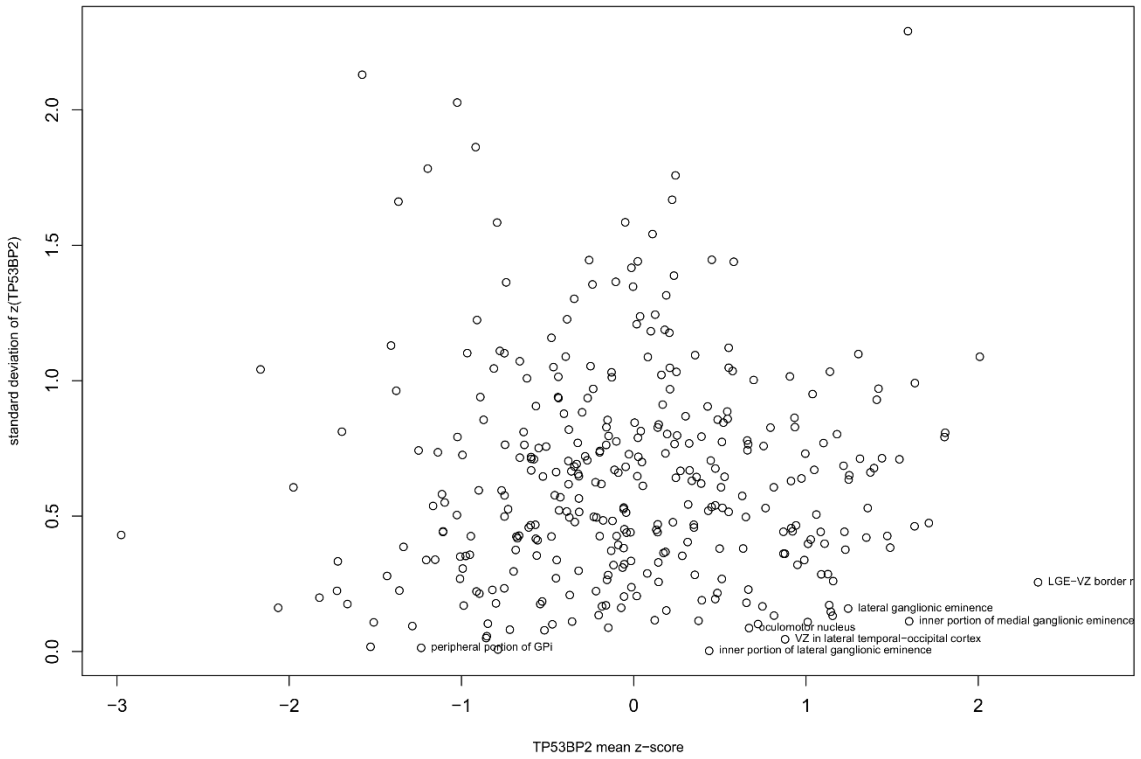


Figure 2.22 *TP53BP2* levels are highly regulated in the ganglionic eminences during prenatal development. Source: Human prenatal LMD microarray; 3 female brains aged 15, 21 and 21 pcw; 1 male brain aged 15 pcw. LGE-VZ, lateral ganglionic eminence – ventricular zone.

2.8 Sexually dimorphic expression of *TP53BP2* in the basal nuclei may underlie male-specific phenotypes

TP53BP2 is highly expressed in the caudate nucleus (part of striatum) in both genders, but dominates in the amygdala in females and the globus pallidus in males. This separation was very consistent across different male brains (**Figures 2.23, 2.24**). This dimorphism suggests *TP53BP2* might play a male-specific role in the globus pallidus and a female-specific role in the amygdala. Strikingly, *Trp53bp2*$^{\Delta 4/+}$ male mice discussed above (official allele name *Trp53bp2*<tm1b(EUCOMM)Hmgu>) showed a significant male-specific increase in voluntary movement in the open-field test (**Figure 2.27**). They travelled at greater speeds and covered more distance than wild-type males, but no difference was observed between *Trp53bp2*$^{\Delta 4/+}$ vs wild-type females. *TP53BP2* is also highly expressed in the nucleus accumbens which plays a role in encoding new motor programs. These observations suggest ASPP2 controls movement activation specifically in males and that this function might involve ASPP2 expression in the basal ganglia. It would be interesting to test whether caudate nucleus and/or globus pallidus size and function is affected in *Trp53bp2*$^{\Delta 4/+}$ males.

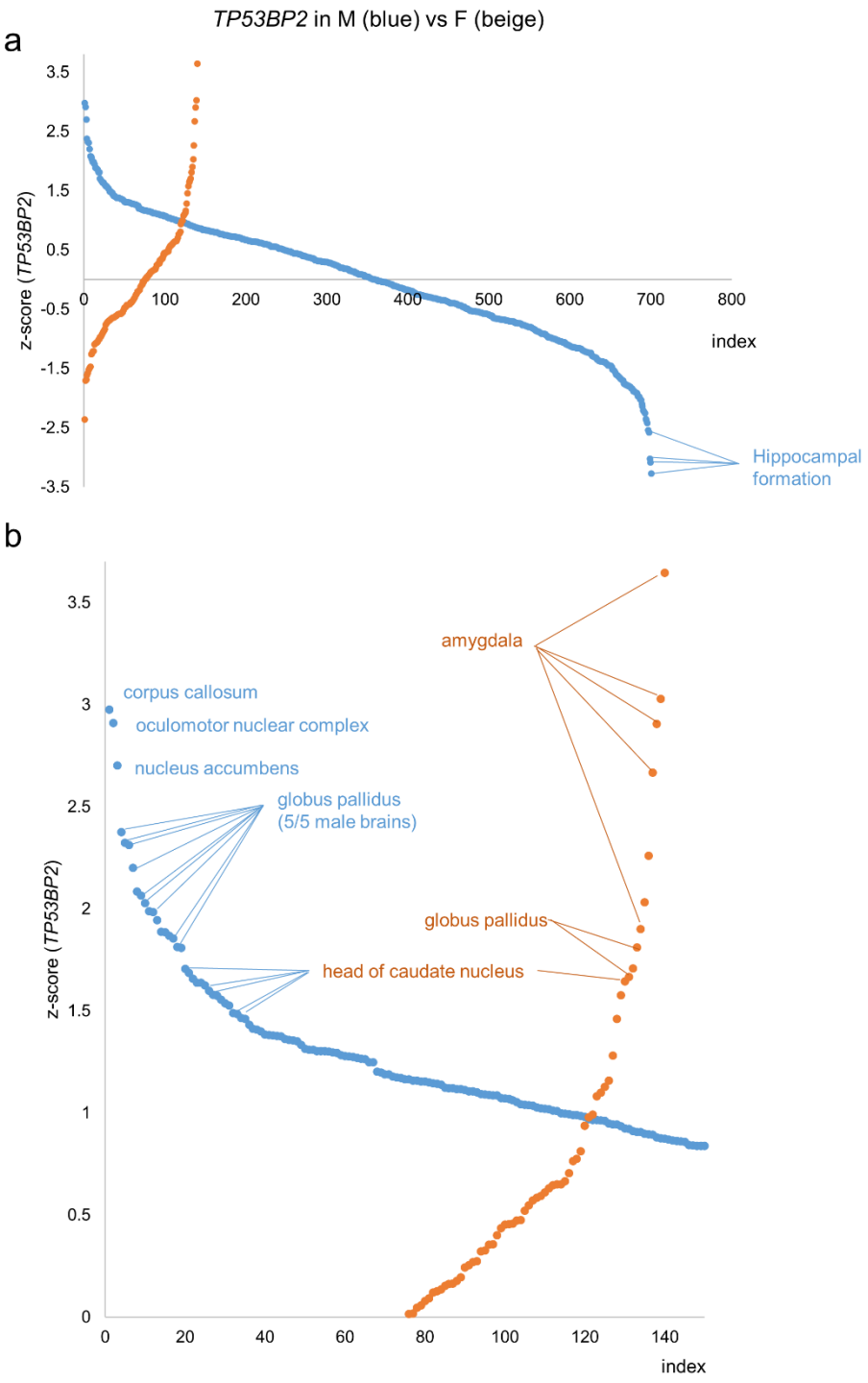


Figure 2.23 ***TP53BP2* expression is spatially dimorphic in human adult brains**. The top region in males (M, blue) is globus pallidus, whereas in females (F, beige) *TP53BP2* is higher in the amygdala. Some nuclei such as caudate nucleus show very similar *TP53BP2* levels in both genders.

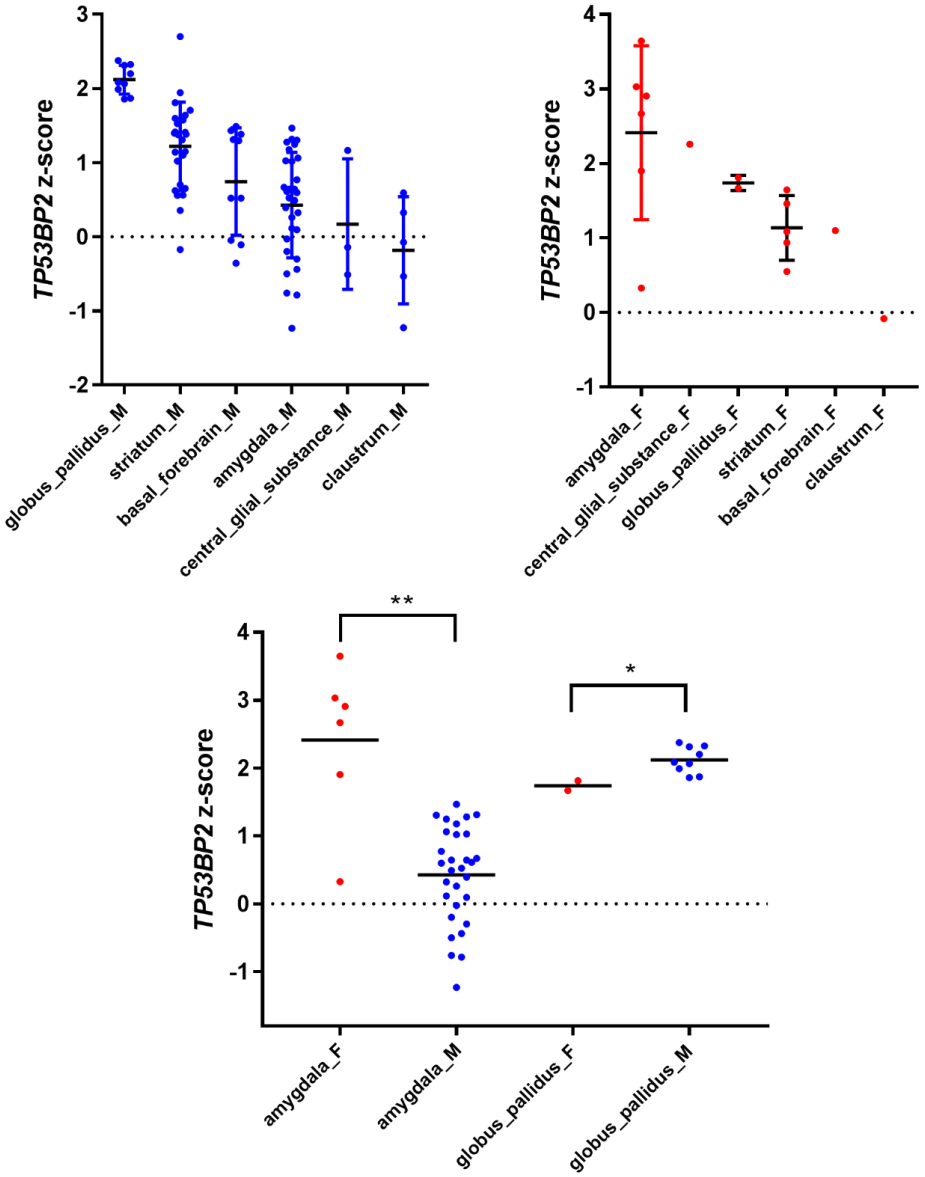


Figure 2.24 Sexual dimorphism of *TP53BP2* expression in the adult human cerebral nuclei. Blue dots represent data points for male brain regions; red dots for female brain regions. Source data retrieved from the Allen Institute; M, males; F, females; **, $p < 0.01$; *, $p < 0.05$; Student's two-tailed t-test.

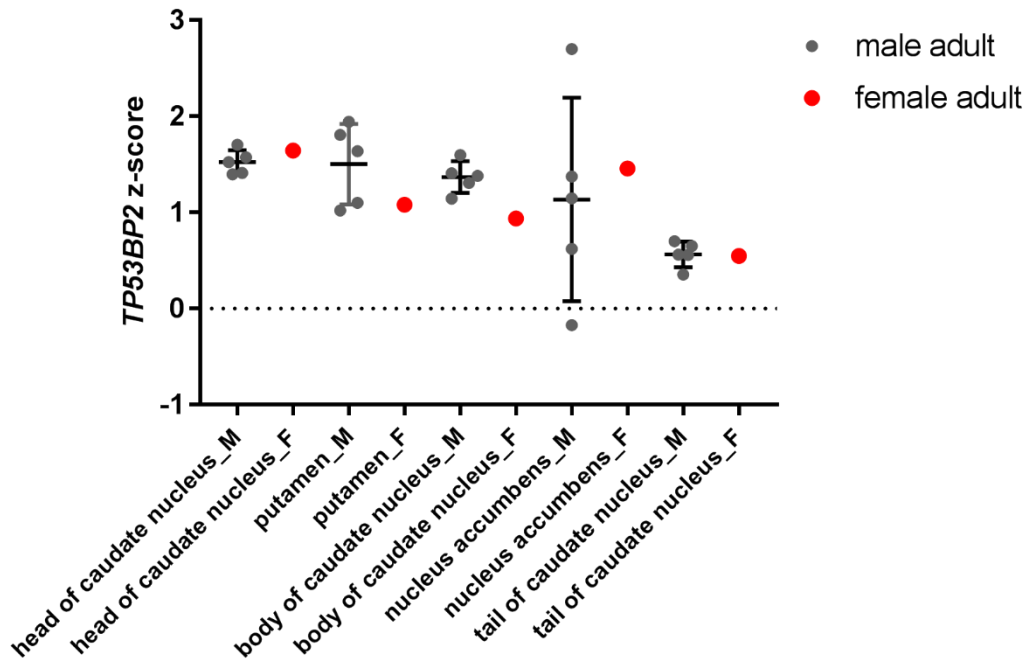


Figure 2.25 *TP53BP2* is very consistently expressed within the caudate nucleus compared with other striatal regions. Data from Allen Institute's Human developed brain transcriptome.

Trp53bp2$^{\Delta 3/\Delta 3}$ males die earlier and show more severe hydrocephalus phenotype than *Trp53bp2*$^{\Delta 3/\Delta 3}$ females. In fact, virtually all the long-surviving (<3 months) adult *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice are female. This suggests *Trp53bp2* is developmentally less redundant in males and explains why heterozygous exon 4-deletion is also more penetrant in males. In addition, all *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice that survive birth show hopping gait, regardless of gender. Given the observed expression pattern of ASPP2 in the basal ganglia, it is likely that ASPP2 deficiency in the basal ganglia causes the hopping gait phenotype in mice.

The phenotypic and expression dimorphism has another interesting connotation. In breast cancer cells, *TP53BP2* levels are repressed by the addition of oestradiol (Broad compound database) and *TP53BP2* is strongly associated with the basal subtype of breast cancer in which it is significantly upregulated

compared with normal tissue and compared with ER^+ breast cancers. Moreover, *TP53BP2* is induced during decidualisation in the mammalian human endometrium which occurs in a low-oestrogen phase of the menstrual cycle. *Trp53bp2* deficiency also causes gonadal abnormalities (data above) and ASPP2 protein is expressed on sperm and regulates sperm development (Zak et al., manuscript in preparation). All these suggest that *TP53BP2* is physiologically negatively regulated by oestrogen in selected tissues, which could play a role in the observed dimorphism in the brain.

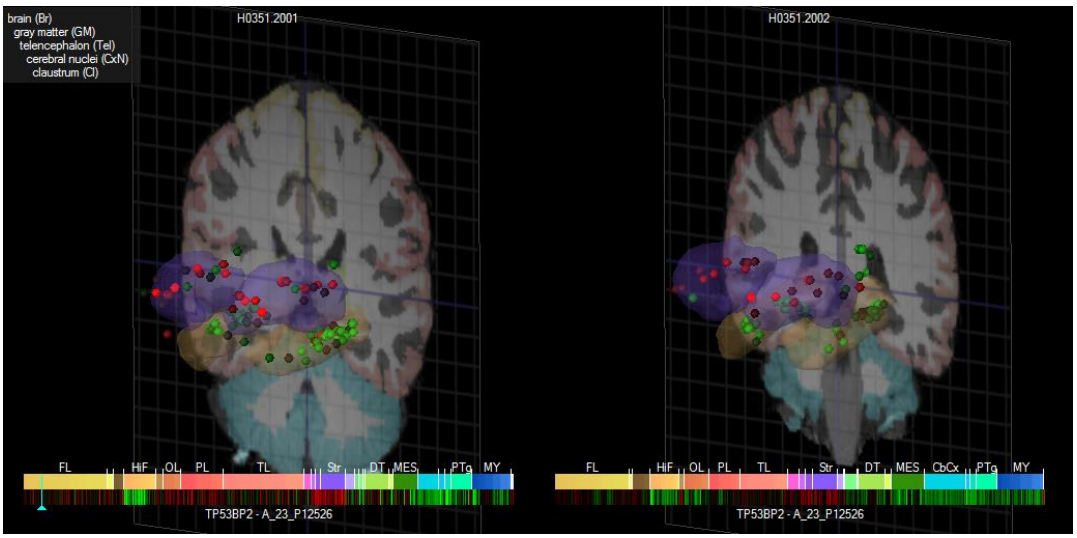


Figure 2.26 *TP53BP2* shows polarity in expression in the putamen and claustrum (violet surface) vs hippocampus (brown surface). Red = z(*TP53BP2*) > 0; green = z(*TP53BP2*) < 0.

Interestingly, the lowest *TP53BP2* levels are found in the hippocampal formation which is immediately adjacent to the putamen and claustrum (**Figure 2.26**). This sets up a sharp spatial polarity of *TP53BP2* expression delineated by the contact surface between the putamen, amygdala and the hippocampal formation. In females the contact surface between the amygdala and hippocampus separates the region of highest and lowest *TP53BP2* expression in the brain. Given ASPP2's role in controlling promoter selectivity of various

transcription factors, it is possible this spatial pattern of expression reflects ASPP2's role in specification of the cerebral nuclei during development.

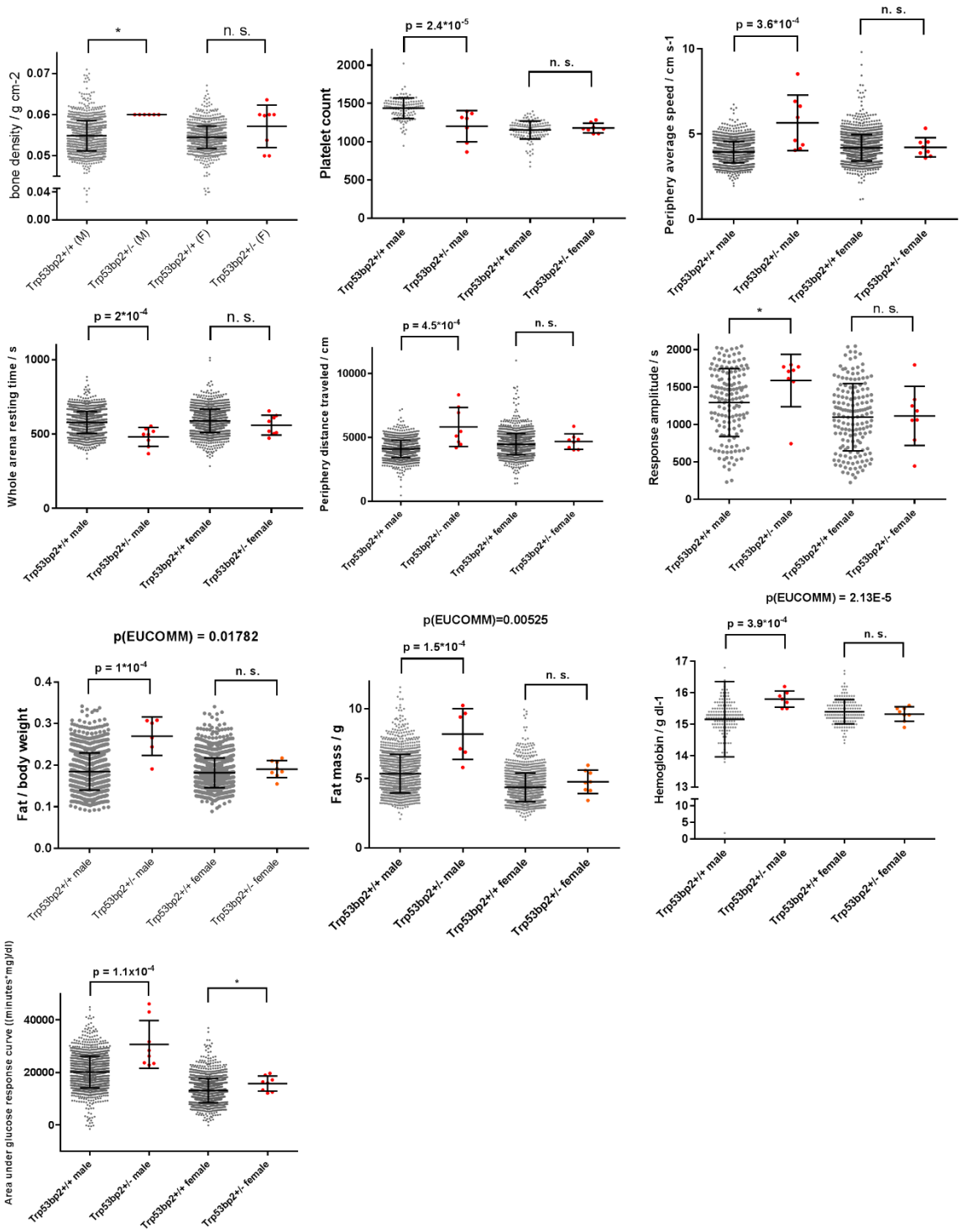


Figure 2.27 *Trp53bp2*-Δ4/+ heterozygous mice have multiple male-specific phenotypes. Statistical tests used: panel Response amplitude – Mann-Whitney-U test; whole-arena resting time: Mann-Whitney U; Fat/body weight – Mann-Whitney-U; area under glucose – Mann-Whitney U; fat mass – Mann-Whitney U.

Potassium – Mann-Whitney U. *, p<0.05

To determine the extent of sexual dimorphism in ASPP2 function in mice, we mined all datasets from IMPC's Δexon4 line that showed a dimorphic phenotype. Numerous parameters showed a statistically significant difference in *Trp53bp2*$^{\Delta 4/+}$ males but not females (**Figure 2.27**). In comparison, only a handful of parameters had a female-specific phenotype; one example being blood potassium levels (**Figure 2.28**). Non-brain phenotypes of Aspp2-deficient mice are also of interest. For example, *Trp53bp2*$^{\Delta 4/+}$ male have a significant increase in haemoglobin levels (**Figure 2.27**). A SNP within the *TP53BP2* gene has been linked with levels of glycosylated haemoglobin (p= 0.000683) (330).

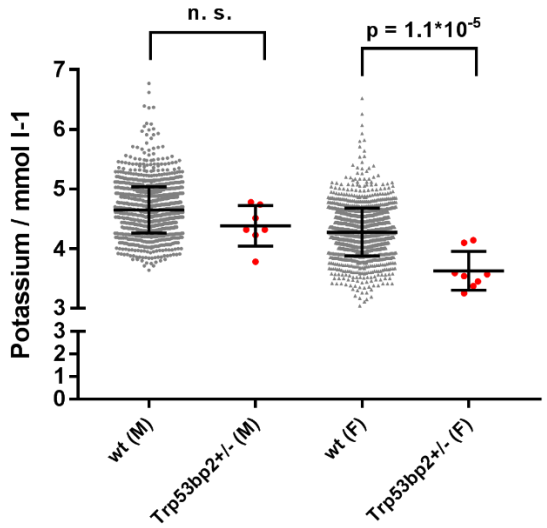


Figure 2.28 Circulating blood potassium levels are decreased in *Trp53bp2*$^{\Delta 4/+}$ females but not in males. Raw data extracted from Phenoview, data release v4.3; Mann-Whitney U test.

Taken together, these results demonstrate that caudate nucleus and lateral ventricles are sensitive to *TP53BP2* dosage in humans and mice, and that

Trp53bp2 controls locomotor activation in male mice. The observation of specifically high expression of *TP53BP2* in the basal ganglia of adult brains suggests the gene might play a general role during CNS development and then assume a more specialized role in the basal ganglia of developed brains. It remains to be determined whether structural changes in the basal ganglia and lateral ventricles cause the observed neurological symptoms in patients with *TP53BP2* CNVs, such as seizures or intellectual disability. Due to ASPP2's multiple roles in CNS development it is difficult to decouple effects caused by early events, such as neural tube closure or proliferation of early neuroepithelial cells, from later events such as development of the caudate nucleus.

The observation of higher *TP53BP2* levels in the globus pallidus of males vs females, the impact of *TP53BP2* copy number on basal ganglia size, the male-specific increase in locomotor activation upon exon4 deletion in mice, together with the 100% penetrant hopping gait in *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice demonstrate that ASPP2 is required for the development of full motor function in mice and that the basal ganglia, especially the caudate nucleus and globus pallidus, are likely to play a role in this process. Indeed, hyperkinetic disorders such as Parkinson's disease have been connected with the impairment of striatal neurons projected to the lateral globus pallidus (331). In monkey models of Parkinson's disease, firing rates of globus pallidus external (GPe) neurons are reduced (332, 333).

These results also have potential relevance in schizophrenia. Studies have shown that schizophrenia patients smaller absolute and relative volumes of white matter in the caudate nucleus relative to healthy subjects (334). *TP53BP2* might play a role in the pathology of schizophrenia as a neighbouring SNP has been

linked with the disease in a GWAS study (rs7539624; $p = 8 \times 10^{-6}$) (335). Interestingly, another GWAS link came in a study of MMR vaccine-related febrile seizures: the SNP rs140084787 with an odds ratio of 3.25 ($p = 6 \times 10^{-6}$) (336). Both of these SNPs lie within the introns of *CAPN2*, suggesting *CAPN2* might also be involved. This gene is also expressed in the caudate nucleus although unlike ASPP2, the CAPN2-deficient mice have no neurological phenotypes.

Multiple questions remain. Firstly, why are both *TP53BP2* deletions and duplications associated with microcephaly? Secondly, what are the dominant deficiencies in the brain in these CNV patients that cause the clinically relevant symptoms, and which (if any) changes are benign? Thirdly, does ASPP2 actively regulate motor functions in the basal ganglia? More mechanistic studies are necessary to answer those questions.

2.9 *PPP1R13B* shows a distinct expression profile from *TP53BP2* during development and in adults

PPP1R13B, the most homologous gene to *TP53BP2* in the human genome, encodes ASPP1. Like ASPP2, ASPP1 contains a C-terminal SH3 domain, 4 Ankyrin repeats, a proline-rich region, a coiled-coil domain and an N-terminal ubiquitin-like/RAS-association domain. In humans, the gene is most highly expressed in the thyroid, followed by testis and cerebellum (**Figure 2.29**)

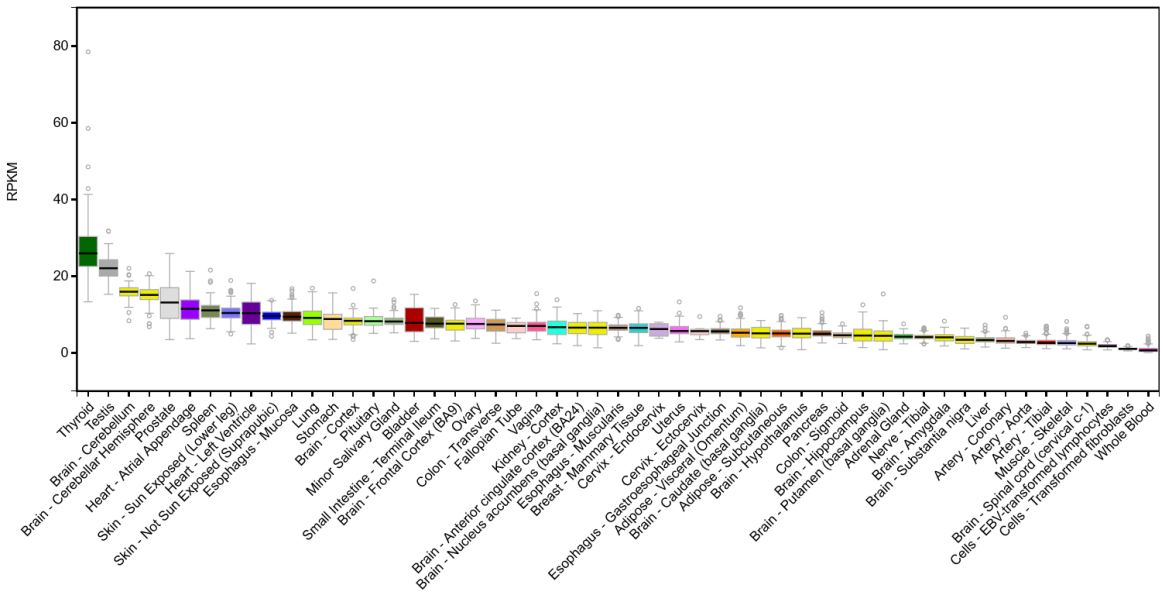


Figure 2.29 mRNA levels of *PPP1R13B* across human tissues. Data retrieved from GTEX Portal v6p.

ASPP1-deficient mice are born at less than the expected rate and show abnormalities in lymphatic vasculature (261). Therefore, damaging variants in *PPPR13B* might also play a role in human developmental diseases. At present, the relationship between genetic variation at the *PPP1R13B* locus and human disease is not known. Interestingly, according to ExAC, *PPP1R13B* is very likely to be haploinsufficient in humans (pLI = 0.98) due to its lower-than-expected number of truncating germline variants in the general population. However, **Table 2.4** shows that *PPP1R13B* hemizygous deletions segregate in the general population and do not seem to be enriched for NDDs, although they were significantly enriched in ALS patients (337). Therefore, more studies are needed to investigate the identity and frequency of variants resulting in loss of one *PPP1R13B* allele.

Interestingly, *PPP1R13B* shows an inverse expression with *TP53BP2* in the human brain and this dichotomy is most striking in the ventricular and subventricular zones (**Figures 2.30, 2.31, 2.32, 2.33**).

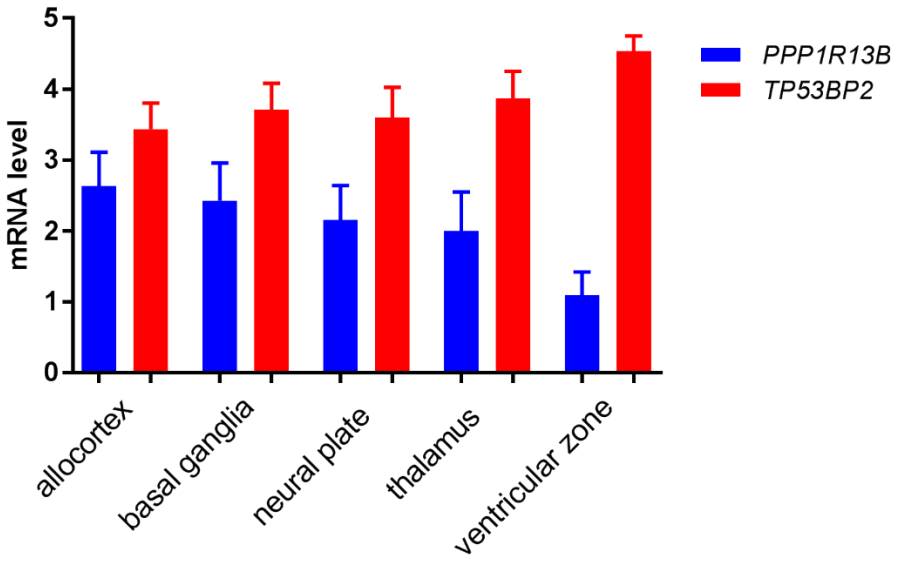


Figure 2.30 In the developing human brain, *PPP1R13B* and *TP53BP2* show significant spatiotemporal anti-co-expression. *PPP1R13B* is lowly expressed in the proliferating ventricular zone in which *TP53BP2* is most highly expressed. mRNA levels retrieved from from Allen Institute's BrainSpan atlas, Developmental Transcriptome. Data pooled across top-level structures from 42 brain specimens; regions are sorted by mean *PPP1R13B* expression.

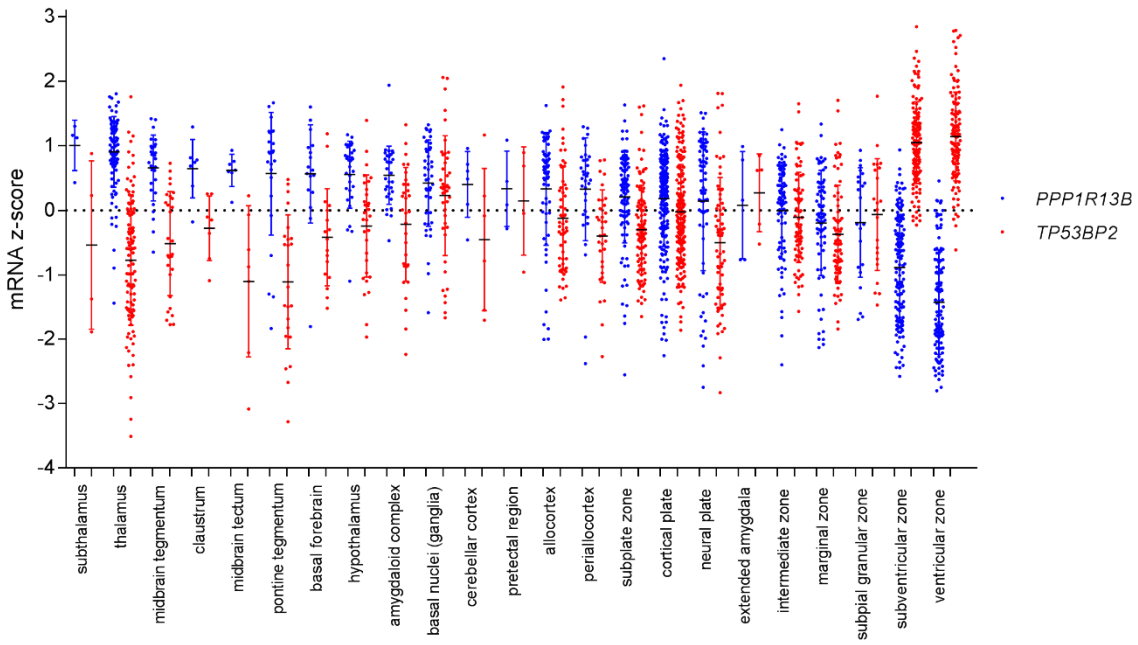


Figure 2.31 mRNA levels of *PPP1R13B* and *TP53BP2* across subregions of the human developing brain. Data taken from the BrainSpan LMD Microarray, a high-resolution of four specimens with ca. 300 distinct anatomical sites. Data from all specimens pooled by regions and subregions sorted by mean *PPP1R13B* levels.

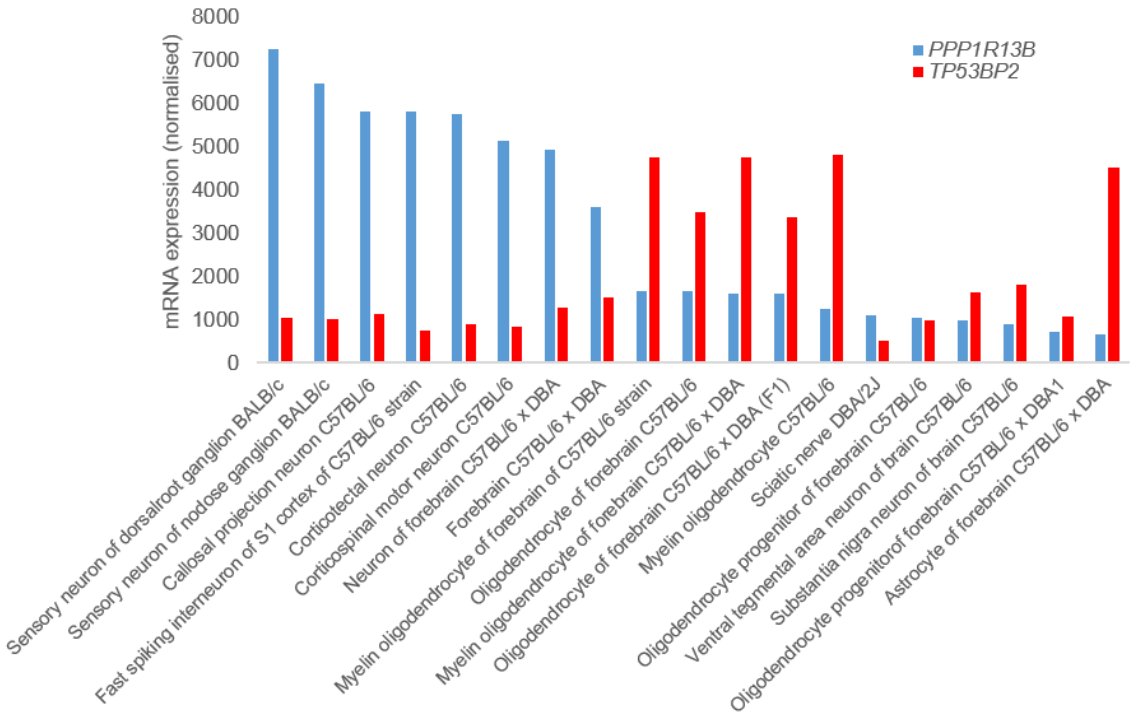


Figure 2.32 *PPP1R13B* vs *TP53BP2* expression across selected neural cell types in the mouse. Data retrieved from Illumina's Correlation Engine and replotted. Cell types sorted by mean *PPP1R13B* level.

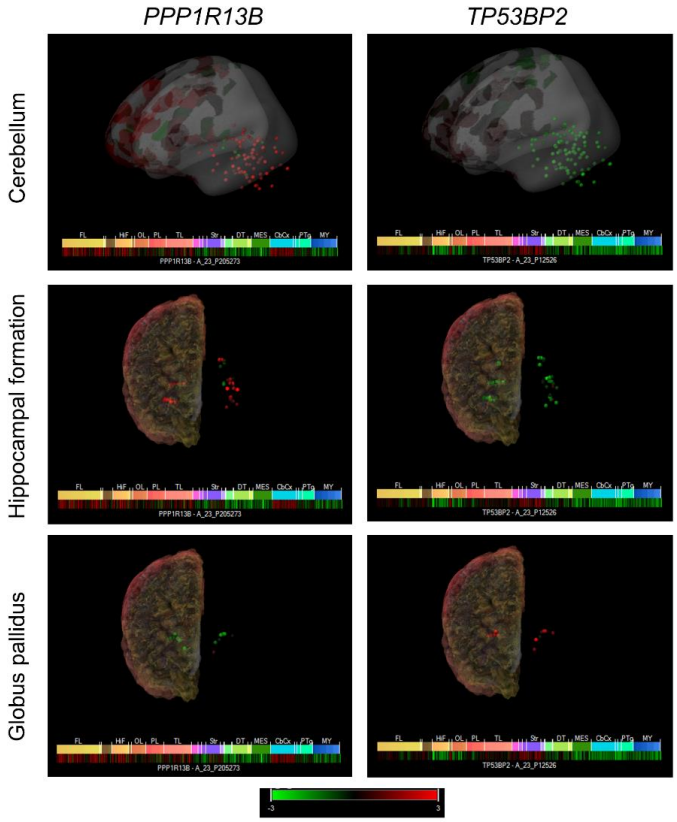


Figure 2.33 *PPP1R13B* and *TP53BP2* are spatially anti-coexpressed in the adult human brain. Representative images of brain regions with highly (red) vs lowly (green) expressed mRNA of *PPP1R13B* and *TP53BP2*. Images were taken from a single specimen, dots represent sub-regions of global region shown on the left. Hemisphere (top) and right hemisphere (middle and bottom) pial surfaces shown for orientation. Image generated using Brain Explorer 2.

This striking dichotomy in spatiotemporal expression between *PPP1R13B* is driven at least partly by cell type specificity. *TP53BP2* is expressed dominantly in neural progenitors in the developing brain, and astrocytes and oligodendrocytes in the developed brain (**Figure 2.32**). In contrast, *PPP1R13B* is most highly expressed in neurons, especially sensory and cerebellar neurons.

A genome-wide spatio-temporal co-expression analysis revealed that *PPP1R13B* is co-expressed with genes significantly enriched in synaptic function (**Table 2.5, Table 2.6**). Regulators of multiple neurotransmitter receptors were represented (e.g. NMDA, GABA receptors) and the correlation was very robust (R>0.8), increasing in statistical significance when more genes were added to the initial list. The enrichment for synaptic genes and genes regulating neurogenesis was observed using both data from the human Prenatal LMD Microarray and the human Developmental Transcriptome. In contrast, *TP53BP2* showed no co-expression with synaptic genes.

Table 2.5 *PPP1R13B* is spatiotemporally co-expressed with synaptic genes in the human brain. Microarrays from the Developmental transcriptome were used as input.

Gene	R	Cell function	Mutations - disease	Link with autism/schizophrenia
STX1B	0.834	Presynaptic protein	fever-associated epilepsy	
PIP5K1C	0.833	Synaptic protein, pain signalling	lethal congenital contractural syndrome=total immobility	3 intronic SNVs in autism
UNC13A	0.829	Synaptic vesicle exocytosis, DAG pathway	ALS	
CAMK2B	0.825	Plasticity of parallel fibre-Purkinje cell synapse		3 LoF, 1 missense in autism

KCNC3	0.824	Motor function in Purkinje cells	Degenerative and developmental CNS phenotypes e.g. ataxia
EPB41L1	0.823		
GRIN1	0.818	NMDA receptor	Schizophrenia, bipolar disorder
SPTBN4	0.817	Binds CAMKII, spectrin	
KIAA0284	0.817		
ABCG4	0.81	Transporter	

Table 2.6 Genes spatiotemporally co-expressed with PPP1R13B in human brains are significantly enriched in synaptic function. The enrichment analysis was run on database version 2016-06-22 using AmiGO2. Top input genes = genes with the highest co-expression coefficient with *PPP1R13B*.

No. of top input genes	GO term (db2016-06-22)	Rank	P value	Fold enrich ment	Genes
50	regulation of neuronal synaptic plasticity	1 (BP)	3.24E-07	58.8	*JPH3, NEURL, SLC8A2, CAMK2B, SYNGR1, UNC13A, GRIN1*
50	neuron part	1 (CC)	3.97E-14	7.9	25/50
50	synapse part	2 (CC)	6.42E-11	11.4	*GRIN1, SRCIN1, ITGA3, NEURL, SLC8A2, SLC6A17, CAMK2B, IQSEC3, STX1B, SYN1, GABBR2, SYNGR1, PIP5K1C, DNM1, UNC13A, SCAMP5, GRIN1*
100	modulation of synaptic transmission	1 (BP)	1.65E-12	13.8	*JPH3, NCDN, PRKACA, RIMS1, CLSTN1, PRKACA, CELF4, CLSTN3, NEURL, SLC8A2, CAMK2B, STX1B, SYN1, PPFIA3, SYNGR1, ATP2B2, SHANK1, UNC13A, GRIN1*
100	regulation of neuronal synaptic plasticity	2 (BP)	3.15E-08	37.8	49/100
100	neuron part	1(CC)	4.48E-21	6.7	42/100
100	synapse	2 (CC)	2.38E-19	8.9	33/100

2.10 Missense mutations in *PPP1R13B* in human developmental diseases

De novo mutations have been implicated in NDDs including autism, schizophrenia and intellectual disability. To determine whether *PPP1R13B*

variants might play a role in these diseases, we mined published studies on *de novo* mutations in NDDs and the ClinVar database for any *PPP1R13B* mutations. Interestingly, multiple *de novo* missense mutations in *PPP1R13B* have been identified in NDDs (**Table 2.7**). Two *de novo* mutations were found in autism spectrum disorder patients: one novel intronic and one rare missense variant (G881R). One *de novo* missense mutation was found in childhood-onset schizophrenia (R72Q). No *de novo* variants in *TP53BP2* or *PPP1R13L* were found in these datasets.

Table 2.7 *De novo PPP1R13B* mutations discovered in NDD patients

Study	Disease	patient	SNV (nucleotide)	SNP ID	SNV (protein)	Inherited/de novo	ExAC AC
DeRubeis 2014 (48)	ASD	NDAR_INVPA860MYW_wes1	14:104205312C>T	rs373141354	G881R	De novo	4/120504 (all European)
Iossifov 2014 [Simons Simplex Collection] (143)	ASD	12603.p1	14:104216291AC>A	-	intronic	De novo	-
Ambalavanan 2016 (338)	Childhood-onset SCZ	COS2720	Chr14: 104251194G>A	-	R72Q	De novo	0/120740

A missense variant such as R72Q may have a dominant effect (poison allele) rather than just gene dosage change effects such as a deletion. *PPP1R13B* could be relatively tolerant to gene dosage but not to missense mutations in certain domains. This is consistent with the observations that *TP53BP2* controls proliferation of neural progenitors whereas *PPP1R13B* is spatio-temporally co-expressed with synaptic genes and preferentially expressed in neurons. Therefore, *TP53BP2* might be more important in controlling the early events in

CNS development and controlling the size of the NPC pool whereas *PPP1R13B* might play a role in synaptic function. Two of the *de novo* mutations are classified as Likely pathogenic based on ACMG's guidelines, with the third being of uncertain significance. Taken together, *PPP1R13B* is highly expressed in neurons, co-expressed with synaptic proteins, and disrupted by likely pathogenic *de novo* mutations in ASD and childhood-onset schizophrenia, which might disrupt ASPP1's function at the neuronal synapse.

2.11 Deleterious ASPP family variants and evolutionary constraint

Given the discrepancy between *PPP1R13B*'s pLI score and the observed frequency of *PPP1R13B* deletions in the general population, and on the other hand a low pLI score (<0.1) for *TP53BP2* whose deletions are associated with severe disease, we mined ExAC's data on the observed vs expected occurrence of LoF variants (**Table 2.8**). **Figure 2.34** depicts the pLI vs z(LoF) spectrum for all genes, z(LoF)being the z-score of observed vs expected number of LoF variants among all genes, and **Figure 2.35** shows all false negatives with curated confirmed haploinsufficiency status in ClinGen. These haploinsufficient genes include those whose deletion causes relatively mild phenotypes that impart less stringent selective pressure (e.g. *LDLR* associated with hypercholesterolemia). However, some genes in this set cause severe developmental phenotypes, such as *MEF2C.* Notably, z(LoF) of *TP53BP2* is the highest among genes with a pLI<0.2, i.e. *TP53BP2* has a higher z(LoF) than all of the false-negatives in this analysis. This makes *TP53BP2* a likely false-negative haploinsufficient gene.

Table 2.8 Germline variants in ASPP genes detected by ExAC

ExAC_subset	Gene	Variant type	Expected no.	Observed no.	z-score	Constraint metric
All_r03	*PPP1R13B*	Synonymous	212.7	191	0.92	pNull=1.57e-9
All_r03	*PPP1R13B*	Missense	433.7	345	2.08	Prec=0.023
All_r03	*PPP1R13B*	LoF	37.7	6	5.12	pLI=0.977
All_r03	*TP53BP2*	Synonymous	165.0	163	0.10	pNull=7.5e-6
All_r03	*TP53BP2*	Missense	379.6	336	1.09	Prec=0.998
All_r03	*TP53BP2*	LoF	37.6	12	4.14	pLI=0.002
All_r03	*PPP1R13L*	Synonymous	212.1	154	2.47	pNull=0.0030
All_r03	*PPP1R13L*	Missense	363.8	271	2.38	pRec=0.994
All_r03	*PPP1R13L*	LoF	20.7	8	2.77	pLI=0.0027
Non-TCGA	*PPP1R13B*	Synonymous	195.6	179	0.76	pNull=2.05e-9
Non-TCGA	*PPP1R13B*	Missense	399.0	316	2.07	Prec=0.0135
Non-TCGA	*PPP1R13B*	LoF	34.7	5	4.97	pLI = 0.987
Non-TCGA	*TP53BP2*	Synonymous	151.8	154	-0.11	pNull=3.6e-5
Non-TCGA	*TP53BP2*	Missense	349.2	303	1.23	Prec=0.999
Non-TCGA	*TP53BP2*	LoF	34.6	12	3.79	pLI=6.8e-4
Non-TCGA	*PPP1R13L*	Synonymous	195.1	141	2.45	pNull=0.0014
Non-TCGA	*PPP1R13L*	Missense	334.6	240	2.58	Prec=0.960
Non-TCGA	*PPP1R13L*	LoF	19.0	6	2.95	pLI=0.039
Non-psych	*PPP1R13B*	Synonymous	192.2	176	0.74	pNull=3.9e-10
Non-psych	*PPP1R13B*	Missense	392.2	300	2.35	pRec=0.0033
Non-psych	*PPP1R13B*	LoF	33.8	4	5.04	pLI=0.997
Non-psych	*TP53BP2*	Synonymous	149.0	147	0.10	pNull=1.3e-5
Non-psych	*TP53BP2*	Missense	343.2	296	1.29	pRec=0.986
Non-psych	*TP53BP2*	LoF	33.7	10	4.01	pLI=0.014
Non-psych	*PPP1R13L*	Synonymous	191.7	140	2.36	pNull=8.0e-4
Non-psych	*PPP1R13L*	Missense	329.0	241	2.45	pRec=0.839
Non-psych	*PPP1R13L*	LoF	18.5	5	3.09	pLI=0.160

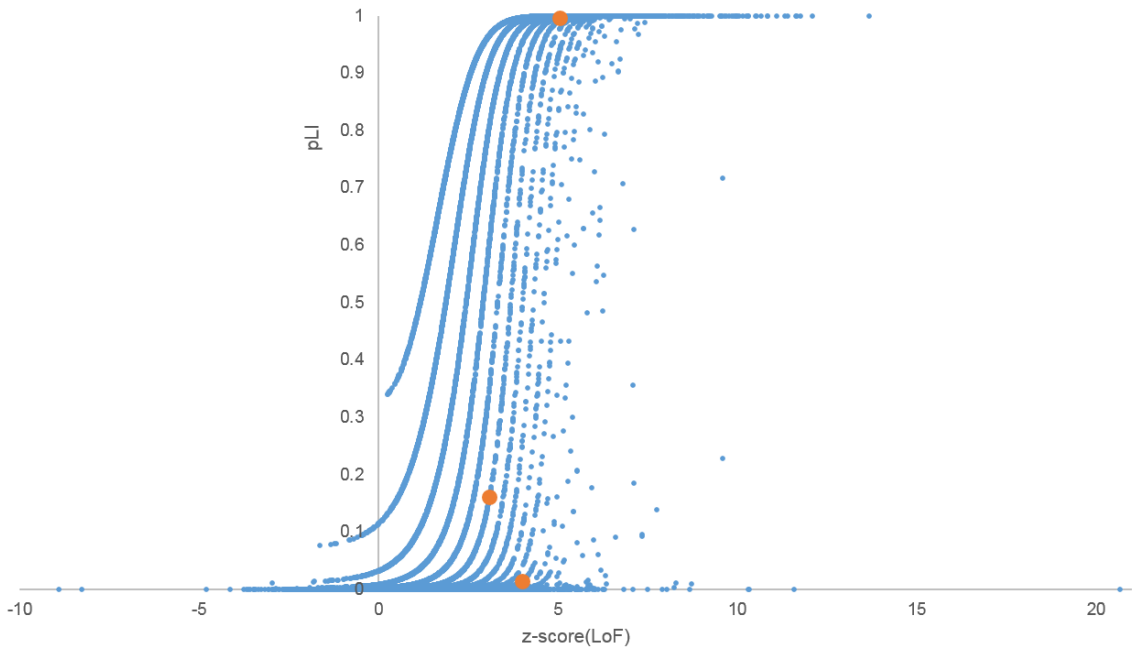


Figure 2.34 Relationship between pLI and LoF z-score for genes in the non-psychiatric ExAC set. Position of ASPP genes highlighted as orange dots.

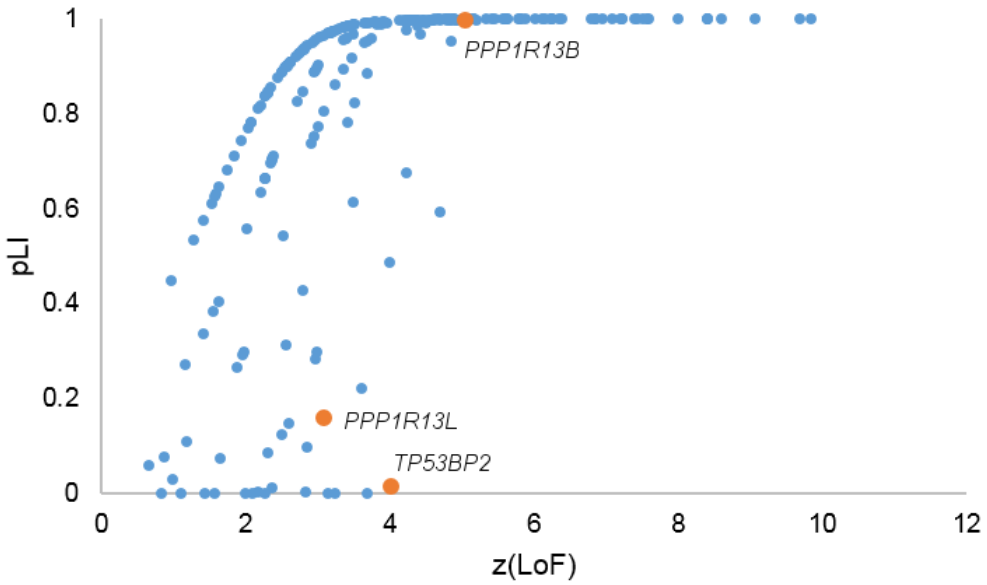


Figure 2.35 pLI vs z(LoF) for ClinGen haploinsufficient genes with sufficient evidence and the ASPP genes for comparison. Data taken from the ExAC non-psych set and ClinGen Dosage Sensitivity Map (retrieved 19/9/16). Some genes (n=7) were missing from the ExAC analysis; remaining n=242 genes plotted.

pLI is a Bayesian-derived score aimed at classifying all genes into 3 groups, and importantly, correcting for a bias in gene size whereby longer genes have higher z(LoF) scores (14). It is likely that this effort to exclude false positives

such as *TTN* produced a class of false negatives disqualified on the basis of their gene size. The ASPP family represents an interesting example of the limitations of the pLI score: whereas *TP53BP2* seems to be a haploinsufficient gene with incomplete penetrance with pLI<0.1 and thus a false negative, *PPP1R13B* has deletions segregating in the general population and thus despite its very high pLI score, is unlikely to be a fully penetrant haploinsufficient genes. This observation also suggests the possibility that certain LoF SNVs might have additional toxicity beyond affecting gene dosage, for example by causing aberrant splicing. It would be interesting to see the mRNA isoform spectrum in patients carrying LoF variants versus those carrying deletions and also normal controls.

2.12 Inverse *PPP1R13B* vs *TP53BP2* expression may have evolved to avoid dosage imbalance in *TP53BP2*-dosage-sensitive tissues

Homozygous deletion of *Trp53bp2* in mice is lethal prior to E16, therefore ASPP2 performs non-redundant functions during embryonic development. In humans, truncating germline variants in *TP53BP2* are observed less frequently than expected for a recessive pattern of effect, and homozygous truncating variants have never been observed. Hemizygous deletions spanning the *TP53BP2* locus are observed at an extremely low frequency in the general population. Therefore, it is safe to assume *TP53BP2* is also non-redundant in human development. The data presented above suggests *TP53BP2* is dosage sensitive with developmental abnormalities associated with both hemizygous deletions and duplications of the gene. What properties of *TP53BP2* make this gene sensitive to dosage?

Birchler et al. demonstrated that single-gene dosage-sensitive modifiers of the white eye colour gene in Drosophila are enriched in transcription factors, chromatin components and signal transduction proteins (339). Tumour suppressor genes were also found to be dosage sensitive. ASPP2 is a transcription co-factor that modifies transcriptional activity of multiple TFs in a promoter-specific manner. ASPP2 also modulates signal transduction of the RAS and Hippo pathways, and the gene is a haploinsufficient tumour suppressor in mice. Therefore, *TP53BP2* belongs to a functional class of genes that is enriched in dosage sensitive genes.

Tolerance of genic CNVs in humans is also negatively correlated with the number of protein-protein interactions of the corresponding protein products, i.e. highly promiscuous proteins are unlikely to be deleted or duplicated in the general population (340). ASPP2 is a protein-protein interaction (PPI) hub that contains multiple PPI domains such as the RAS-association domain, SH3 domain or Ankyrin repeats. In fact, ASPP2 is not known to contain any catalytic, DNA-binding or lipid-binding domains, suggesting its sole role is a protein interactor. In the BioGRID database, the average number of non-redundant interactions per human gene is 13.3 whereas 110 interactions were detected for ASPP2. Therefore, ASPP2 interacts with a higher number of proteins than a protein selected at random. On this basis, CNVs containing *TP53BP2* are less likely to be tolerated than CNVs containing less promiscuous genes.

We showed above that *PPP1R13B* and *TP53BP2* are spatiotemporally anti-coexpressed in the human developing brain. The most striking difference is observed in the ventricular and subventricular zones where *TP53BP2* is expressed at very high levels and *PPP1R13B* at low levels. Given the high homology of these genes, the likely evolutionary origin of *PPP1R13B* is a gene

duplication of *TP53BP2*. If so, the most ancient protein-coding *PPP1R13B* orthologue (i.e. not pseudogene) must have had an overlapping spectrum of protein-protein interactions. This is also consistent with the notion that ASPP2 is the ancient ASPP protein from which ASPP1 and iASPP were derived later in evolution. The most ancient ASPP is most likely to be dosage sensitive, by virtue of having the most conserved functions.

In yeast and *Arabidopsis*, it has been demonstrated that following whole-genome duplication events causing tetraploidisation, genes return to diploidy in a non-random fashion, with macromolecular complexes remaining for longer periods than other genes. Given *TP53BP2*'s dosage sensitivity, the introduction of a close homologue with an overlapping interactome would create an imbalance in ASPP2's protein-protein interactions resembling a pure duplication. One solution to this problem is the positive selection of mutations in both genes that uncouple the competing dosage-sensitive nodes. However, such changes take place over many generations. A more immediate solution involves modulating the two genes' expression such that *PPP1R13B* does not compete with *TP53BP2* in tissues and developmental time points where *TP53BP2* is dosage sensitive. This is akin to an epigenetic compensation for a duplication.

We propose that the observed inverse expression of *PPP1R13B* and *TP53BP2* in the developing brain evolved as a means to compensate for the imbalance introduced by the original gene duplication of *TP53BP2*. Given that *TP53BP2* duplication seems to be associated with neurological phenotypes, it is likely that CNS development is among the processes most sensitive to *TP53BP2* dosage. However, it would be interesting to investigate which other tissues display this pattern of expression. In addition to inverse expression, protein truncation is

another mechanism to minimise competition between homologous proteins. iASPP shares the Ank-SH3 domains and the conserved Pro-rich region, but possesses a unique N-terminal Pro-rich domain. If *PPP1R13L* evolved from a duplication of *TP53BP2* in an ancestral organism, the endowment with a unique N-terminal domain would disqualify iASPP from participating in ASPP2's junctional complexes, which rely on the N-terminal coiled-coil domain. It is ASPP2's extranuclear function which is required for the regulation of cell polarity via Par3 which is non-redundant in brain development. Overall, ASPPs seem to have evolved distinct patterns of expression and domain structures to avoid dosage imbalance in tissues sensitive to *TP53BP2* (or the ancestral ASPP) dosage.

2.13 Generation of *Ppp1r13b* knockout and conditional-knockout mice

In order to evaluate the non-redundant role of ASPP1 in mammalian development, a model system is needed. The International Knockout Mouse Consortium has generated libraries of targeting vectors and embryonic stem cells with targeted loci for several thousand genes. A targeting vector for *Ppp1r13b* was generated by EUCOMM and used to generate ES cell clones positive for the tm1a cassette (**Figure 2.36**). The tm1a allele has conditional potential and a lacZ reporter, allowing the transcriptional activity at the gene promoter to be assayed *in vivo* (**Figure 2.36**).

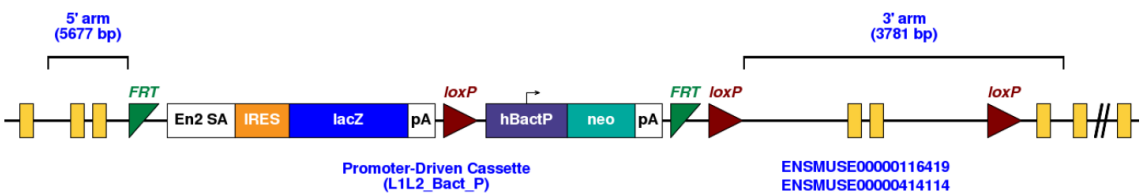


Figure 2.36 tm1a(EUCOMM)Hmgu allele structure

ES cells on the C57BL/6N background and positive for the Ppp1r13b<tm1a(EUCOMM)Hmgu> allele were obtained from EUCOMM. Two independent tm1a(EUCOMM)Hmgu allele-positive clones were received: codes HEPD0653_6_F08 (abbreviated F08) and HEPD0653_6_07 (abbreviated C07). Long-range PCR was performed to test for the presence of the 5' and 3' targeting homology arms (**Figure 2.36**) using genomic DNA purified from the ES cells. The presence of the correctly sized amplicons confirmed the targeting of the homology arms to the correct genomic location. This experiment confirmed data later released through EUMMCR (https://www.eummcr.org/order/cart) that shows both clones were positive for the correctly-integrated 5' and 3' arms.

The transgenic core led by Dr Ben Davies performed ES cells culturing, blastocyst injections and delivered 12 chimaeric offspring from F08-injected females and 9 chimaeric offspring from C07-injected females. The agouti coat colour was used to estimate degree of chimaerism. Eight of the twelve F08 chimaeras were >50% chimaeric whereas eight of the nine C07 chimaeras showed >50% chimaerism. The male chimaeras were mated with wild-type C57BL/6J females for approximately 6 months. No agouti offspring were obtained among 50 pups from the C07 chimaera matings but 21 agouti pups were obtained from the F08 matings.

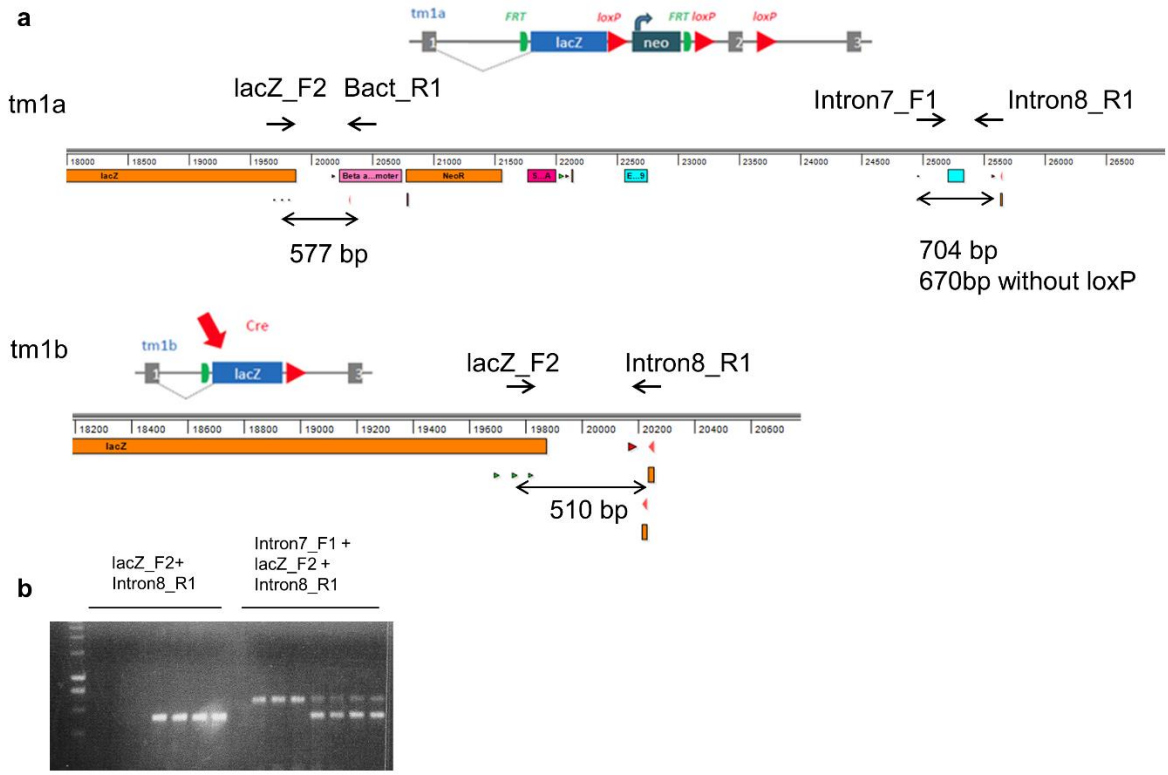


Figure 2.37 Genotyping strategy for tm1a and tm1b mice

The F08 pups were screened for the presence of the tm1a cassette using primers developed on the chimaeric parent males which contain some tm1a-positive cells (**Figure 2.37**). The tm1a heterozygous mice confirmed by genotyping were first used to set up crosses with PGK-Cre and Flp recombinase lines to allow for tm1a cassette manipulation. Subsequently, tm1a heterozygous-heterozygous matings were established to assess the phenotype of the homozygous offspring. Neonatal deaths were observed in multiple heterozygous x heterozygous matings. However, genotyping some of these pups revealed that neonatal death was not specific to tm1a/tm1a homozygous pups. This suggests the original ES cells may have contained chromosomal abnormalities independent of the targeted allele which can cause neonatal lethality and/or developmental abnormalities.

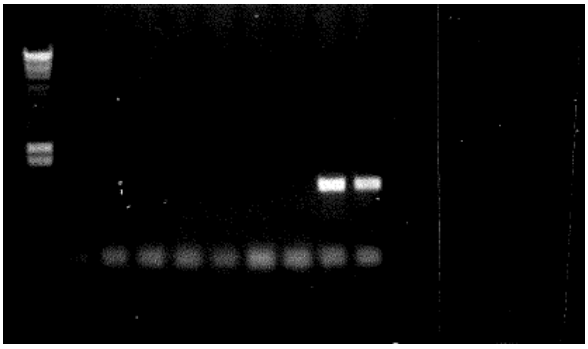


Figure 2.38 Genotyping PCR to detect presence of the tm1b allele. Primers used were lacZ_F2 and Intron8_R1 as depicted in Figure 1.x. Last two lanes were positive for tm1b.

To avoid *Ppp1r13b*-independent abnormalities, *Ppp1r13b*$^{tm1a/+}$ mice were back-crossed to a pure C57BL/6J line (S. Zhong, work in progress). In addition, *Ppp1r13b*-knockout mice (carrying the *Ppp1r13b*<tm1b(EUCOMM)Hmgu> allele) were obtained through the tm1a – PGK-Cre cross and the deletion confirmed by PCR (**Figure 2.38**). Similarly, mice carrying the *Ppp1r13b*<tm1c(EUCOMM)Hmgu> allele, which have a loxP-flanked copy of *Ppp1r13b*, were generated by from the tm1a – Flp cross.

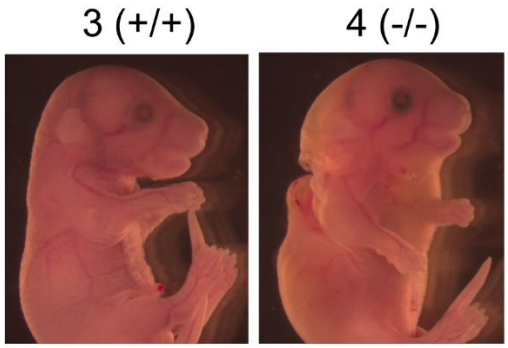


Figure 2.39 Oedema in E17.5 *Ppp1r13b*$^{-/-}$ embryos. The embryo's skin shows dissection damage, however the massive oedema is still visible. The numbers shown are arbitrary indices.

Heterozygous matings were established using the tm1b heterozygous mice. Two litters were examined at E17.5 embryonic age. One litter contained no tm1b/tm1b homozygous, i.e. *Ppp1r13b*$^{-/-}$ mice. The second litter contained two

tm1b/tm1b mice including one showing severe oedema (**Figure 2.39**). This result agrees with an earlier report by Hirashima et al. of subcutaneous oedema in mid-gestation *Aspp1*$^{-/-}$ embryos (261).

Discussion

We have described the discovery and assessment of candidate pathogenic variants in congenital diseases with varying genetic architectures (**Table 2.9**). 1q41q42 deletions have an established pathogenicity and the present work refined the likely contribution of *TP53BP2* haploinsufficiency towards the clinical phenotype. *TP53BP2* deletions are absent from published cohorts of disease-free individuals and occur in patients as *de novo* variants, suggesting these deletions have a high penetrance.

Table 2.9 Summary of variants studied in this chapter and evidence of pathogenicity

Variant	Phenotype	*De novo* / inherited	ACMG evidence codes (SNV)	ACMG classification	Notes
1q41q42 deletions	1q41q42 microdeletion syndrome; encephalocele	Most *de novo*	-	Pathogenic	Brain MRI consistent with *TP53BP2* dosage effect
1q41q42 duplication 1:223820857-224076362dup	Microcephaly; developmental delay, seizures, vision problems	Inherited (maternal)	-	Uncertain significance	Brain MRI consistent with *TP53BP2* dosage effect
1q41q42 duplication 1:222694279-224034322dup	Microcephaly, developmental delay, craniofacial, dental abnormalities, abnormal behaviour	Inherited (maternal)	-	Uncertain significance	Distal breakpoint within *TP53BP2*
PPP1R13B R72Q	Childhood-onset schizophrenia	*De novo*	PS2, PM2	Likely pathogenic	
PPP1R13B 14:104216291 AC>A	ASD	*De novo*	PS2, PM2	Likely pathogenic	
PPP1R13B G881R	ASD	*De novo*	PS2, PP3	Uncertain significance	

Chromosomal deletion-associated human disorders often have variable phenotypes and severity among individuals, and the vast majority of pathogenic deletions include multiple genes, hampering the identification of causal genes. Experimental model systems are, therefore, needed to aid the identification of causal genes in human developmental disorders. Here we have identified *TP53BP2* as a candidate gene responsible for brain abnormalities in 1q41q42 microdeletion syndrome and show that ASPP2-deficient mice also develop CNS abnormalities with 100% penetrance, and phenotypic differences depending on background. Notably, *TP53BP2* is the only gene in the critical region for brain abnormalities of the syndrome for which the gene-deficient mouse model has a CNS phenotype.

Recently, a 1q41q42 microdeletion case with brain abnormalities was reported whose deletion included the last exon of *FBXO28*, suggesting *FBXO28* might also be responsible for these abnormalities (341). A mouse deficient in FBXO28 is currently unavailable. Importantly, Case 6 in our report (**Figure 2.1, Table 2.1**) showed occipital encephalocele yet harbours a deletion that includes *TP53BP2* but excludes *FBXO28*. *FBXO28* lies outside the existing SRO for brain abnormalities, but it is possible that there are multiple dosage-sensitive genes for brain development in the 1q41q42 region and future studies are needed to investigate this. For example, patients with non-overlapping *DISP1* and *FBXO28* deletions both showed seizures (99, 100), despite the fact that *DISP1* is not contained in the critical region of the syndrome.

Few mouse models exist that are deficient in human 1q4-located genes and phenocopy the abnormalities associated with 1q4 deletions to a significant extent.

The ASPP2-deficient mice described here show craniofacial, eye, heart and urogenital abnormalities in addition to CNS defects, all of which are associated with 1q41q42 deletions. Interestingly, among the 31% of patients with *TP53BP2* deletions who presented with eye abnormalities, strabismus and optic nerve hypoplasia were common. Our HREM analysis of *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos revealed hypoplastic or absent optic nerve and abducens nerve among 33% and 67% of B6 and BALB/c embryos, respectively. Given that the abducens nerve controls eye movement through the lateral rectus muscle, it would be interesting to test whether patients with strabismus and *TP53BP2* deletions have any abducens nerve abnormalities. We note that our ability to determine penetrance precisely is limited by the low number of *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos (n = 3) on the BALB/c background analysed by HREM, due to the demanding nature of this technique. Overall, these observations might explain the significant phenotypic heterogeneity associated with 1q41q42 deletions: there are likely multiple dosage-sensitive genes in the region whose haploinsufficiency is both incompletely penetrant and variably expressed. The observation of NTDs in some of the ASPP2-deficient mice suggests it would be informative to investigate whether mutations involving *TP53BP2* might contribute to NTDs in humans.

To date, three different *Trp53bp2*-deficient mouse lines have been generated: exon 3 deletion mice reported by our group (*Trp53bp2*$^{\Delta 3/\Delta 3}$) (ref. 105), exon 4 deletion mice generated by the International Mouse Knockout Consortium (IKMC) (*Trp53bp2*$^{\Delta 4/\Delta 4}$) (ref. 342), and mice with deletion of exons 10-17 (*Trp53bp2*$^{\Delta 10\text{-}17/\Delta 10\text{-}17}$) (ref. 104). Homozygous deletion of exon 4 or exons 10-17 is lethal prior to embryonic day 9 with 100% penetrance, whereas deletion of exon 3 results in neonatal lethality with background-dependent penetrance. The exon 3

deletion is an in-frame deletion causing only the major isoform of *Trp53bp2* to be deleted completely; low levels of shorter forms of *Trp53bp2* can be expressed. Therefore *Trp53bp2*$^{\Delta 3}$ acts as a hypomorphic allele, which has allowed us to discover developmental effects of ASPP2 deficiency that could not be observed with full-knockout alleles due to their early embryonic lethality.

Imaging of heterozygous *Trp53bp2*$^{\Delta 3/+}$ embryos showed a lower frequency of structural abnormalities than observed in patients with heterozygous *TP53BP2* deletions. One possible explanation is that the exon 3 deletion is a hypomorphic allele, compared to the complete deletion allele in the patients. In addition, more than one gene is deleted in each 1q41q42 microdeletion. Therefore, the concomitant deletion of multiple haploinsufficient genes might exacerbate the phenotype, as is the case in other known contiguous gene deletion syndromes (44, 45).

At the protein level, the ability of ASPP2 to regulate apicobasal polarity and p53-dependent apoptosis is likely important for its role in neural tube closure. Recently, inappropriate activation of p53 during development was linked with the pathogenesis of CHARGE syndrome, a disorder characterized by abnormalities including developmental delay, heart defects, coloboma, cranial nerve abnormalities, choanal atresia and inner ear abnormalities (343, 344). The phenotypic spectrum in the CHARGE syndrome mouse model shows substantial overlap with the ASPP2-deficient mice (including neural tube defects, short lower jaw, coloboma and heart defects), suggesting dysregulated p53 activity might be responsible for these phenotypes in ASPP2-deficient mice. However, while p53-deficient and p53-overactivated embryos show primarily exencephaly (131, 343),

ASPP2-deficient embryos develop mainly spina bifida or craniorachischisis, pointing to the importance of p53-independent functions of ASPP2.

ASPP2 is also an important regulator of epithelial plasticity and apicobasal polarity (106, 130). Recent studies of Grainyhead-like 2 mutants have shown that dysregulated epithelial mesenchymal transition (EMT) can cause NTDs (345). It is possible that loss of epithelial character together with the loss of apicobasal polarity disrupt neural tube closure in ASPP2-deficient embryos. ASPP2-deficient embryos also showed overgrowth of neural tissue both in the brain and the spinal cord. The surplus of neuroepithelial cells might also contribute to NTDs and potentially other CNS phenotypes in these embryos through mechanical obstruction. Regardless of the exact cellular mechanism, the observed phenotypical similarities between ASPP2 deficiency in mice and humans highlight the importance of ASPP2 in controlling CNS development in both mice and humans.

These results may have implications to other diseases. Rare variants in *TP53BP2* have been linked with Tetralogy of Fallot (ToF), a congenital heart disease featuring VSD (346). Given the observation of VSD in ASPP2-deficient mouse embryos, patients with *TP53BP2* deletions and patients with rare missense *TP53BP2* variants, ASPP2 likely plays a causal role in preventing VSD in humans. It would be interesting to investigate the mechanism by which pathogenic *TP53BP2* variants disrupt heart development leading to the VSD phenotype.

Interestingly, *TP53BP2* duplications, associated with microcephaly and developmental delay, were found to be likely pathogenic. Consistently, *TP53BP2* duplications are absent from public databases of normal individuals and *TP53BP2*

is within the top 1% of genes predicted to exhibit triplo-sensitivity in the ExAC cohort. Ventricular dysmorphism is observed in both *TP53BP2* duplication and deletion patients. However, whereas *TP53BP2* deletions are associated with enlarged bodies and volumes of the lateral ventricles, *TP53BP2* duplication patient showed an effacement of the bodies of the lateral ventricles. These opposing effects of *TP53BP2* deletion and duplication on LV morphology may be related to ASPP2's role in controlling the differentiation of ganglionic eminence cells into the striatum, resulting in too few and too many striatal cells, respectively. Spatiotemporal pattern of *TP53BP2* expression in the developing brain supports the notion of *TP53BP2*'s dosage sensitivity in the ganglionic eminence, in which *TP53BP2* levels are high and very consistent between individuals. In support of this hypothesis, ganglionic eminences in *Trp53bp2*$^{\Delta 3/\Delta 3}$ mouse embryos show loss of cell polarity and general cell disorganisation. In addition, *Trp53bp2*$^{\Delta 4/+}$ males showed hyperactivity in voluntary movement, and all *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice which survive birth exhibit hopping gait, presenting a possible link between ASPP2 and striatal control of motor function.

Of note is the increased severity of ASPP2 deficiency-related phenotypes in mice, as *TP53BP2* deletions have been observed at an approximate 2.3 male-to-female ration among 1q41q42 microdeletion patients. Together with the striking sexual dimorphism in *TP53BP2* expression within human brains, these data suggest *TP53BP2* deletions may be more penetrant in males than females. Given that *TP53BP2* deletions are absent from existing cohorts of disease-free individuals, large scale studies of CNVs in the general population are necessary to ascertain whether *TP53BP2* deletions can occur in asymptomatic females.

Additionally, *de novo* single-nucleotide variants in *PPP1R13B*, which encodes ASPP2's sibling protein ASPP1, were identified in autism and schizophrenia patients. We showed that *PPP1R13B* is expressed in a spatiotemporally inverse manner with respect to *TP53BP2* in the developing brain, most strikingly in the ventricular zone where *TP53BP2* levels are high, which may be reflect an evolutionary epigenetic compensation for *TP53BP2*s dosage sensitivity in this tissue. *PPP1R13B* is expressed in neurons and spatiotemporally co-expressed with synaptic genes in the human brain. Accordingly, three *de novo* mutations in *PPP1R13B* have been identified in childhood-onset schizophrenia and autism spectrum disorder, diseases associated with synaptic dysfunction. ASPP1-deficient mice have been generated and will aid in the characterisation of ASPP1's function in the brain and other tissues.

Taken together, *TP53BP2* and *PPP1R13B* show distinct patterns of expression and distinct disease associations. In accordance with being an orthologue of the putative ancestral ASPP, *TP53BP2* shows dosage sensitivity in development and *TP53BP2* CNVs are associated with congenital disorders of the brain, and potentially heart and urogenital abnormalities. *PPP1R13B* is more redundant in development, however its neuronal expression implicate ASPP1 *de novo* mutations in neurodevelopmental diseases with a synaptic dysregulation.

Materials and Methods

Patient recruitment

Patients with 1q4 deletions and duplications were enrolled and de-identified data collected following the UK NHS Research Ethics Committee approval (09/H0606/78, Amendment 3), and in accordance with the Institutional Review Board protocols of the participating institutions.

Comparative genomic hybridisation

DNA isolated from patients' blood were analysed by chromosomal microarray following the participating institutions' standard protocols. With the exception of Case 7, all new 1q microdeletion and microduplication cases have been deposited on the DECIPHER database (https://decipher.sanger.ac.uk/).

Brain MRI and morphometry

De-identified brain MRI and/or CT scans were obtained in DICOM format from radiology departments where the scans were recorded. *Normal MRI controls used in the preparation of this article were obtained from the Pediatric MRI Data Repository created by the NIH MRI Study of Normal Brain Development.* To quantify lateral ventricles (LV) volume, MRI scans were first converted to Nifti format using MRIConvert (University of Oregon, USA; http://lcni.uoregon.edu/downloads/mriconvert) and subsequently analysed by ALVIN (347). Multiple scans in each series were used to compute LV volume to produce Figure 2b-left. Axial T1 scans only were used to compute LV volumes for

genotype comparison as shown in Figure 2b-right. Out of the fourteen 1q41q42 microdeletion MRI scans, two patient scans (Rosenfeld 9 and Case 7) were unsuitable for morphometry: Case 7 is a prenatal case and the scan for Rosenfeld 9 cannot be converted into Nifti. Therefore, the morphometric analysis of LV volume was performed with 12 patient scans.

Mice

Experiments with mice received UK Home Office approval, Project licence number 30/2862 and fully complied with the UK Home Office guidelines. For this study, *Trp53bp2*$^{\Delta 3}$ mice in a pure C57BL/6J, mixed C57BL/6Jx129SvJ and pure BALB/c backgrounds were crossed to generate wild type, *Trp53bp2*$^{\Delta 3/+}$ and *Trp53bp2*$^{\Delta 3/\Delta 3}$ embryos (348). Pregnant females were sacrificed at the indicated time points during pregnancy and embryos were collected and genotyped using the following primers: 5'-CTCCACCCCAGGAAATTACA-3' (intron3), 5'-CGGTTTGGAAGTCAAAGGAA-3' (exon 3), and 50-GGACCGCTATCAGGACATA-30 (neomycin resistance gene).

Embryo sexing

The following primers for ZFY were used for sexing by PCR: 5'-GACTAGACATGTCTTAACATCTGTCC-3' and 5'-CCTATTGCATGGACAGCAGCTTATG-3'. The following GADPH primers were used as controls for the PCR reaction: 5'-TGGCTACAGTAACCGAGTGGT-3' and 5'-CTTTGAGTGGAAGCCGAAGT-3'.

Mouse micro-computed tomography (micro-CT)

Embryos generated from mice as described above were first sacrificed on ice-cold PBS, washed in phenol red-free Hank's salts with EDTA, removed from their yolk sac, their umbilical cord cut, and the freed embryo fixed in 4% formaldehyde for at least 24 h. Each embryo was then immersed in 2 ml Lugol solution (L6146, Sigma-Aldrich, St Louis, MO, USA) for 72 hours at 4°C. All incubations included the replacement of staining solution after the first 24 hours. Four embryos were embedded in 1% agarose (Web Scientific, Crewe, UK) in a dedicated sample holder and scanned using a micro-CT scanner (SkyScan 1172, Bruker, Kontich, Belgium) at a nominal isotropic pixel size of 4.9 µm (scan parameters: X-ray source voltage – 48 kV; current – 208 mA; filter – a 0.5 mm Al filter; number of projections – 900). The total scan time was 8 hours for four embryos.
The images projections were reconstructed with the scanner-built-in reconstruction software 'NRecon' and downsampled offline by a factor of two, resulting in a 9.8 µm isotropic voxel size. Image analysis (i.e. 3D segmentation and rendering) was performed in Amira (FEI Ltd, Hillsboro, OR, USA).

High resolution episcopic microscopy

HREM images of E14.5 embryos were obtained as previously described (324). We used the episcopic procedures for sectioning and image capture using a previously described arrangement of microtome and magnification optic (349).

Mouse echocardiography

Adult female BALB/c *Trp53bp2*$^{\Delta 3/\Delta 3}$ mice (n=8) and age matched wild type controls (n=7) received echocardiographic examination at a mean age of 28 weeks. Mice were lightly anaesthetised using 1-1.5% isoflurane in 100% medical oxygen, kept

warm on a homeothermic blanket and imaged using a VisualSonics Vevo 2100 (Toronto, Canada) with 22-55 MHz transducer. B-mode trans-thoracic short-axis and long-axis images were obtained to assess left ventricular volumes and function and pulmonary artery Doppler flow as a measure of right-sided function. Images were obtained and analysed using Vevo 2100 v1.6.0 software by a single operator blinded to genotype.

Trp53bp2 exon 4 deletion mice (Trp53bp2$^{\Delta 4}$)

Publicly available phenotypic data for *Trp53bp2*$^{\Delta 4/\Delta 4}$ and *Trp53bp2*$^{\Delta 4/+}$ mice were mined from the International Mouse Phenotype Consortium (IMPC) URL https://www.mousephenotype.org/data/genes/MGI:2138319 release 4.3, last updated 12 May 2016, and from Phenoview (https://www.mousephenotype.org/phenoview), updated 18 May 2016. Raw data were re-plotted using Graphpad Prism version 6.05 for Windows, (GraphPad Software, La Jolla, CA, USA). Experiment design by IMPC followed protocols of the International Mouse Phenotyping Resource of Standardized Screens (IMPReSS) (https://www.mousephenotype.org/impress). For statistical testing, results from IMPC's Mixed Model framework, linear mixed-effects model were adopted. P values below 0.05 were deemed significant.

Visualisation and statistical analysis

Data for panels in Figure 2.1a and 2.1c; were generated using the UCSC Genome Browser (http://genome.ucsc.edu/) (350), genome release hg19. Plots in panels Figure 2.1b, 2.3b, 2.2b and Figure 2.12a-d were generated using Graphpad Prism version 6.05 for Windows, (GraphPad Software, La Jolla, CA, USA). Chromosome

plots in Figure 2.8 were generated using Phenogram (http://visualization.ritchielab.psu.edu/phenograms/plot) (351). Statistical tests were performed in GraphPad Prism (see above) and R (352).

Human brain expression data

Human brain mRNA expression data was mined from the Allen Brain Atlas (http://www.brain-map.org/). 3D visualisation of expression data was carried out using Brain Explorer available through the same portal.

Human germline variant data

Expected and observed frequencies of human truncating, missense SNVs were retrieved from the ExAC Browser (http://exac.broadinstitute.org/), v0.3.1 (14).

Primer design and DNA visualisation

DNA sequence manipulation including primer design and gene visualisation was performed using SeqBuilder (DNASTAR; Madison, WI, USA). Primer melting temperatures were computed using Oligocalc (http://biotools.nubic.northwestern.edu/OligoCalc.html) (353).

Ppp1r13b targeted allele mouse generation

Ppp1r13b knockout and conditional-knockout mice were generated as described in section 1.13. Targeting vector cloning and ES cell generation was performed by the EUCOMM consortium at the Helmholtz Zentrum Munchen (HMGU). ES cell culturing and blastocyst injections into pseudopregnant recipients were performed by the Trasgenic Core led by Dr Ben Davies. Chimaeric males were mated with

wild type females until F1 offspring obtained from multiple litters. For generation of tm1b and tm1c allele mouse lines, the PGK-Cre and Flp transgenes were removed by back-crossing to wild type mice and selecting Cre and Flp transgene-negative tm1b- and tm1c-positive offspring, respectively.

CHAPTER 3: RARE VARIANTS IN NEURAL TUBE DEFECTS

In this chapter, we explore the hypothesis that rare variants in *TP53BP2* and/or its direct binding partners *TP53*, *TP73* and *PARD3* contribute to NTD risk in humans. A case-control targeted sequencing study is performed in a cohort of 340 cases comprising mostly prenatal NTDs, and 240 controls of matched age and ethnicity.

RESULTS AND DISCUSSION

3.1 Rare nonsynonymous *TP53BP2* variants are present in both neural tube defect cases and matched controls

We showed in Chapter 2 that *TP53BP2* deletions are associated with NTDs and *Trp53bp2* deficiency causes NTDs in mice. Therefore, we decided to test whether mutations in the *TP53BP2* gene are also associated with NTDs in humans. 340 NTD cases of East Asian ethnicity and 240 matched controls were subjected to targeted next-generation sequencing of the *TP53BP2* gene locus. We focused our attention on rare variation of potentially large effect, rather than common variants of small effect. After filtering out common variants and variants of low-confidence we observed 4 distinct non-synonymous rare variants in the cases and 3 in the controls (**Figure 3.1a**). These SNVs were validated by Sanger re-sequencing (**Figure 3.2a**).

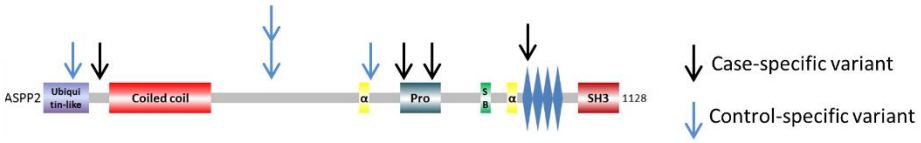
There was no significant burden of *TP53BP2* rare variants in cases compared with the controls. We also observed no difference in the distribution of missense mutations along the protein sequence of ASPP2 (**Figure 3.1b**). All four case-specific variants are novel, that is, not previously reported in ExAC or other

publicly available databases. In contrast, only 2 of the 3 missense variants in controls were novel. Therefore, we tested whether the case-specific variants had a deleterious effect on ASPP2 function. Given that ASPP2 was originally discovered as a binding partner of p53, the ability of mutant ASPP2 proteins enhance p53-dependent transactivation of the *BAX* promoter was assayed. All case-specific mutants were still able to stimulate p53-dependent transactivation of *BAX*, however ASPP2 K102E reached a limit of approximately 18-fold luciferase activity over the baseline and this activity did not increase with higher doses of ASPP2, whereas wild-type ASPP2 and the other 3 mutants were able to achieve higher degrees of transcriptional activity (**Figure 3.1c**). This resulted in a statistically significant reduction in trans-activation activity for ASPP2 K102E at 1000 ng plasmid used (**Figure 3.1c**).

A

Gene Name	Variant (genomic, hg19)	Variant (protein)	Frequency in cohort / controls (n = 337 / 240)	Novel?	NTD clinical details	Functional assay
TP53BP2	chr1:223998183T>C	K102E	1 / 0	YES	Occipital encephalocele; spina bifida aperta	**Reduced activity on *BAX***
	chr1:223984144A>C	I693M	1 / 0	YES	Craniorachischisis	Active on *BAX*
	chr1:223983881G>A	S781L	1 / 0	YES	Spinal dysraphism	Active on *BAX*
	chr1:223981006T>G	E948A	1 / 0	YES	Anencephaly, spina bifida cystica	Active on *BAX*
TP53BP2	chr1:224001983C>T	R77H	0 / 1	YES	-	-
	chr1:223987705G>A	R455C	0 / 2	NO	-	-
	chr1:223984238T>C	N662S	0 / 1	YES	-	-

B



C

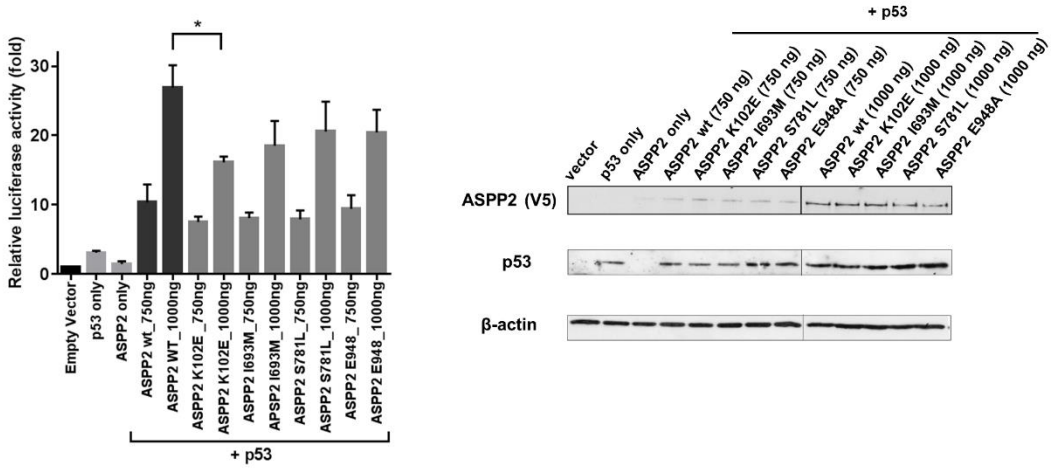
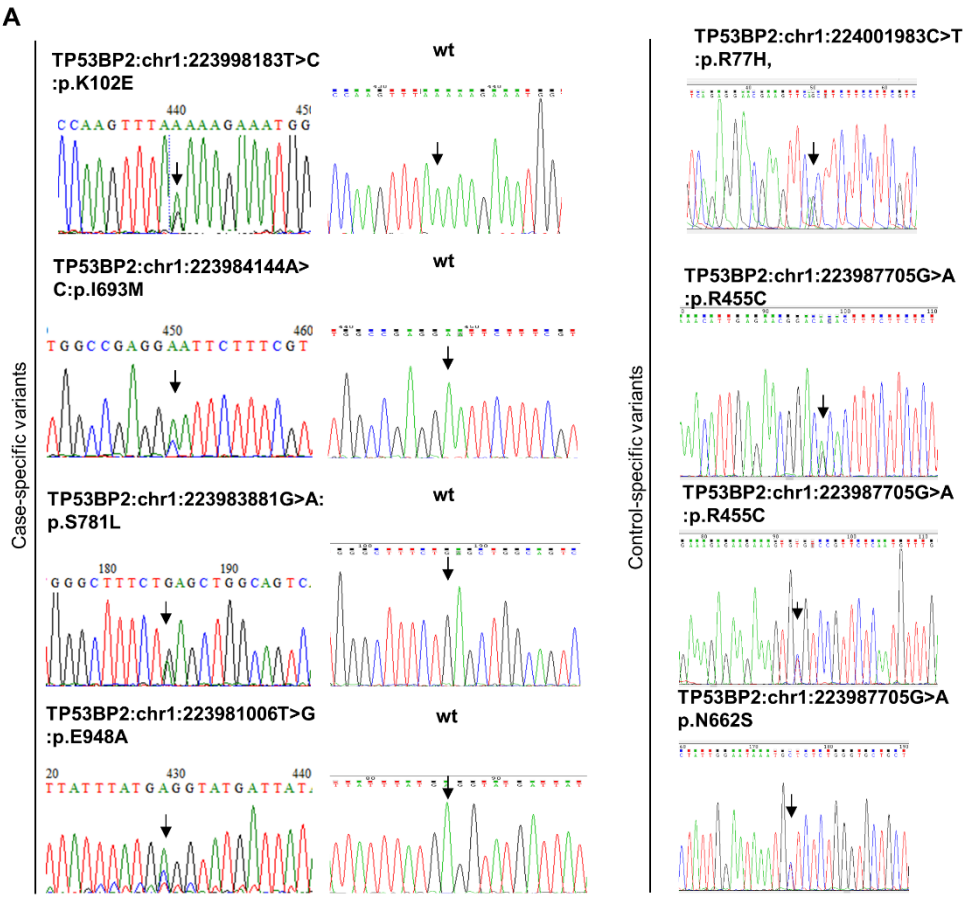


Figure 3.1 Rare nonsynonymous variants in *TP53BP2* identified in neural tube defect patients and matched controls. (a) details on nonsynonymous variants found and the associated clinical phenotype; (b) domain map of missense variant positions in cases (black) and controls (blue); Novel = variant not reported in public databases including dbSNP and ExAC; (c) luciferase assay of p53 co-transactivation of the BAX promoter: black = wild type ASPP2; dark grey = ASPP2 mutants; Saos2 cells used with the dual *Renilla*-firefly luciferase system. Pro, proline-rich region; α, alpha-helical domain; *, $p < 0.05$, Student's two-tailed t-test.

A



B

Gene Name	Variant (protein)	Variant (genomic, hg19)	Known SNP?	NTD case specific?	Control-specific?	PolyPhen V2 (Humdiv)	PROVEAN
TP53BP2	K102E	chr1:223998183T>C	-	YES		0.027 (benign)	-1.069 (Neutral)
	I693M	chr1:223984144A>C	-	YES		0.716 (poss. damaging)	-0.905 (Neutral)
	S781L	chr1:223983881G>A	-	YES		0.020 (benign)	-1.406 (Neutral)
	E948A	chr1:223981006T>G	-	YES		**0.984 (prob. damaging)**	**-4.952 (Del.)**
TP53BP2	R77H	chr1:224001983C>T	-		YES	1.000 (prob. damaging)	-2.036 (Neutral)
	R455C	chr1:223987705G>A	-		YES	0.999 (prob. damaging)	-1.857 (Neutral)
	N662S	chr1:223984238T>C	-		YES	0.001 (benign)	0.002 (Neutral)

Figure 3.2 Re-sequencing of rare nonsynonymous variants discovered in NTD cases and matched controls. (a) Sanger re-sequencing traces for case-specific variants with matching wild-type controls (middle); control-specific variants (right). (b) Details on computational predictions of variant effect by PolyPhen v2 and PROVEAN.

3.2 A burden of missense *TP53* and *TP73* variants in neural tube defects

We next asked whether proteins directly modulated by ASPP2 are disrupted by rare variants in NTDs. ASPP2 enhances transcriptional activity of p53

family transcription factors on pro-apoptotic promoters. P53 and p73 have both been implicated in CNS development: p53-deficient and mutant mice show NTDs; p73-deficient mice show defects in CNS development. In contrast, p63 is dispensable in CNS development (354). Therefore, we sequenced *TP53* and *TP73* genes in human NTD cases by NGS. Focusing on rare nonsynonymous variants, 9 variants case-specific variants were validated by Sanger re-sequencing (**Table 3.1; Figure 3.3**) whereas only one variant in controls. In *TP53* we identified 2 nonsynonymous variants in a total of 3 NTD cases versus no nonsynonymous variants in controls (**Table 3.1**). Together, these 12 missense variants in cases versus 1 in controls represent a significant burden of rare variants in the two p53-family transcription factors in NTDs ($p = 0.0197$, Fisher's exact two-tailed test). However, this association would not survive multiple hypothesis testing correction. Therefore, a larger cohort is required to confirm the association of *TP53* and *TP73* rare variants with NTD risk.

Five of the case-specific variants are novel. One of these, p73 V190F, showed near-complete loss of transcriptional activity on the *BAX* promoter (**Figure 3.4**). This mutation lies within the DNA-binding domain and targets a highly conserved residue, suggesting V190F disrupts p73's ability to bind DNA. Although the other mutants were transcriptionally active at similar levels as wild-type p73 within the margin of error, this result demonstrates that some NTD patients harbour loss-of-function mutations in *TP73*.

Table 3.1 Details on rare *TP53* and *TP73* variants discovered in NTD cases. Genome reference is hg19; Refseq reference transcript: TP73:NM_005427 except for R414W which affects a residue absent in the major isoform of *TP73*.
*TP73:NM_001204185:exon11:c.1240C>T: p.R414W, TP73:NM_001204187:exon11:c.1240C>T:p.R414W; For AC of novel variants, AC for the nearest neighbouring variant was used

Gene	SNV (genomic)	SNV (protein)	SNP ID	gene region	Sanger validated?	Freq cases (null:het:hom)	Allele count (cases)	Freq controls (null:het:hom)	Allele count (ExAC-all)	Allele count (ExAC-East Asian)	Variant type	Internal case ID
TP53	chr17:7577102T>C	G279E		exonic	yes	371|1|0	1/744	221|0|0	0/117922	0/8404	novel rare	TJ-QS75
TP53	chr17:7573954A>G	E358V		exonic	yes	367|2|0	2/738	213|0|0	1/116208	0/8486	rare	A2364; SYZ1
TP73	chr1:3644710C>A	P335T		exonic	yes	305|1|0	1/612	50|0|0	2/118934	1/8568	rare	A1408
TP73	chr1:3639971C>A	Q224K		exonic	yes	340|1|0	1/682	223|0|0	0/118930	0/8588	novel rare	A1489
TP73	chr1:3647593C>T	P483L		exonic	yes	337|1|0	1/676	213|0|0	0/114016	1/8400	rare	A2111
TP73	chr1:3649438T>A	I569N		exonic	yes	340|1|0	1/682	223|0|0	0/107654	0/8138	novel rare	A2554
TP73	chr1:3638723G>T	V190F		exonic	yes	338|1|0	1/678	174|0|0	5/119526	0/8620	novel rare	A2563
TP73	chr1:3648061G>A	E507K		exonic	yes	340|1|0	1/682	199|0|0	0/120876	0/8648	novel rare	A2841
TP73	chr1:3648065A>G	Y508C	rs369898417	exonic	yes	340|1|0	1/682	199|0|0	1/120876	0/8648	rare	A2882
TP73	chr1:3647534C>T	R414W*	rs150268231	exonic	N/A	337|1|0	1/676	213|0|0	17/102000	0/7576	rare	A1477
TP73	chr1:3646709G>A	V448M		exonic	yes	338|1|0	1/678	201|0|0	0/23724	0/1270	novel rare	A1639

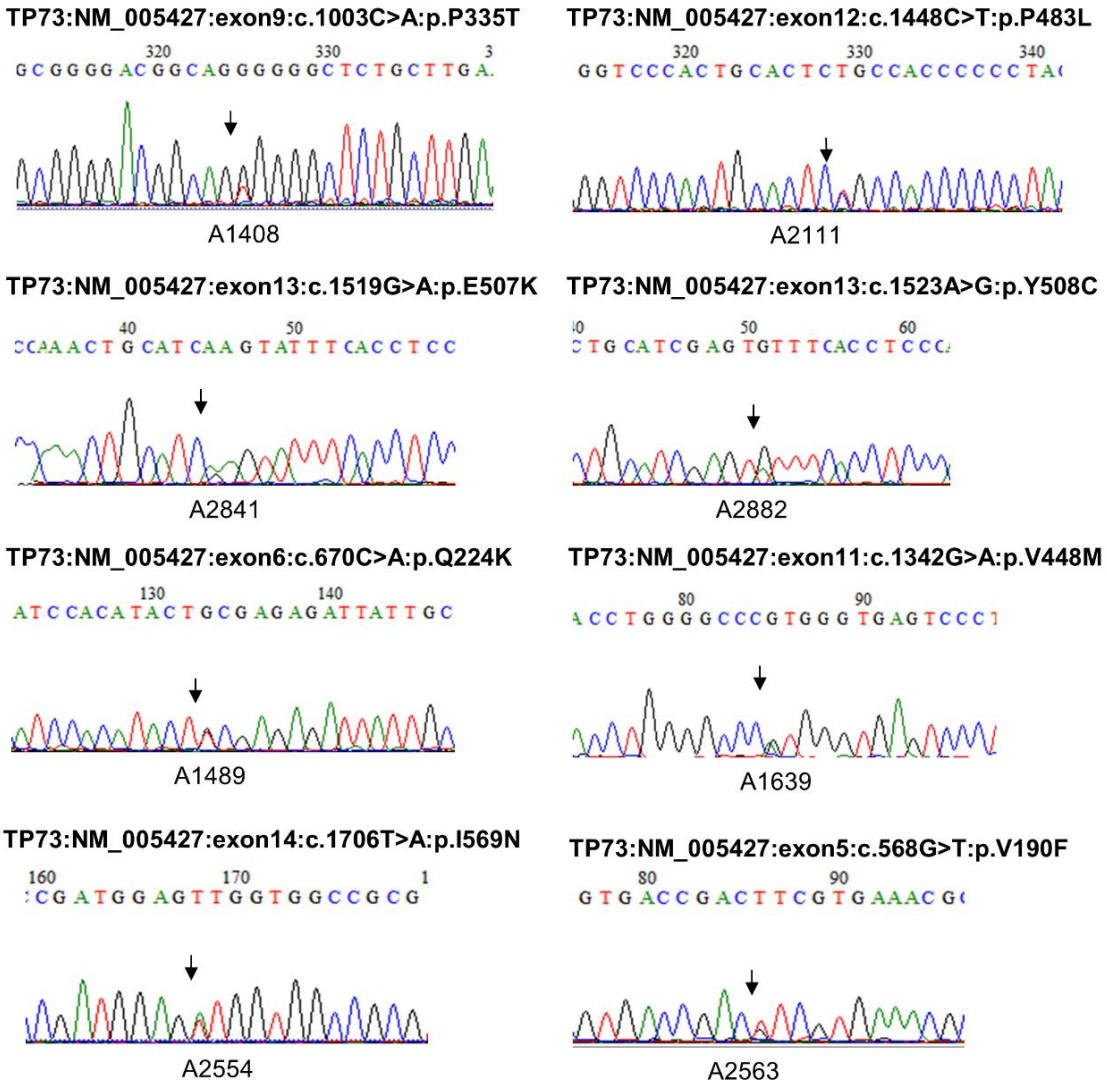


Figure 3.3 Sanger re-sequencing of case-specific nonsynonymous rare variants in *TP73*.

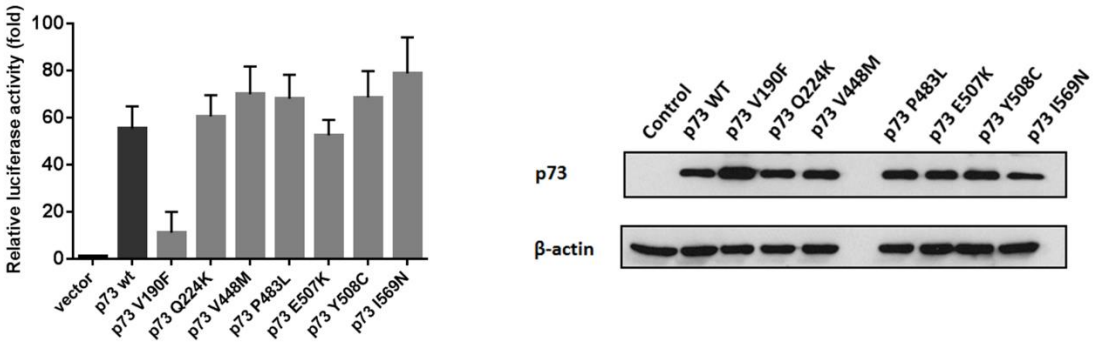


Figure 3.4 Transcriptional activity of NTD patient-derived p73 mutants. Left: relative luciferase activity on the BAX promoter; right: protein levels of the transfected mutants measured by Western blot.

3.3 A missense *TP53* variant is significantly enriched in neural tube defects and absent in unaffected siblings

When referenced to the major p53 isoform, the case-specific p53 variants identified encode p53 G279E and E358V. G279E is a DNA binding domain mutation observed in over 50 human tumours (source: IARC TP53 database; http://p53.iarc.fr/) as a somatic mutation, but absent from ExAC and other public databases, meaning its germline occurrence in the general population is less than 1 in 60,000. In our assay of *BAX* promoter transactivation the mutant protein showed no transcriptional activity (**Figure 3.5**). The recurrence of the mutation in the somatic cancer genome together with its absence or low abundance in the germline demonstrate this loss-of-function mutation has pro-tumour effects and is likely deleterious in normal tissues. Therefore, we sequenced the proband's parents and observed that the variant was paternally inherited from an unaffected father (**Figure 3.6**). This observation demonstrates that the mutation is not completely dominant, however it could be a variant that increases NTD risk.

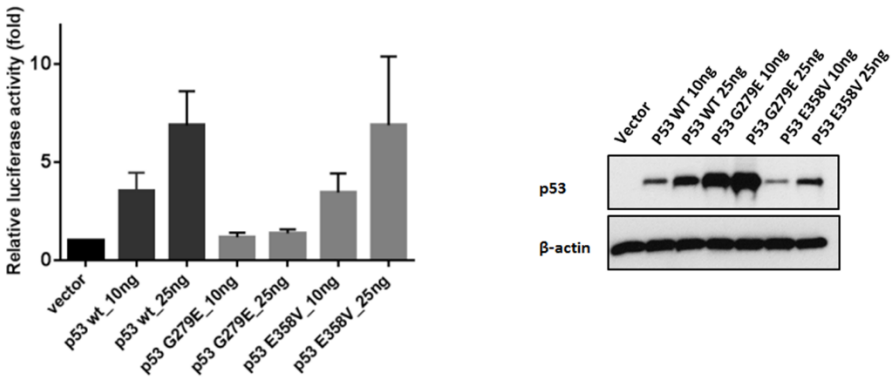


Figure 3.5 Transcriptional activity of NTD patient-derived p53 mutants on the *BAX* promoter. Left: relative luciferase activity on the BAX promoter; right: protein levels of the transfected mutants measured by Western blot.

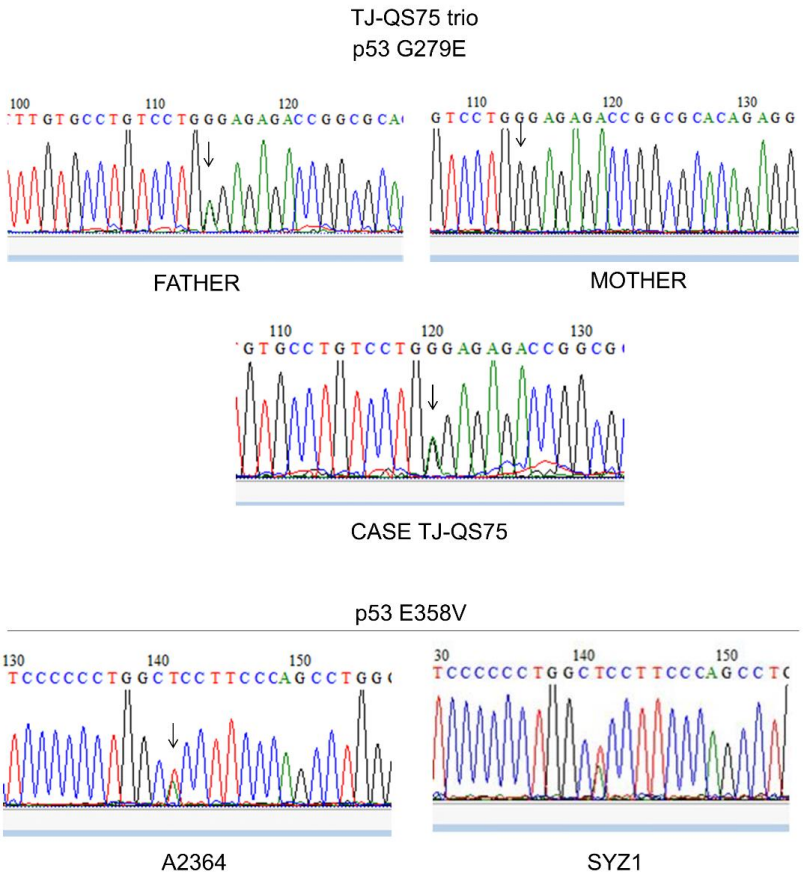


Figure 3.6 Sanger re-sequencing of case-specific nonsynonymous *TP53* variants

The second case-specific variant encoding p53 E358V was observed in two unrelated patients with spina bifida (**Table 3.1, Figure 3.6**). The variant is absent from all East Asian ExAC genomes as well as our internal matched controls, making it significantly enriched in this NTD cohort over the general population ($p = 4.1 \times 10^{-4}$, two-tailed Fisher's exact test). Moreover, the variant was absent in the proband's two unaffected siblings (**Figure 3.7**), although it was inherited from an unaffected father. The affected patients both had spina bifida, with the one also having encephalocele. Taken together, these data identify chr17: 7573954T>A as a candidate pathogenic variant in human spina bifida.

P53 E358 was able to transactivate the *BAX* promoter at wild-type p53 levels (**Figure 3.5**), suggesting its potential pathogenicity is independent of p53's transcriptional activity. Interestingly, the variant was also detected in the germline of an adult-onset sarcoma patient (patient identified in study first described in (355)). This raises the possibility that E358V is also a cancer-predisposing variant, causing the Li-Fraumeni syndrome. However, the mutant protein shows normal transcriptional activity on multiple target promoters (D. M. Thomas, *personal communication*). E358 is located within the tetramerisation domain of p53. Interestingly, loss of tetramerisation function through L344P mutation abolishes the oncogenic effects of p53 mutant D281G (356). It would be interesting to test whether E358V affects p53 tetramer formation. Regardless, more data are needed to assess its pathogenicity in cancer.

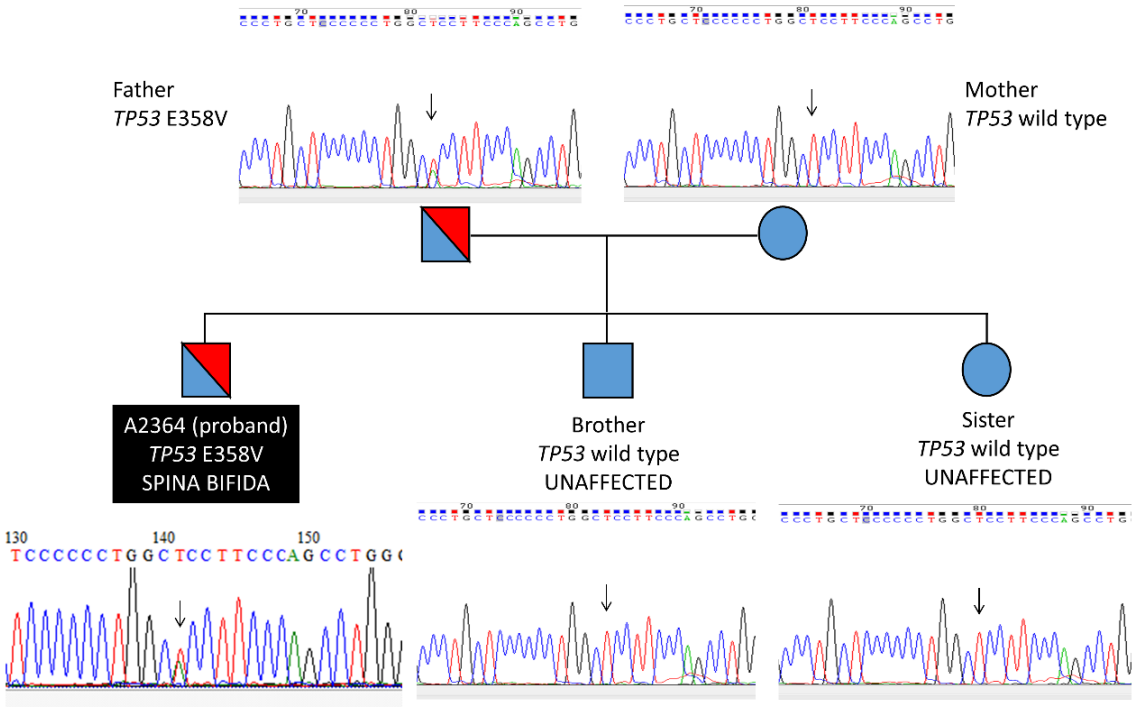


Figure 3.7 Pedigree of p53 E358V carrier. The variant segregates with the disease among the siblings, although the parent carrier is unaffected.

The observation of a recurrent rare missense variant specific to NTD cases prompted us to investigate which recurrent NTD-specific variants in other genes have been identified so far. **Table 3.2** shows NTD-specific nonsynonymous rare variants observed in at least 2 independent NTD cases, and their significance of enrichment compared with the ethnicity-matched ExAC cohort. When corrected for multiple hypothesis testing, only 4 variants fall within $p < 0.05$: VANGL1 S83L; TP53 E358V; CELSR1 N2309 and DVL1 T582P. Many of the remaining recurrent variants are in fact present at comparable frequencies in the general population. Thus, applying these criteria, chr17: 7573954T>A (p53 E358V) is a novel recurrent NTD-enriched variant which, unlike the other three significantly enriched variants, affects a gene not associated with planar cell polarity. It would be informative to sequence *TP53* in a larger cohort to validate the variant's genetic association with spina bifida.

Table 3.2 Recurrence statistics of NTD nonsynonymous variants in published literature and this study. EVS, exome variant server; ctrls, controls; ExAC, exome aggregation consortium; MHS, multiple hypothesis testing; ENF, European Non-Finnish; AFR, African; EUR, European; LAT, Latino. Variants identified in this study are shaded.

Variant	Frequency in NTD cohort	In-study ctrls	EVS	ExAC	ExAC Ethnic specific	p (Ethnic specific)	p (MHS corrected)	Ethnic odds ratio	Reference	Variant details
VANGL1 S83L	3/673	0/1187	3/6503	9/60702	4/33369	**2.57E-04**	**4.88E-03**	37.16	(357)	1:116206325 C / T (rs146695372)
TP53 E358V	2/337	0/222	0/6503	1/58104	0/4243	**4.05E-04**	**7.70E-03**	inf	this study	17:7573954 T / A
CELSR1 N2230S	2/473	0/639	0/6503	0/37424	0/20808	**4.97E-04**	**9.44E-03**	Inf	(358)	
DVL1 T582P	2/473	0/150	0/6503	1/51616	1/28285	**8.08E-04**	**1.53E-02**	119.01	(359)	1:1271791 T / G
FZD6 R405Q	2/473	0/639	96/6503	229/60702	9/33370 (ENF) AFR 209/5203	**9.94E-03**	1.89E-01	15.67	(80)	8:104337548 G / A (rs150760762)
MTR M495I	2/237	0/239	45/6503	88/60681	3/5777	**1.45E-02**	2.76E-01	16.22	(82)	1:237001869 G / A (rs140424701)
VANGL1 T251M	2/673	0/1348	0/6503	11/60517	10/33267	**2.28E-02**	4.34E-01	9.88	(357)	1:116206829 C / T

MTHFD1 P328L	2/238	0/235	15/6503	84/60696	76/33363	1.07E-01	2.03E+00	3.69	(82)	14:64892766 C / T (rs147367593)
MTHFR R519C	2/233	0/229	33/6503	91/60436	13/5761	0.1145	2.18E+00	3.80	(82)	1:11852412 G / A (rs45496998)
PARD3 G1249S	3/337	0/222	0/6503	16/60645	15/4327	1.38E-01	2.62E+00	2.57	this study	10:34400423 C / T
GGH S121G	2/232	0/231	0/6503	16/60114	15/5699	1.43E-01	2.71E+00	3.27	(82)	8:63938855 T / C
MTHFD1 V345I	2/234	0/235	9/6503	20/60704	18/5203 (AFR)	2.13E-01	4.05E+00	2.47	(82)	14:64892816 G / A (rs143413657)
MTR R52Q	2/213	0/206	68/6503	498/60692	31/5778	0.3299	6.27E+00	1.75	(82)	1:236966848 G / A (rs12749581)
ALDH1L1 G23D	4/232	0/236	56/6503	543/60689	37/5788 (LAT); 441/33351 (EUR)	0.6623	1.26E+01	1.35	(82)	3:125879755 C / T (rs143122118)
PARD3 T861S	3/337	0/222	0/6503	31/60548	30/4322	7.30E-01	1.39E+01	1.28	this study	10:34625160 T / A
AHCY I92V	2/232	0/223	43/6503	444/58293	346/32037	1	1.90E+01	0.80	(82)	20:32881908 T / C (rs11552695)
GART P641A	2/239	0/239	113/6503	325/60693	295/5202 (AFR); 22/5789 (LAT)	0.2475	4.70E+00	2.20	(82)	21:34889697 G / C (rs34588874)

3.4 Genetic disruption of Par3-aPKC binding in human neural tube defects

ASPP2 regulates apicobasal polarity of neuroepithelial cells via Par3. Polymorphisms in the *PARD3* gene which encodes Par3 have been previously associated with NTDs in humans (128). However, it is unknown whether rare variants in *PARD3* contribute to NTD risk. Therefore, we performed targeted sequencing of *PARD3* in the NTD cohort and matched controls, and validated rare nonsynonymous variants by Sanger re-sequencing. Six nonsynonymous rare (MAF < 1%) variants were identified that were specific to NTD cases. Nine more variants were identified that were specific to controls (n = 5) or observed in both NTD cases and controls (n = 4) (**Figures 3.8, 3.9 and Table 3.3**).

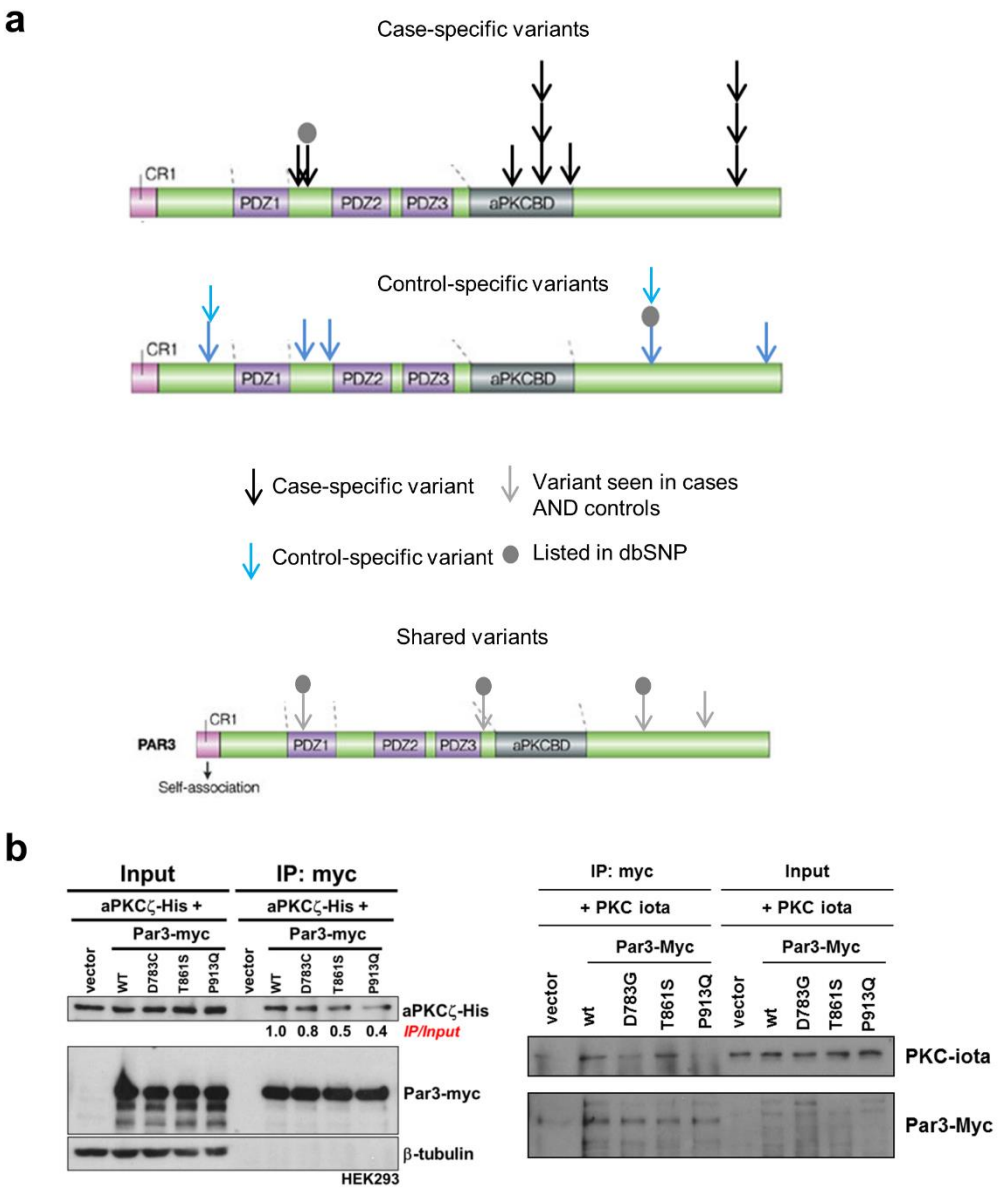


Figure 3.8 NTD-specific missense variants are enriched in the aPKC-binding domain of Par3 and disrupt Par3's binding to aPKC. (a) Position of Par3 residues mutated in NTD cases (top, black arrows) vs matched controls (below, blue arrows), and variants detected in both cases and controls (bottom, grey arrows). (b) Binding of wild-type Par3 vs three aPKC-BD mutants to aPKCζ (left) and aPKCι (right). Hek293 cells were transfected with the aPKC and Par3 plasmids and Par3 was immunoprecipitated using anti-Myc tag antibody (left). Right: Par3 and aPKCι proteins were expressed using in vitro translation and subjected to the immunoprecipitation assay as per aPKCξ. Domain map of Par3 adapted from Macara et al. (360) with permission.

Interestingly, although the overall burden of nonsynonymous variants was comparable between cases and controls, the missense variants showed a striking

difference in distribution along the Par3 sequence (**Figure 3.8**). Three of the observed case-specific missense variants, one of which was observed in three unrelated cases, are located in the aPKC-binding domain (aPKCBD) of Par3. Two of these are novel: D783G and P913Q. No missense variants in aPKCBD of Par3 were observed in controls or shared between cases and controls. These data demonstrate there is an enrichment of Par3-aPKCBD missense variants in NTDs over controls.

Table 3.3 Rare nonsynonymous variants in *PARD3* identified in neural tube defect patients. Freq, frequency.

SNV (genomic)	SNV (protein)	SNP ID	Sanger validated?	Freq cases (null:het:hom)	Allele count (cases)	Freq controls (null:het:hom)	Allele count (ExAC-all)	Allele count (ExAC-East Asian)	Variant type
chr10:34671833C>T	R345H		yes	339|1|0	1/680	222|0|0	1/121390	0/8652	novel rare
chr10:34671821C>T	R349H	rs199923448	yes	339|1|0	1/680	222|0|0	8/121398	7/8652	rare
chr10:34630624T>C	D783G		yes	339|1|0	1/680	222|0|0	0/121282	0/8652	novel rare
chr10:34625160T>A	T861S		yes	337|3|0	3/680	222|0|0	31/121096	30/8644	rare
chr10:34620149G>T	P913Q		yes	340|1|0	1/682	223|0|0	0/121349	0/8646	novel rare
chr10:34400423C>T	G1249S		yes	336|3|0	3/678	222|0|0	16/121290	15/8654	rare

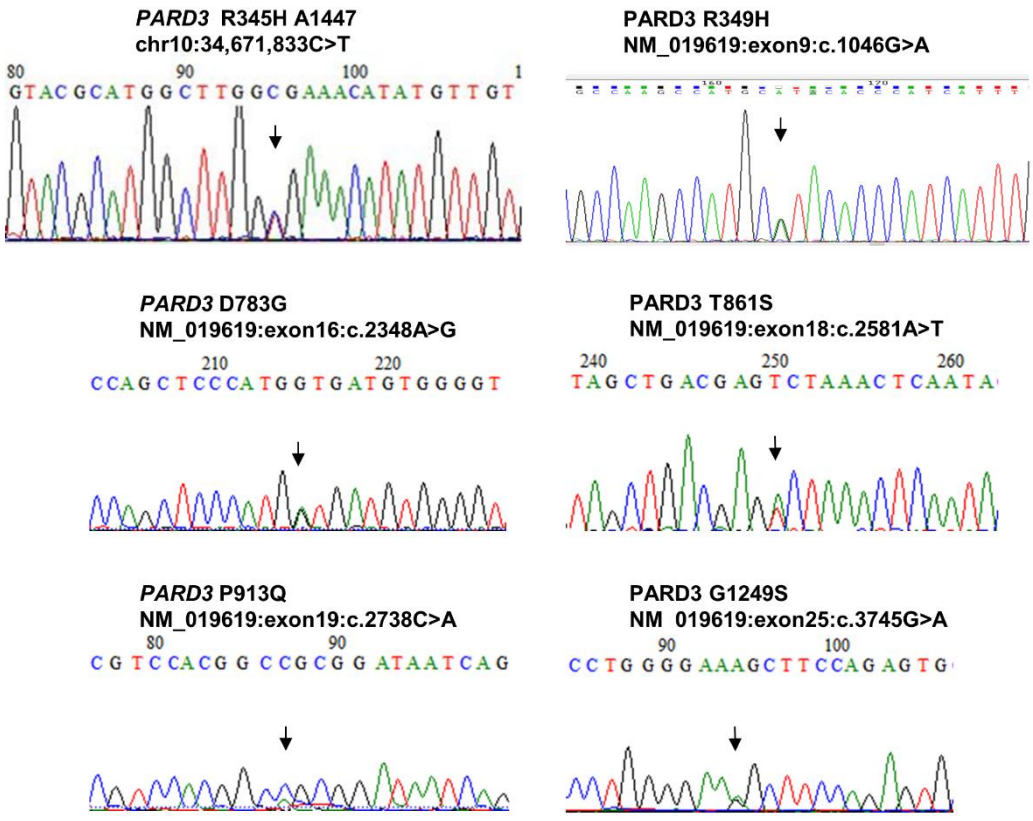


Figure 3.9 Sanger re-sequencing traces validating rare *PARD3* variants discovered in NTD cases by Illumina sequencing. Variants which failed validation are not shown.

Given the enrichment of aPKCBD variants among cases, we asked whether these mutations affected Par3's ability to bind aPKC. Two isoforms of aPKC exist in humans: aPKCι (iota) and aPKCζ (zeta). Immunoprecipitation of exogenously transfected Par3 mutants in Hek293 cells produced lower levels of aPKCξ bound to Par3 mutants than to the wild type. This effect was particularly severe for Par3 P913Q which bound less than 50% of the amount of aPKCξ bound by wild type Par3 (**Figure 3.8b**). Similarly, *in vitro*-translated Par3 mutants bound less PKCι than wild type, with the exception of T861S. Strikingly, Par3 P913Q pulled down virtually no PKCι in this assay (**Figure 3.8b**). These results show that aPKCBD mutants disrupt Par3's ability to bind aPKC, with P913Q being particularly deleterious.

Discussion

We performed targeted sequencing of 4 genes in a cohort of 340 NTD cases and 220 matched controls. We observed similar frequency of missense *TP53BP2* mutations in cases and controls, suggesting *TP53BP2* deletions, such as the deletion observed in our encephalocele case presented in Chapter 2, contribute more significantly to NTD risk than *TP53BP2* SNVs. However, a total of 12 *TP53* and *TP73* missense variants in cases versus one *TP73* missense variant in controls were found and validated in our assay. These variants represent an overall burden of missense variants in these two p53-family transcription factors in NTDs ($p = 0.0197$, Fisher's exact test). An increased sample size will be crucial in confirming the role of rare *TP53* and *TP73* variants in human NTDs.

We also observed an enrichment of missense variants in the aPKCBD of Par3 among NTD cases, whereas such variants were absent from controls. Experimentally, these mutations disrupted Par3's binding to both isoforms of aPKC, implicating the genetic disruption of the Par3-aPKC interaction in the aetiology of NTDs. In epithelial cells, Par3 is known to regulate apical constriction through aPKC-mediated phosphorylation of ROCK (124). Apical constriction is an important process during neural tube closure, necessary for the folding of the neuroepithelial sheet that gives rise to the neural tube. It is possible that rare variants in aPKCBD weaken Par3's ability to mediate apical constriction. These heterozygous rare variants might be acting as hypomorphic alleles, reducing the amount of Par3 protein in the cell capable of binding aPKC. Given the association of polarity gene CNVs with NTDs (81), and the strong evidence of pathogenicity of *de novo* truncating variants in *SHROOM3 (119)*, it is likely that a polygenic burden

of variants disrupting apicobasal polarity and/or the process of apical constriction contributes to NTD risk. Thus, monoallelic disruption of Par3-aPKC binding may predispose foetuses to NTDs by reducing the effective dosage of Par3 during apical constriction.

Table 3.4 Summary of disease-associated variants identified. Question mark indicates the uncertainty surrounding the Par-aPKC binding assay as a well-established functional assay.

Variant	Phenotype	*De novo* / inherited	ACMG evidence codes	ACMG classification	Notes
TP53 G279E	Spina bifida	Inherited	PS3, PM2, PP3	Likely pathogenic	
TP53 E358V	Spina bifida	Inherited / n/a	PS4, PM2	Likely pathogenic	Co-segregation with disease; 2 unrelated cases
TP73 V190F	Craniorachischisis	N/A	PS3, PM2, PP3	Likely pathogenic	
PARD3 D783G	Craniorachischisis	N/A	PM2, PS3?	Likely pathogenic (?)	Enrichment of aPKCBD variants in cases
PARD3 P913Q	Craniorachischisis	N/A	PM2, PP3, PS3?	Likely pathogenic (?)	Enrichment of aPKCBD variants in cases

Our functional analysis of NTD-associated p53 and p73 mutations also identified loss-of-function alleles, p53 G279E and p73 V190F, that inactivate the proteins' transcriptional activity. Recently, a significant genetic overlap was found between genes targeted by *de novo* mutations in developmental disorders and driver genes in cancer (361, 362). Importantly, the functional impact of mutations identified in these genes was not always concordant, particularly for tumour suppressor genes. Therefore, it is possible that a subset of *TP53* mutations increase susceptibility to NTDs. The notion that these mutations might be acting differently to somatic cancer mutations and cancer susceptibility mutations seen in the Li-Fraumeni syndrome is supported by the fact that the recurrent variant we observed as significantly enriched in NTDs (p53 E358V) does not affect p53's

transcriptional function. It would be illuminating to investigate whether increased risk of NTDs exists among Li-Fraumeni patients. Interestingly, truncating germline variants have a high haploinsufficiency score in ExAC (pLI = 0.91), suggesting they are under negative selection which could have a developmental as well as cancer susceptibility component.

Among the individual variants identified, *TP53* variants G279E, E358, and *TP73* variant V190F can be classified as Likely pathogenic following the ACMG guidelines (**Table 3.4**) (ref. (65)). *PARD3* variants P913Q and D783G would qualify for this category if the functional test of Par3-aPKC binding is considered a well-established assay. Given the subjective nature of this statement, we mark the associated evidence types with a question mark (**Table 3.4**).

Altogether, our results support a role for a polygenic burden of rare missense variants in the aetiology of NTDs. Deleterious variants in multiple genes, such as *TP73* V190F or *PARD3* P913Q, are likely to have additive effects on NTD risk. WES or WGS studies of larger NTD cohorts will be instrumental in order to formally confirm the association of *TP53*, *TP73* and *PARD3* rare variants with NTD risk.

Materials and methods

Patient recruitment

Human NTD samples and associated controls were collected following an approval by the Committee of Medical Ethic in the Capital Institute of Pediatrics (Beijing, China) and the study was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. Human samples were obtained from the Northern China, a region with a NTD prevalence of 199.38/10,000 based on the local epidemiologic surveillance data collected during January 2002 and December 2004 (109). The Committee of Medical Ethic in the Capital Institute of Paediatrics (Beijing, China) approved and the study was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The enrolled pregnant women were diagnosed by trained local clinicians using ultrasonography, and then registered in a database. As in our previous study (363), medical abortions were carried out by obstetricians to terminate NTD-affected pregnancies. Pathologic diagnosis of NTDs was performed by an experienced pathologist according to the International Classification of Disease, Tenth Revision. Written informed consent was obtained from the parents on behalf of foetus. The normal embryos aborted for non-medical reasons from the same region were used as geography-matched controls. Routine prenatal check-ups, questionnaire interviews and autopsies were completed for controls, and any embryos that harboured any pathological malformation or intrauterine growth retardation were excluded from the control group. Samples were first isolated, subdivided and immediately stored at -80℃. During transportation, the tissues were kept on dry ice.

DNA isolation, targeted gene sequencing and variant calling

Genomic DNA was extracted from muscle, skin, brain or lung. Twenty-five milligrams of tissue were used for DNA extraction with the DNeasy Blood and Tissue kit (69504, Qiagen, Germany) and stored at -80℃. The concentration and purity of DNA samples were determined by measuring their absorbance at 260 and 280 nm. Genomic DNA was sheared using Covaris™ Adaptive Focused Acoustics system (Covaris, MA). Samples were prepared for sequencing using the Truseq DNA Sample preparation Kit (Illumina, Inc, San Diego, CA) following the manufacturer's standard procedure. Custom capture oligos were designed using SureDesign website (https://earray.chem.agilent.com/suredesign/, Agilent technologies). The genes of interest were captured using the SureSelect Target Enrichment kit (Agilent Technologies, USA) with the following details. The hybridization reactions were carried out on an AB2720 Thermal Cycler (Life Technologies Corporation, USA) with following hybridization conditions: Incubated the hybridization mixture for 16 or 24 hours at 65°C with a heated lid at 105°C. After the hybridization reactions, the genes of interest were captured with magnetic beads (Invitrogen, USA). The capture production was enriched with following cycling conditions: 98ºC for 30 s; 10 cycles of 98ºC for 10 s, 60ºC for 30 s, 72 ºC for 30 s; 72ºC for 5 min. Several libraries were pooled and subjected to bridge amplification on cBot (Illumina, Inc, San Diego, CA) following the manufacturer's standard cluster generation protocols. Following hybridization of sequencing primer, base incorporation was carried out on MiSeq Benchtop Sequencer (Illumina, Inc, San Diego, CA) following the manufacturer's standard sequencing protocols. The mean depth of coverage was 254.34x, and on average,

86.73% of targeted bases had sufficient coverage and quality for variant calling. The average read length was 251 bp. Sequence reads were aligned to the hg19 iteration of the reference human genome using BWA, version 0.5.9 (364). Base score recalibration and local realignment for indel (insertion or deletion) detection and duplicate removal were performed with GATK (365). SNVs were called using GATK (Samtools pileup (MapQ>=30,reads>=2) (version 1.1) (366) and Varscan (version 2.3) (367), short indels (insertions and deletion) were calling using Varscan with loose standard (min-coverage =1, min alterative allele reads=1, min-var-freq > 0.03). Genotypes were called using Bayes estimation and Binomial test. Variants were annotated with ANNOVAR (368) and searched in databases inluding dbSNP, 1000 genomes, Exome Variant Server, Exome Aggregation Consortium. Variants with a minor allele frequency of >= 0.01 in dbSNP and/or ExAC were termed common variants, others rare variants. Rare variants absent in all databases were termed novel rare variants. All missense rare variants were validated by Sanger sequencing with custom primers (**Table 3.5**).

NB: Protein coding variants in *TP53BP2* were referenced to the Uniprot accession Q13625 which belongs to the full-length 1128-residue ASPP2 protein. However, the corresponding RefSeq transcript NM_001031685 encodes a 1134-residue protein and there is no reference RefSeq transcript with a correctly annotated start codon. Therefore the residue numbers given in this manuscript are referenced to the Uniprot sequence while the nucleotide numbers are referenced to the Refseq transcript.

Table 3.5: Primers for Sanger re-sequencing.

Gene symbol	Variant	Nucleotide position	Forward primer	Reverse primer

TP53BP2	K102E	chr1:223998183	TGTTAGCGAGTACAAGG	ATTCACCAGGAACTTTT
TP53BP2	I693M, S781L	chr1:223984144	TCCCTGCTTTGTCATTTA	GGTTCTTCTGGGTTATTCT
TP53BP2	E948A	chr1:223981006	CCTATTTGCCACTTCTGATT	CCTGGATGACAGAGCGAGAC

Plasmids and cloning

Single nucleotide variants were cloned by site-directed mutagenesis using whole-plasmid amplification with partially overlapping primers, amplification by Phusion Hot Start Polymerase II (Thermo Fisher Scientific, Waltham, MA), followed by digestion by DpnI (New England Biolabs, Ipswich, MA) for 3 h. The digested DNA was used to transform competent cells (Bioline, Taunton, MA) and the resulting colonies screened for the introduced mutation by Sanger sequencing. Plasmid DNA for mutant screening was prepared using the Plasmid miniprep kit (Qiagen, Hilden, Germany); plasmid DNA for transfections was prepared using the Plasmid Maxiprep kit (Sigma Aldrich, St Louis, MO). Luciferase plasmids, p53, ASPP2 expression plasmids were obtained as described previously (103, 107).

Luciferase assays

Twenty-four hours after plating, cells were transfected with the following plasmid combinations for specific experiments. H1299 cells were lysed in Passive Lysis Buffer (PLB) 20 h after transfection. Saos-2 cells were washed 1 time with PBS and supplemented with fresh growth medium 20 h post-transfection and subsequently lysed with PLB 40 h after transfection. Cells containing PLB were stored at -80°C and thawed in order to achieve homogenous cell lysates. Luciferase assays were performed using the Dual-luciferase Reporter Assay System (Promega, Madison, WI) according to the manufacturer's instruction. The Luminescence module of the GloMax®-Multi+ Microplate with Instinct® plate

reader was used with 5-second integration. In this assay, the activities of firefly (*Photinus pyralis*) and Renilla (*Renilla reniformis* or sea pansy) luciferases are measured sequentially from a single sample. Renilla was used as an internal control for transfection efficiency. For each transfection, empty vector (pcDNA3) was used to control the total amount of DNA transfected. Firefly luminescence signal was normalized based on the *Renilla* luminescence signal. Mutant vs wild-type fold change values at each plasmid concentration were analysed using a two-tailed parametric t-test and corrected for multiple hypothesis testing by multiplying the p-value by the number of tests. Samples with a corrected $p < 0.05$ were deemed significant.

Western blotting

To quantify protein levels expressed in luciferase assays, twenty-four hours (H1299) or forty-eight hours (Saos-2) following transfection, cells were washed with PBS and lysed with Passive Lysis Buffer (PLB) (Luciferase buffer, Promega). Extracts were centrifuged at 16000 g for 20 minutes at 4^{o}C. Protein concentrations were measured by Bradford assay (Bio-Rad). Protein sample was mixed with sample buffer before being loaded on an SDS-acrylamide-gel (8% manually carted Bis-Tris gel). Prestained protein marker (New England Biolabs) was used to determine the size of the protein. The SDS-polyacrylamide gel electrophoresis was initially run at 80V (250 mM Tris-base, 250 mM MOPS, 5mM EDTA, 0.5 % SDS and 5 mM Sodium Bisulfite) before being increased to 150 V after the visualisation of the marker bands. Subsequently, the proteins were transferred to nitrocellulose membranes and blocked with 5% milk in TBS containing 0.1% tween-20 (TBST) for 1 hour before being incubated with specific primary antibody overnight at 4°C. The following day, membrane was washed three times for 10

minutes with TBST before being incubated for 1hr at room temperature with horseradish peroxidase-conjugated secondary antibody (Dako, Agilent, Santa Clara, CA, USA). For visualisation, membrane was washed with TBST and then developed using the ECL western blotting system (GE Healthcare, Chicago, IL, USA) on an autoradiography film.

Contributions

The following individuals directly contributed to the work presented in this chapter: Jaroslav Zak, Huili Li, Isaraphorn Pratumchai, Yihua Wang, Stefanie Werner, Ting Zhang and Xin Lu. J. Z. and X. L. designed study; H. L. and T. Z. designed the clinical study; J. Z. performed most experiments and analysed data; T. Z. and H. L. sequenced and re-sequenced patient samples and provided clinical data; I. P. produced Western blots in Figures 3.1c, 3.4 and 3.5, and contributed to luciferase assay readings in Figures 3.1c, 3.4 and 3.5 ; Y. W. produced Western blot in Figure 3.8b; S. W. contributed to luciferase assay readings in Figures 3.4 and 3.5.

CHAPTER 4: IDENTIFYING DRIVER GENES IN HUMAN CANCER

RESULTS AND DISCUSSION

In this chapter, two fundamental questions are explored, using the ASPP genes as an example:

1) Using publicly available data, how can we assess the likelihood that a given gene plays a role in human cancer, and if it does, with pro- or anti-tumour effects?

   Results section 4.1 uses an unbiased approach to assessing a gene's role in cancer and its classification as a pro- or anti-tumorigenic gene.

2) How can we utilise genomic data integration to identify rare candidate driver mutations in cancer and their molecular effects?

   Results section 4.2 described analyses of genetic interactions between mutations, and associations between mutations and mRNA levels to identify candidate driver mutations and determine their mechanism of action.

Results section 4.1: An unbiased approach to assessing the role of individual genes in human cancer

Table 4.1 Genomic data sources and types of evidence available for an unbiased analysis of the role of individual genes in cancer. List up to date as of 10 September 2016. The column 'causal on its own' determines whether the type of evidence available from the specified data source can lead to an established causal role for a given gene without referring to other analyses or data sources. NT, normal tissue; TP, primary tumour.

Data source	URL/API	Molecular level	Type of data	Species	Level	Causal evidence
Candidate Cancer Gene Database	ccgd-starrlab.oit.umn.edu/about.php	Genomic (insertions)	Hits from forward genetic screen	Mouse	*In vivo* model	Yes
cBio	CGDS R package	Somatic mutation	Somatic mutations aggregated from studies	Human	Patient	No
COSMIC	cancer.sanger.ac.uk/	Somatic mutation	Somatic mutation list	Human	Patient	No
ICGC	Multiple options e.g. Python API	Somatic mutation	Somatic mutation list	Human	Patient	No
TCGA	Firebrowse API	Somatic mutation	Somatic mutation list	Human	Patient	No
TCGA	Firebrowse API	Somatic mutation	MutSig2CV	Human	Patient	Potentially
TCGA	Firebrowse API	Somatic CNV	GISTIC2	Human	Patient	Potentially
Oncomine	GUI	mRNA	Microarrays TP vs NT	Human	Patient	No
TCGA	Firebrowse API	mRNA	RNAseq TPvsNT	Human	Patient	No
TCGA	Firebrowse API	Clinical correlation	mRNA vs clinical	Human	Patient	No
GWAS studies	www.ebi.ac.uk/gwas/	Germline SNPs	GWAS	Human	Patient	Potentially
COSMIC		Somatic mutation	Somatic mutation list	Human	Cell line	No
ACHILLES sgRNA screen	GUI	Genomic (insertions)	Cell line dependence	Human	Cell line	Yes
CTRP	GUI	mRNA	Cell line sensitivity	Human	Cell line	No

4.1.1 Analysing the composition of classes of somatic mutations

Tumour suppressor genes are frequently inactivated by truncating mutations (e.g. *PTEN*; *NF1*) whereas oncogenes tend to show missense driver mutations (e.g. *KRAS*; *IDH1*). Vogelstein et al. proposed a ratiometric method to identify which cancer driver genes targeted by driver mutations in cancer (Mut-drivers) act as oncogenes and which as tumour suppressor genes (151). The 20/20 rule requires at least 20% of the recorded mutations to be at recurrent positions and missense for oncogene classification, or >20% of the recorded mutations to be inactivating for TSG classification (151). To evaluate whether the ASPP family genes – *PPP1R13B*, *TP53BP2* and *PPP1R13L* – are more likely to be oncogenes or TSGs, nonsynonymous mutations in these genes were collected and manually curated to avoid duplicates, mis-annotated mutations, erroneous reference transcript selection and other caveats present in public mutation databases.

A snapshot of the manual database corresponding to v78 of COSMIC, v1.2.4 of cBioPortal, r23 of the ICGC Data Portal, data release 2.0 of the GDC plus legacy mutations from earlier releases, there were a total of 835 nonsynonymous somatic mutations in the ASPP genes. *PPP1R13B* and *TP53BP2* showed a comparable percentage of truncating mutations: 20% and 19% respectively (**Figure 4.1)** This proportion of truncating mutations was significantly larger than for *PPP1R13L* (**Figure 4.1**).

To rule out biases from including legacy mutations which might have been removed for quality control reasons, the analysis was repeated solely with mutations indexed by the four databases mentioned above at their synchronised

data freeze points. We achieved the same results, and also note that >90% of the manually curated mutations are contained in at least one of the current releases of the public databases. These results show that based on their mutation pattern, *PPP1R13B* and *TP53BP2* are more likely to be TSGs than oncogenes, and that they fall on the borderline of qualification as Mut-driver tumour suppressors using the ratiometric method by Vogelstein et al.

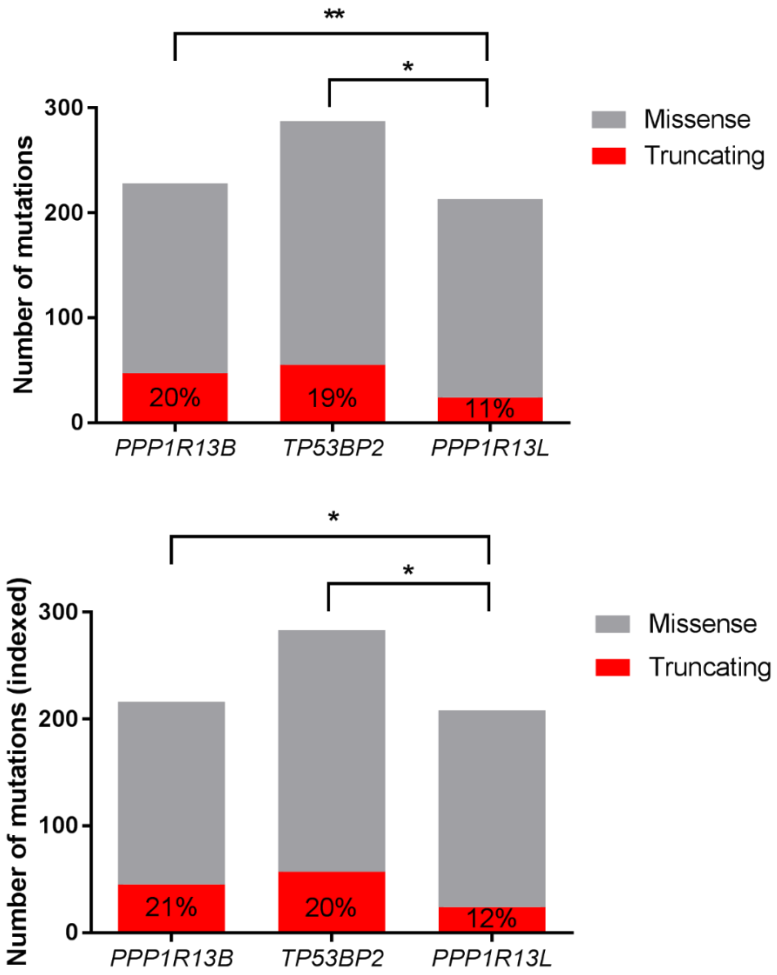


Figure 4.1 *PPP1R13B* (ASPP1) and *TP53BP2* (ASPP2) show significantly higher proportion of truncating somatic mutations among nonsilent mutations compared to *PPP1R13L*. Data based on snapshot of COSMIC v78, cBio v1.3.4 and ICGC r22 plus legacy mutations (top) and without legacy mutations (bottom). Fisher's two-tailed exact test, * $p < 0.05$; ** $p < 0.01$.

To test for the association of nonsilent ASPP mutations with *TP53* status, we compared the frequency of mutations in each ASPP gene in wild-type- versus mutant-*TP53* tumours (**Figure 4.2**). Mutations in *TP53BP2* were significantly more enriched in wild-type-*TP53* tumours compared to *PPP1R13L* mutations. This suggests *TP53BP2* mutations might be genetically interacting with *TP53* mutations.

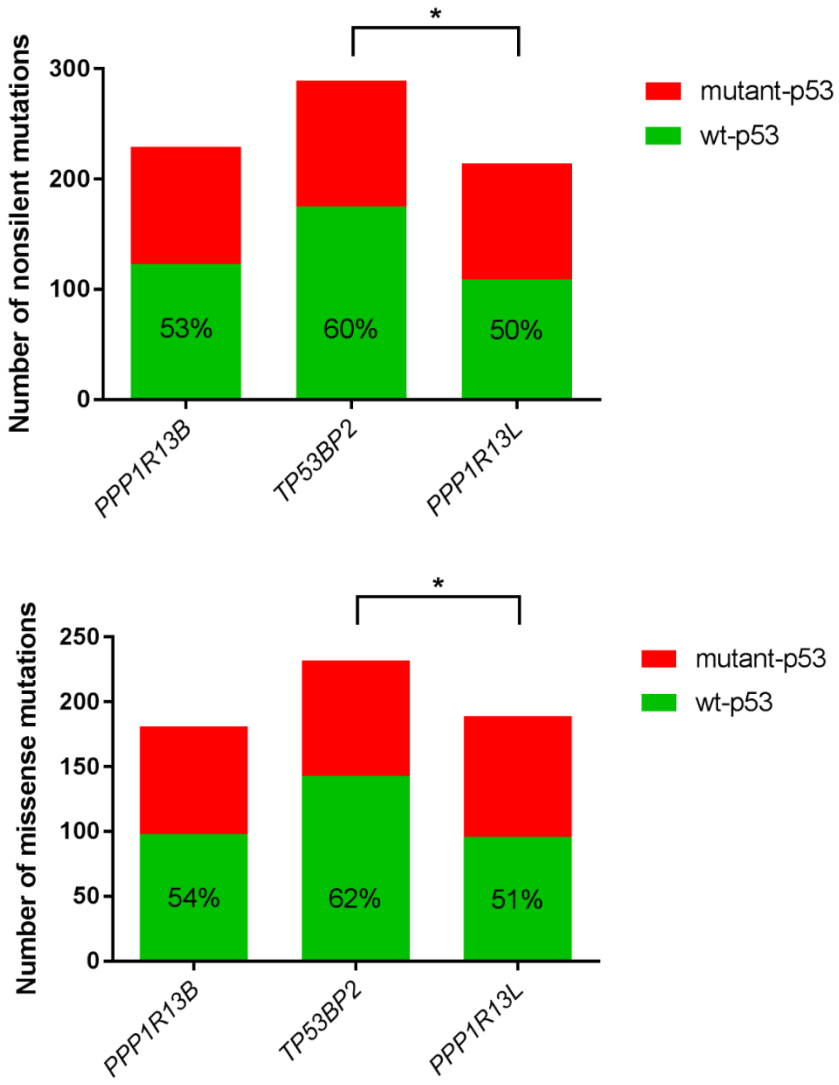


Figure 4.2 Frequency of nonsilent (top) and missense (bottom) ASPP mutations in tumours harbouring wild-type vs mutant *TP53*. The source mutation set is identical to that used in Figure 3.5 top panel. * $p < 0.05$, Fisher's exact two-tailed test; wt, wild type.

4.1.2 Analysing the frequency of nonsilent/silent mutations

The vast majority of recognised driver mutations are nonsynonymous. Therefore, one can expect to be able to identify driver genes simply by ranking genes by the ratio of nonsilent/silent mutations. To test this hypothesis, nonsilent/silent ratios across 31 TCGA studies available through the Firebrowse API were collected for 18,310 genes and the distribution of ratios plotted (**Figure 4.3**). Ratios for *PPP1R13B*, *TP53BP2* and *PPP1R13L* were shown as blue, orange and grey lines, respectively. As a positive control, distribution of ratios for genes from the Sanger Cancer Gene Census was plotted (**Figure 4.3**, red line). Although some genes from the Cancer Gene Census showed high nonsilent/silent ratios (notably a local peak around 9), the overall distribution varies very little from the distribution of ratios for all 18,310 genes. The overlap of the distributions shows that the majority of Cancer Census genes have nonsilent/silent ratios that fall within the bulk of the genome-wide distribution. Notably, the nonsilent/silent ratios of *PPP1R13B* and *TP53BP2* are higher than the median of the Cancer Gene Census distribution so the ratios are consistent with a Mut-driver role for the genes. Altogether, these data demonstrate that while some recognised cancer genes do show high enrichment for nonsilent over silent somatic mutations, the ratio alone does not have sufficient sensitivity to detect cancer driver genes.

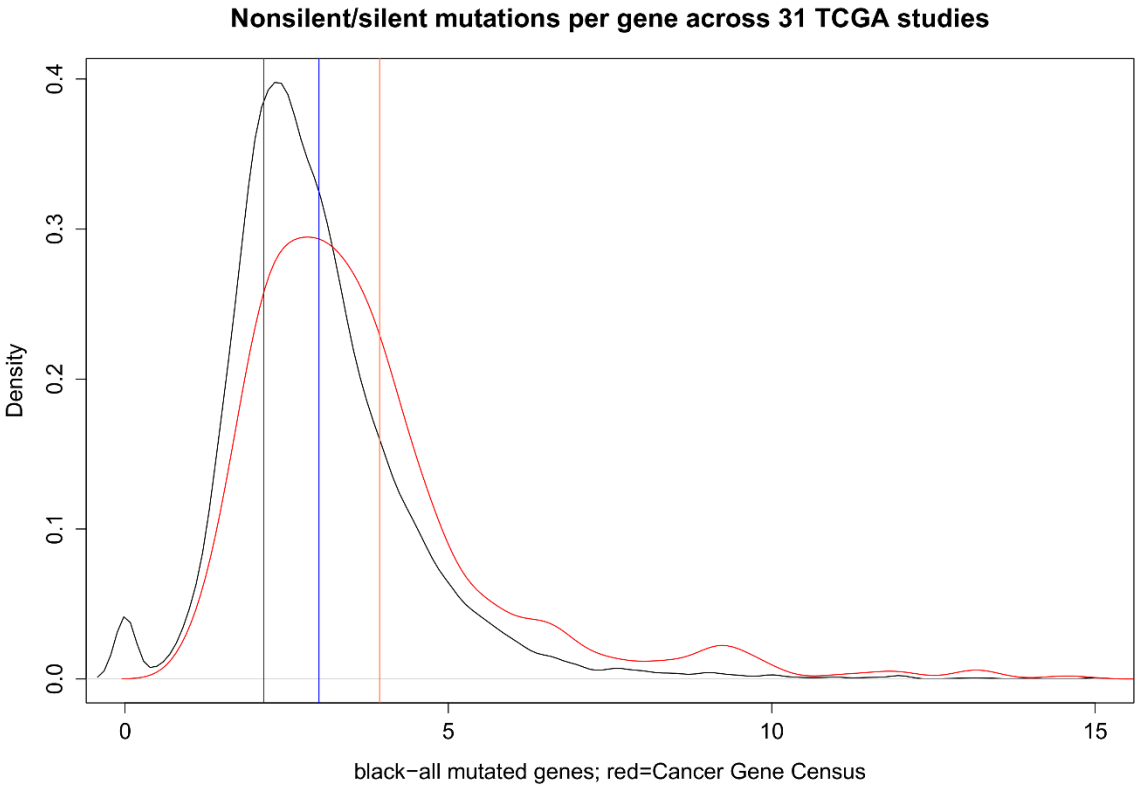


Figure 4.3 Density plot of nonsilent/silent somatic mutations per gene collected from 31 TCGA studies: distributions of curated cancer genes vs all genes do not significantly differ. The formula used to compute the ratio is nnon/(nsil+1) to avoid infinite values. Black distribution depicts 18,310 genes with at least 1 somatic mutation detected; red distribution depicts genes from the Sanger Cancer Gene Census. Vertical lines show ratios for *ASPP1*, *ASPP2* and *iASPP* genes marked in blue, orange and dark grey, respectively. Firehose version used: 2016_01_28. Data retrieved via Firebrowse API.

4.1.3 Analysing the occurrence of loss of heterozygosity for truncating mutations

The classical tumour suppressor gene model classifies TSGs as recessive genes by requiring a 'second hit', i.e. biallelic inactivation, to initiate tumorigenesis. Genes conforming to this recessive TSG model are therefore expected to show loss of heterozygosity accompanying truncating variants whereas this phenomenon does not occur in oncogenes. We examined the copy number of *PPP1R13B*, *TP53BP2* and *PPP1R13L* in tumours carrying truncating mutations in the respective genes, for which copy number status was available. Four tumours

with *PPP1R13B* truncating mutations accompanied by loss of *PPP1R13B* heterozygosity (LoH) were found (**Table 4.2**). Three tumours with truncating *TP53BP2* mutations with *TP53BP2* LoH, but no tumours with the same concomitant alterations in *PPP1R13L*, were found. The observation of recurrent truncating mutation + LoH events supports the role as TSG for *PPP1R13B* and *TP53BP2*.

Table 4.2 Truncating mutations in *PPP1R13B* and *TP53BP2* with loss of heterozygosity. Copy number status retrieved from cBio v1.2.4.

Gene	Tumour type	Mutation	ASPP1 CN	Patient barcode
PPP1R13B	Liver	I322fs	ShallowDel	TCGA-ED-A97K
PPP1R13B	Liver	X955_splice	ShallowDel	TCGA-G3-A25S
PPP1R13B	Lung adenocarcinoma	R81.	ShallowDel	TCGA-55-1592
PPP1R13B	Lung squamous	E64.	ShallowDel	TCGA-63-5131
TP53BP2	Bladder	S769.	ShallowDel	TCGA-DK-A3X1
TP53BP2	Prostate	K513fs	ShallowDel	TCGA-M7-A724
TP53BP2	Lung squamous	P898fs	ShallowDel	TCGA-60-2711

Gene	No. of truncating mutations with LoH	Tissue of origin
PPP1R13B	4	Lung, liver
TP53BP2	3	Lung, bladder, prostate
PPP1R13L	0	-

4.1.4 Mining mouse tumorigenesis transposon screens

An important problem in cancer genomics is distinguishing causal driver genes from genes whose dysregulation merely reflects the molecular landscape

shaped by the driver mutations. For example, many proteins are under-expressed in tumour tissue over normal tissue but only some of them actively inhibit tumorigenesis. Therefore, experimental models are needed to connect discrete genetic events with direct effect on tumour growth. Retrotransposon insertion screens in mice are one such model. In a mouse with a cancer-predisposing mutation transposon insertions at genic sites can initiate and/or promote tumour growth. Over-represented sites of insertion reliably identify known tumour suppressor genes in the relevant tissue type.

We mined the Candidate Cancer Gene Database (http://ccgd-starrlab.oit.umn.edu/) for screens in which insertions at any of the *ASPP* genes were over-presented. *Ppp1r13b* was identified as a hit in 7 independent mouse mutagenesis screens (**Table 4.3**). Interestingly, 3/7 screens were specific to colorectal cancer. This suggests *Ppp1r13b* is a tumour suppressor in colorectal cancer and potentially other tissues.

Table 4.3. **Hits from mouse transposon screens**. Retrieved from papers indexed through CCGD, last updated 5 Sept 2016.

Gene	Cancer type	Carcinogenesis induction	*Aspp1* insertions / tumours	Rank	Reference
Ppp1r13b	Colorectal	AhCre *Apc*$^{flox/flox}$	40	329	(369)
Ppp1r13b	Colorectal	VillinCre T2/Onc	6	61	(370)
Ppp1r13b	Colorectal	*Smad4*$^{-/-}$ and *Kras*^{G12D}:SB +*Smad4*^{KO}:SB +*p53*^{R172H}:SB			(371)
Ppp1r13b	Gastric	*Smad4*$^{+/-}$	9	507	(372)
Ppp1r13b	Liver	HBV+	28	1394	(373)
Ppp1r13b	Peripheral nerve sheath	*Trp53*^{R270H} and/or *Cnp-EGFR*	16	611	(374)

Ppp1r13b	Pancreatic	*Kras*^{G12D}	20	1475	(375)

The observations of LoH alleles in human tumours suggest there is added benefit to the loss of both alleles in some tissues. However, this does not imply *PPP1R13B* and *TP53BP2* must be purely recessive tumour suppressors. As demonstrated in two independent studies in mice, single copy loss of *Trp53bp2* is sufficient to significantly increase the incidence of tumorigenesis. Therefore, it is possible that the genes are haploinsufficient tumour suppressors which have to be fully inactivated only under specific conditions, for example in specific tissues. This is consistent with the observation that colon cancer, which shows the strongest evidence of tumour suppressor status of ASPP1, shows truncating mutations of *PPP1R13B* without loss of heterozygosity. This suggests single-copy loss is sufficient to reduce tumour suppressive effects of ASPP1 in colon cancer.

Solimini et al demonstrated that hemizygous deletions that occur frequently in cancer are enriched in genes with anti-proliferative effects (160). They suggest these deletions help optimise the proliferative potential of cancer cells. Therefore, monoallelic deletions of *PPP1R13B* could be acting as mini-drivers.

4.1.5 Analysing mRNA expression in tumour vs normal tissue

Tumour suppressor genes are genetically and/or epigenetically inactivated in cancer whereas oncogenes are typically upregulated at the mRNA level. To analyse whether a gene is more likely a TSG or an oncogene, or shows no evidence of significant mRNA dysregulation, we analysed mRNA levels in tumours and associated normal tissue samples from TCGA and microarrays curated by

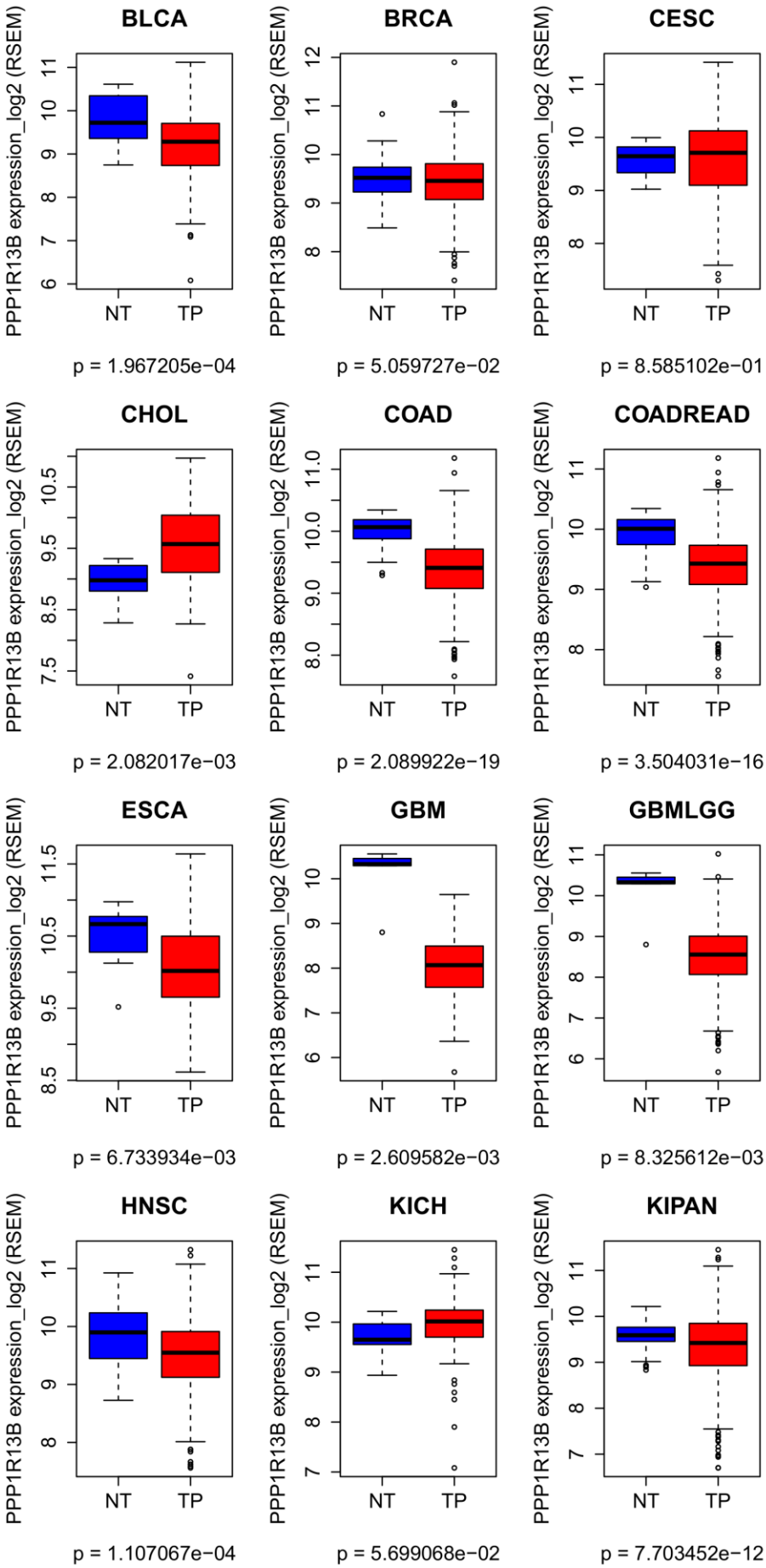
Oncomine. Only tumour types with normal tissue samples available were included in this analysis.

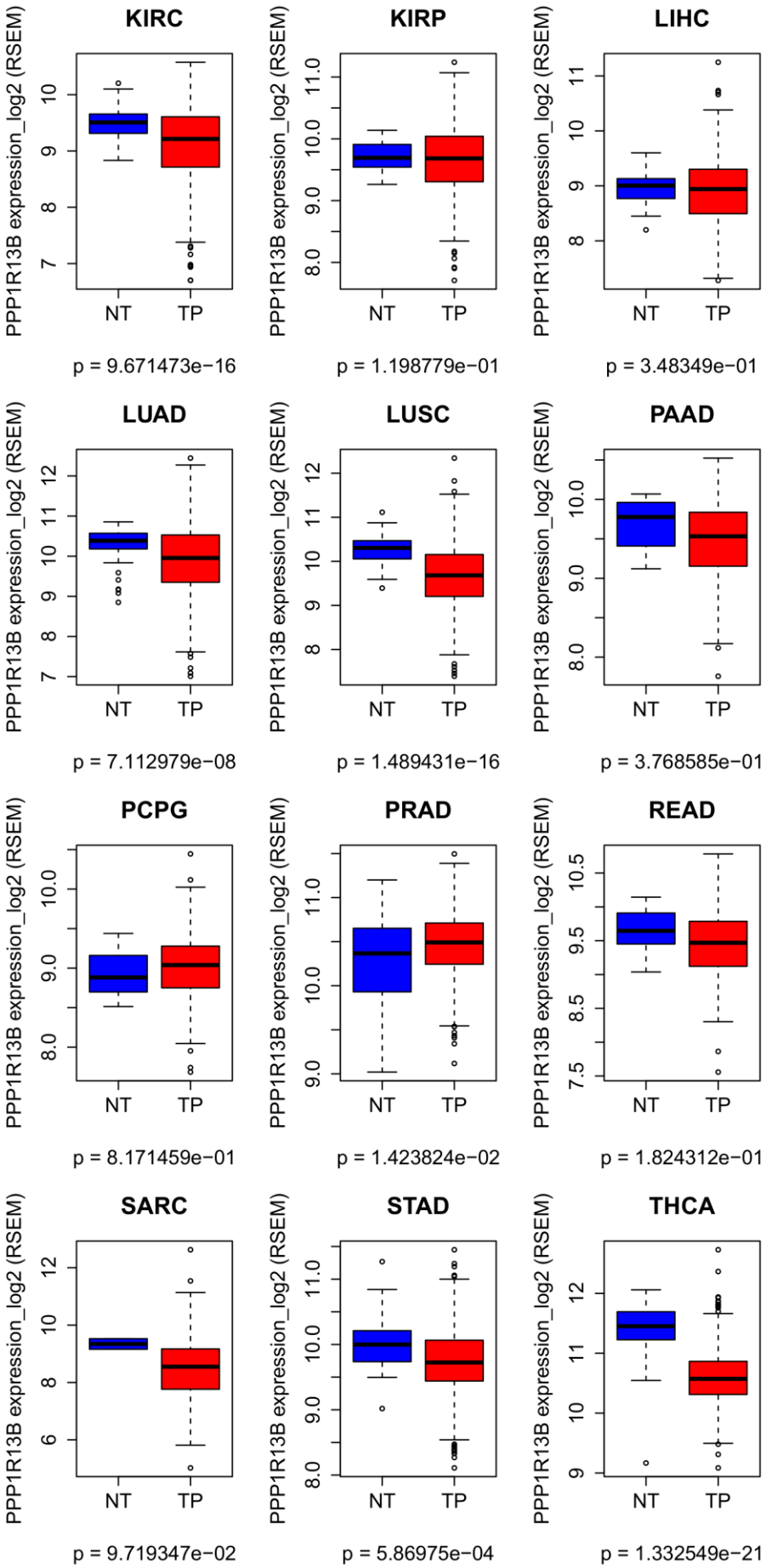
PPP1R13B was significantly differentially expressed in 10 TCGA tumour studies and down-regulated in all 10 (**Table 4.4, Figure 4.4**). Across Oncomine datasets, applying a p-value threshold for *PPP1R13B* differential expression of $p < 1x10^{-5}$ and an absolute fold-change > 1.5 yielded 19 tumour-versus-normal studies, all showing *PPP1R13B* downregulation in the tumours (**Table 4.5**). This striking concordance suggests *PPP1R13B* acts as a TSG across a broad number of tissues.

In contrast to *PPP1R13B*, *TP53BP2* shows evidence of both over- and under-expression in TCGA datasets: 7 studies with a significant difference (after applying a multiple testing correction), out of which *TP53BP2* was under-expressed in 5 (**Table 4.4, Appendix Figure S8**). Oncomine studies meeting the $p < 1x10^{-5}$ and $|FC| > 1.5$ thresholds comprise 7 datasets with under-expression of *TP53BP2,* and 7 with over-expression of the gene (**Table 4.5**). Therefore, mRNA expression alone is not conclusive with respect to *TP53BP2*'s global classification as a TSG or oncogene. It is possible that *TP53BP2* has both tumour suppressive and oncogenic effects whose relative dominance is lineage-dependent.

PPP1R13L shows a high degree of concordance among TCGA studies: significantly up-regulated in 11 studies and downregulation in a single study (**Table 4.4, Appendix Figure S9**); Oncomine studies show 12 vs 6 tumour studies with over- vs under-expressed *PPP1R13L*, applying the same criteria as above (**Table 4.5**). Based on the overall concordance of these studies, *PPP1R13L* mRNA levels in tumours correspond to an oncogene rather than a TSG.

Overall, published microarray and RNAseq datasets demonstrate uniform down-regulation of *PPP1R13B* in tumour vs normal tissue, mainly upregulation of *PPP1R13L* in tumours, and an ambiguous pattern for *TP53BP2* with both over- and under-expression in individual tumour types (**Table 4.6, Figure 4.5**).


BLCA
PPP1R13B expression_log2 (RSEM)
NT
TP
p = 1.967205e−04
BRCA
p = 5.059727e−02
CESC
p = 8.585102e−01
CHOL
p = 2.082017e−03
COAD
p = 2.089922e−19
COADREAD
p = 3.504031e−16
ESCA
p = 6.733934e−03
GBM
p = 2.609582e−03
GBMLGG
p = 8.325612e−03
HNSC
p = 1.107067e−04
KICH
p = 5.699068e−02
KIPAN
p = 7.703452e−12


KIRC
PPP1R13B expression_log2 (RSEM)
NT
TP
p = 9.671473e−16
KIRP
p = 1.198779e−01
LIHC
p = 3.48349e−01
LUAD
p = 7.112979e−08
LUSC
p = 1.489431e−16
PAAD
p = 3.768585e−01
PCPG
p = 8.171459e−01
PRAD
p = 1.423824e−02
READ
p = 1.824312e−01
SARC
p = 9.719347e−02
STAD
p = 5.86975e−04
THCA
p = 1.332549e−21

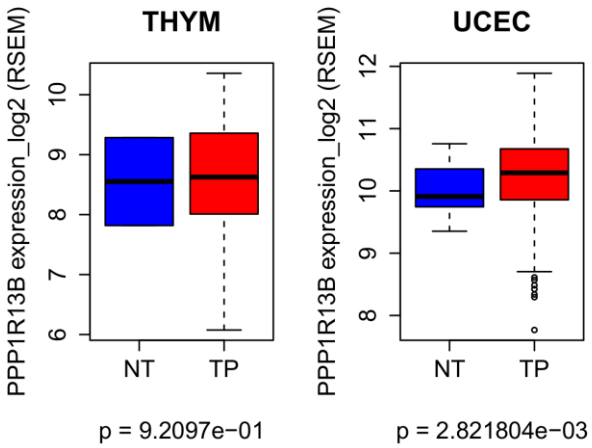


Figure 4.4 mRNA levels of *PPP1R13B* in TCGA primary tumours (TP) vs normal tissue (NT) across 23 tissue types and 3 aggregated tissue types. Student's two-tailed t-test; BLCA, bladder carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; COADREAD, colon and rectal adenocarcinoma; ESCA, oesophageal carcinoma; GBM, glioblastoma; GBMLGG, glioblastoma and lower-grade glioma; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe carcinoma; KIPAN, pan-kidney cancer; KIRC, kidney clear-cell carcinoma; KIRP, kidney papillary cell carcinoma; LIHC, liver carcinoma; LUSC, lung squamous cell carcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectal adenocarcinoma; SARC, sarcoma; STAD, stomach adenocarcinoma; THCA, thyroid carcinoma; THYM, thymoma; UCEC, endometrial carcinoma.

Table 4.4 mRNA levels in primary tumour vs matched normal tissue across 23 TCGA studies and 3 tissue aggregates. mRNA quantified by RSEM and scaled by log2 transform; Student's two-tailed t-test corrected for multiple hypothesis testing by multiplying by the number of tests for each gene (n=26). Studies reaching significant differential expression highlighted in red. Log2FC, $\log_2$ of fold change; p(adj), adjusted p value. Blue shading = significantly underexpressed; orange shading = significantly overexpressed; red shading = statistically significant.

TCGA study	*PPP1R13B* log2FC	*PPP1R13B* p (adjusted)	*TP53BP2* log2FC	*TP53BP2* p (raw)	*PPP1R13L* log2FC	*PPP1R13L* p (adj)
BLCA	-0.60	5.11E-03	-0.33	6.41E-01	1.51	1.26E-01
BRCA	-0.08	1.32E+00	0.28	2.00E-09	0.24	7.20E-01
CESC	0.06	2.23E+01	0.28	4.38E-01	3.04	2.57E-01
CHOL	0.53	5.41E-02	0.36	6.20E+00	2.73	5.83E-13
COAD	-0.60	5.43E-18	0.02	1.67E+01	1.04	5.44E-17

COADREAD	-0.51	9.11E-15	0.01	1.96E+01	1.16	8.15E-19
ESCA	-0.43	1.75E-01	0.03	2.06E+01	1.41	6.12E-01
GBM	-2.09	6.78E-02	-0.35	2.14E-01	1.23	4.49E-01
GBMLGG	-1.56	2.16E-01	0.09	8.63E+00	1.07	7.55E-01
HNSC	-0.35	2.88E-03	0.02	1.97E+01	0.02	2.33E+01
KICH	0.21	1.48E+00	-0.63	9.83E-04	-1.19	2.08E-07
KIPAN	-0.25	2.00E-10	-0.20	3.29E-05	1.03	1.43E-32
KIRC	-0.38	2.51E-14	-0.31	1.43E-07	1.59	1.81E-32
KIRP	-0.08	3.12E+00	0.12	1.98E+00	0.61	1.53E-04
LIHC	-0.05	9.06E+00	0.54	5.37E-04	0.74	3.57E-12
LUAD	-0.40	1.85E-06	-0.36	6.19E-05	0.17	5.93E-01
LUSC	-0.59	3.87E-15	-0.57	1.02E-08	0.71	3.06E-12
PAAD	-0.21	9.80E+00	-0.13	3.40E+00	0.55	1.28E+01
PCPG	0.07	2.12E+01	-0.16	7.31E+00	-2.30	4.51E-01
PRAD	0.18	3.70E-01	0.04	1.16E+01	-0.05	1.44E+01
READ	-0.17	4.74E+00	-0.02	2.24E+01	1.66	6.90E-04
SARC	-0.89	2.53E+00	0.34	5.81E+00	0.76	9.03E-01
STAD	-0.29	1.53E-02	0.29	8.93E-02	0.54	2.30E+00
THCA	-0.79	3.46E-20	-0.11	6.26E+00	0.85	1.18E-23
THYM	0.09	2.39E+01	-0.15	1.49E+01	0.62	2.33E+00
UCEC	0.24	7.34E-02	0.07	8.85E+00	1.36	6.00E-13
Significant DE (padj)		10/26		7/26		12/26
No. significant studies over/under-expressed	0/10		2/5			11/1

Table 4.5. Tumour vs normal mRNA expression from Oncomine. Methodology: p value threshold < 1e-5; TCGA studies excluded; cancer vs normal only; underexpression listed first, overexpression next. Total datasets available for cancer vs normal: n=190 although this includes some mis-classified cancer vs cancer sets. FC, fold change (linear scale; sign added for direction). Student's t-test on log2expression (adopted from Oncomine).

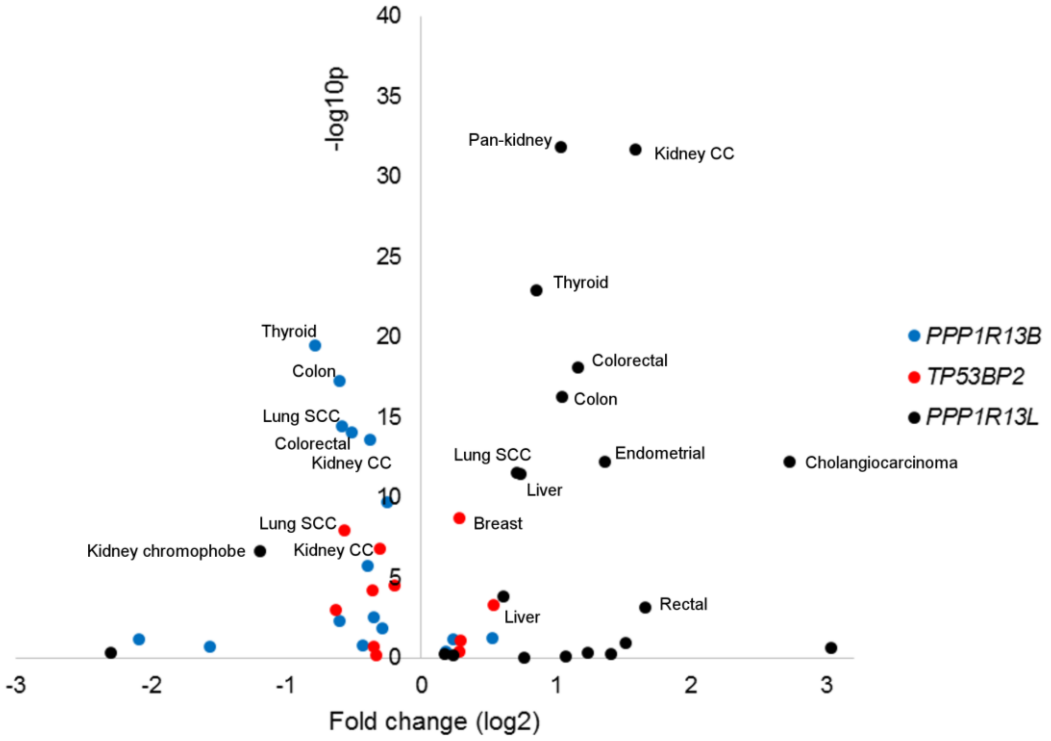
Tissue	Cancer	Study	Gene	FC	P value	Gene rank
Blood	Acute myeloid leukaemia (AML)	Haferlach	*PPP1R13B*	-1.213	5.30E-63	796
Blood	T cell acute lymphoblastic leukaemia (T-ALL)	Haferlach	*PPP1R13B*	-1.281	7.80E-32	806
Blood	Chronic myelogenous leukaemia (CML)	Haferlach	*PPP1R13B*	-1.237	1.20E-20	1626
Brain	Glioblastoma	Sun	*PPP1R13B*	-2.153	2.73E-17	921
Head and neck	Oral SCC	Peng	*PPP1R13B*	-1.512	1.68E-12	453
Brain	Oligodendroglioma	Sun	*PPP1R13B*	-1.68	5.22E-11	1052
Stomach	Gastric adenocarcinoma	Chen	*PPP1R13B*	-1.562	6.53E-11	145
Colon	Colorectal carcinoma	Skrzypczak	*PPP1R13B*	-1.701	3.94E-10	668
Testis	Mixed germ cell tumour	Korkola	*PPP1R13B*	-2.221	1.89E-09	999
Testis	Embryonal carcinoma	Korkola	*PPP1R13B*	-2.454	2.29E-09	479
Lung	NSCLC	Hou	*PPP1R13B*	-1.965	3.12E-09	1066
Brain	Anaplastic astrocytoma	Sun	*PPP1R13B*	-1.827	4.52E-09	343
Head and neck	SCC	Ginos	*PPP1R13B*	-1.628	1.11E-08	771
Esophagus	Adenocarcinoma	Kim	*PPP1R13B*	-1.673	2.30E-08	2613
Testis	Teratoma	Korkola	*PPP1R13B*	-1.976	1.05E-07	1294
Lung	SCC	Hou	*PPP1R13B*	-1.63	1.21E-07	2993
Brain	Glioblastoma	Bredel	*PPP1R13B*	-2.72	3.92E-07	742
Skin	Melanoma	Talantov	*PPP1R13B*	-1.522	9.46E-06	1021
Kidney	Clear-cell RCC	Beroukhim	*PPP1R13B*	-1.554	9.81E-06	804
Liver	Hepatocellular carcinoma	Roessler	*TP53BP2*	3.151	3.53E-83	44
Kidney	Renal oncocytoma	Jones	*TP53BP2*	-2.346	7.45E-15	402
Breast	Invasive breast carcinoma	Finak	*TP53BP2*	-4.112	2.73E-14	2158
Esophagus	Esophageal adenocarcinoma	Kim	*TP53BP2*	-2.684	8.46E-12	1537
Lymph	Peripheral T-cell lymphoma	Piccaluga	*TP53BP2*	-2.648	6.83E-11	463
Liver	Hepatocellular carcinoma	Chen	*TP53BP2*	1.68	7.58E-10	600

Lung	SCC	Hou	*TP53BP2*	-1.57	5.13E-09	2400
Pancreas	Ductal adenocarcinoma	Badea	*TP53BP2*	1.613	1.43E-08	1476
Bladder	Superficial bladder cancer	Modlich	*TP53BP2*	6.093	1.76E-08	3
Kidney	Papillary renal cell carcinoma	Jones	*TP53BP2*	1.671	2.43E-07	951
Blood	Chronic lymphocytic leukaemia	Haslinger	*TP53BP2*	-1.981	2.83E-07	129
Cervix	Cervical cancer	Pyeon	*TP53BP2*	2.186	1.06E-06	1676
Salivary gland	Adenoid cystic carcinoma	Frierson	*TP53BP2*	3.374	4.42E-06	225
Lung	Lung carcinoid	Bhattacharje e	*TP53BP2*	-6.272	8.71E-05	1808
Esophagus	Esophageal carcinoma	Kim	*PPP1R13L*	-5.076	1.26E-26	91
Bladder	Infiltrating urothelial carcinoma	Sanchez-Carbayo	*PPP1R13L*	2.561	1.18E-15	115
Kidney	Renal clear cell carcinoma (hereditary)	Beroukhim	*PPP1R13L*	2.451	1.67E-12	142
Bladder	Superficial bladder cancer	Sanchez-Carbayo	*PPP1R13L*	4.774	2.68E-12	933
Head and neck	SCC	Ginos	*PPP1R13L*	-2.196	8.03E-11	465
Lung	Lung adenocarcinoma	Selamat	*PPP1R13L*	1.58	1.01E-10	1300
Rectum	Rectal adenocarcinoma	Gaedcke	*PPP1R13L*	1.563	1.70E-10	3039
Pancreas	Ductal adenocarcinoma	Badea	*PPP1R13L*	1.657	7.24E-10	587
Colon	Colorectal adenocarcinoma	Skrzypczak	*PPP1R13L*	1.668	1.55E-09	769
Kidney	Renal clear cell carcinoma (non-hereditary)	Beroukhim	*PPP1R13L*	2.367	1.81E-09	239
Skin	Melanoma	Riker	*PPP1R13L*	-6.182	7.84E-08	9
Testis	Mixed germ cell tumour	Korkola	*PPP1R13L*	1.629	5.16E-07	2072
Kidney	Renal clear-cell carcinoma	Jones	*PPP1R13L*	1.6	5.67E-07	1768
Colon	Colon adenoma	Sabates-Bellver	*PPP1R13L*	1.726	1.04E-06	3247
Rectum	Rectal adenoma	Sabates-Bellver	*PPP1R13L*	2.392	1.19E-06	560
Esophagus	SCC	Hu	*PPP1R13L*	-2.445	3.33E-06	922

Breast	Ductal carcinoma	Richardson	*PPP1R13L*	-2.159	4.76E-06	1036
Skin	Melanoma	Talantov	*PPP1R13L*	-4.781	5.47E-06	789

Table 4.6 Summary of tumour vs normal RNAseq and microarray data. TSG, tumour suppressor gene.

Gene	Source	Threshold	No. studies underexpressed	No. studies overexpressed	Concordance rate	Dominant class
PPP1R13B	Oncomine	P<1e-5; abs(FC)>1.5	19	0	100%	TSG
PPP1R13B	TCGA	Padj<0.05	10	0	100%	TSG
TP53BP2	Oncomine	P<1e-5; abs(FC)>1.5	7	7	50%	N/A
TP53BP2	TCGA	Padj<0.05	5	2	71%	TSG
PPP1R13L	Oncomine	P<1e-5; abs(FC)>1.5	6	12	67%	Oncogene
PPP1R13L	TCGA	Padj<0.05	1	11	92%	Oncogene



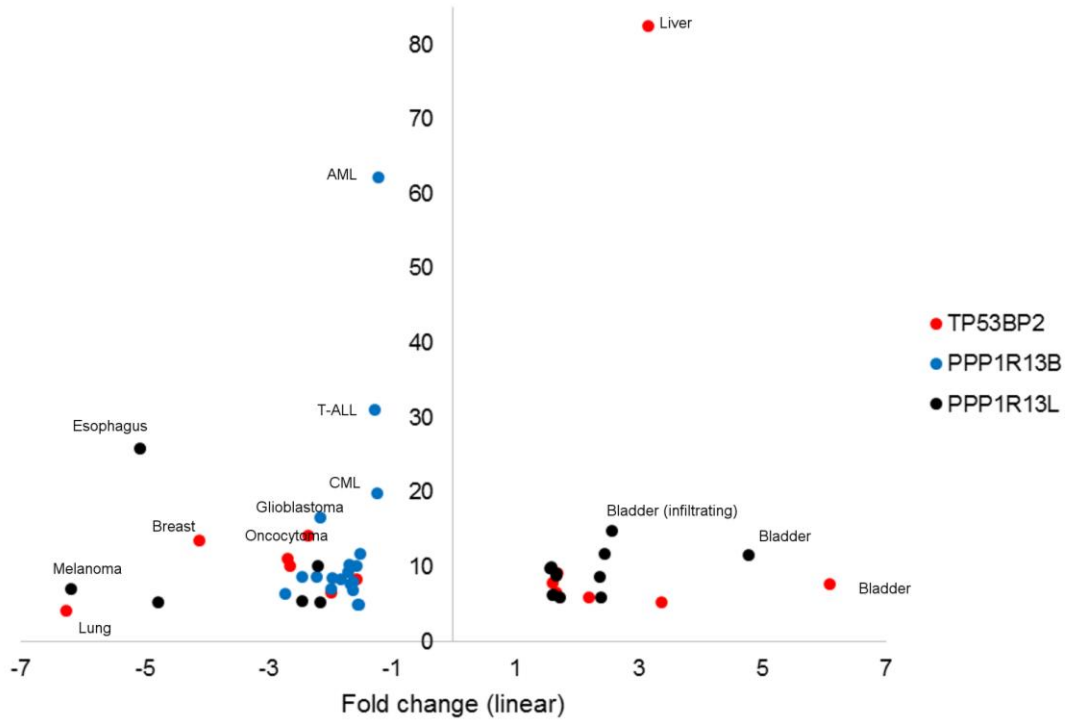


Figure 4.5 Volcano plots of TCGA (top) and Oncomine (bottom) mRNA data. AML, acute myeloid leukaemia; CML, chronic myeloid leukaemia; T-ALL, T-cell acute lymphoblastic leukaemia.

4.1.6 Mining genome-wide CRISPR sgRNA screen of cancer cell line viability

In addition to genetic studies in mice, human cell lines offer another opportunity for investigating causal effects of individual genes on cancer growth. Traditionally, knock-down techniques using short-interfering RNA (siRNA) or short-hairpin RNA (shRNA) reagents have been used to assess the effect of gene loss on cancer cell line growth. However, these techniques result in incomplete reduction in gene expression, and frequent off-target effects. With the advent of genome editing, genes can be removed at the DNA level, eliminating the uncertainty over mRNA dosage effects (376, 377).

A recent example of utilising the genome editing system CRISPR-Cas9 to probe the functional genome in cancer cell lines is a study by project ACHILLES which screened 33 cell lines with genome-wide sgRNA libraries (378). Mining this publicly available data revealed little cell line dependence on *PPP1R13B* for most cell lines in the panel (more positive values indicate less cell line gene-dependence) (**Figure 4.6**). Surprisingly, some cell lines showed a more negative response to the targeting of *TP53BP2* – the tails of the gene's distribution reaches wider than the distribution of *PPP1R13L*'s sgRNA scores, suggesting a subset of cell lines is more sensitive to the deletion of *TP53BP2* than *PPP1R13L,* for example the pancreatic cancer cell line PANC0327 (*TP53BP2* score= -1.05). This observation might help explain why *TP53BP2* is upregulated in certain tumour types. However, a higher number cell lines needs to be screened in order to achieve a global survey of the ASPP genes' effect on cancer cell line fitness.

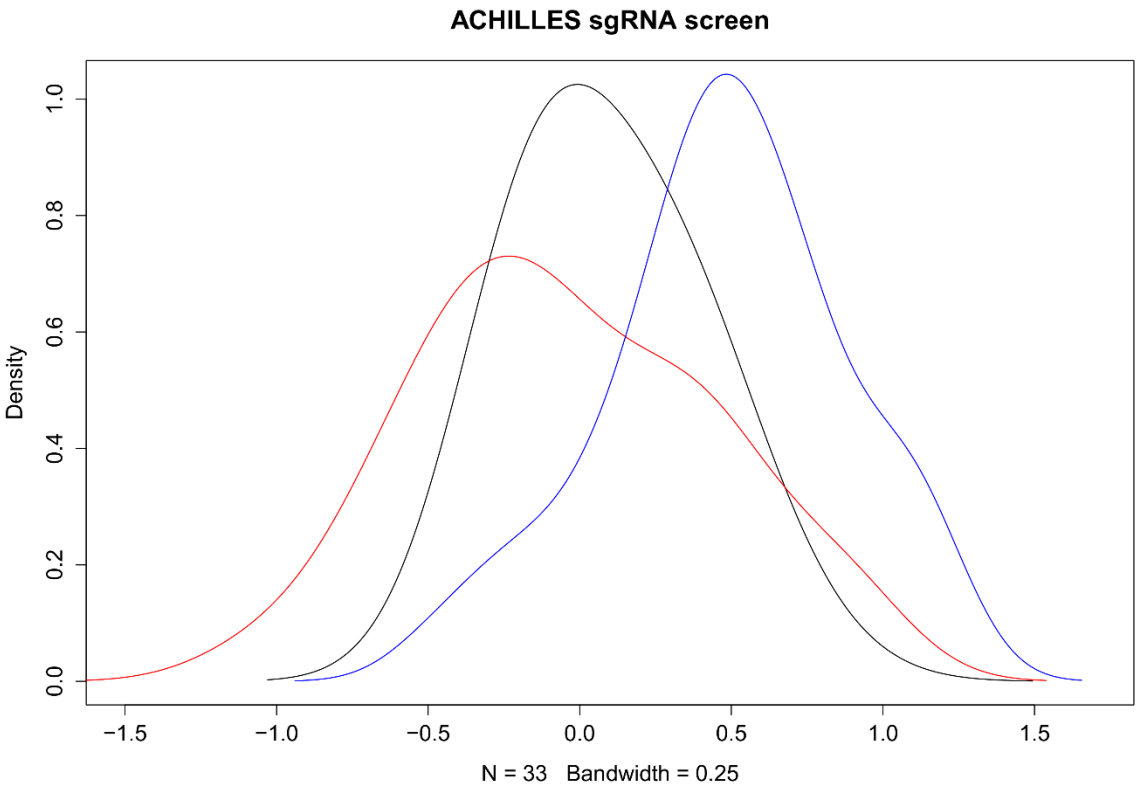


Figure 4.6 Distribution of ASPP gene sgRNA scores in the ACHILLES CRISPR screen: x-axis indicates sensitivity to sgRNA integration: positive values indicate less sensitivity. Blue = *PPP1R13B*; red = *TP53BP2*; black = *PPP1R13L*.

4.1.7 Analysing the correlation between mRNA levels and cell line chemosensitivity

An important hallmark of oncogenes is mediating resistance to cell death, including cell death induced by anti-cancer drug treatment. Conversely, functional tumour suppressor pathways enhance the efficiency of drug action by triggering cell death or growth inhibition in response to drug treatment. Therefore, NCI's Cancer Target Discovery 2 Portal (https://ctd2.nci.nih.gov/dataPortal/) along with an accompanying paper by Rees et al. (379) were mined for information on gene correlation with drug sensitivity in cancer cell lines. First, the all-cancer-cell line correlations were mined and data points reaching experiment-wide significance were plotted (**Figure 4.7**). Strikingly, a large cluster of drugs correlates with poorer

cell line sensitivity (high z-scores) as a function of increasing *PPP1R13L* expression (**Figure 4.7**). In contrast, several drugs correlated with chemosensitivity in cell lines expressing high *PPP1R13B*. These associations reaffirm *PPP1R13B*'s profile as a candidate TSG whereas *PPP1R13L*'s as a candidate oncogene in human cancer.

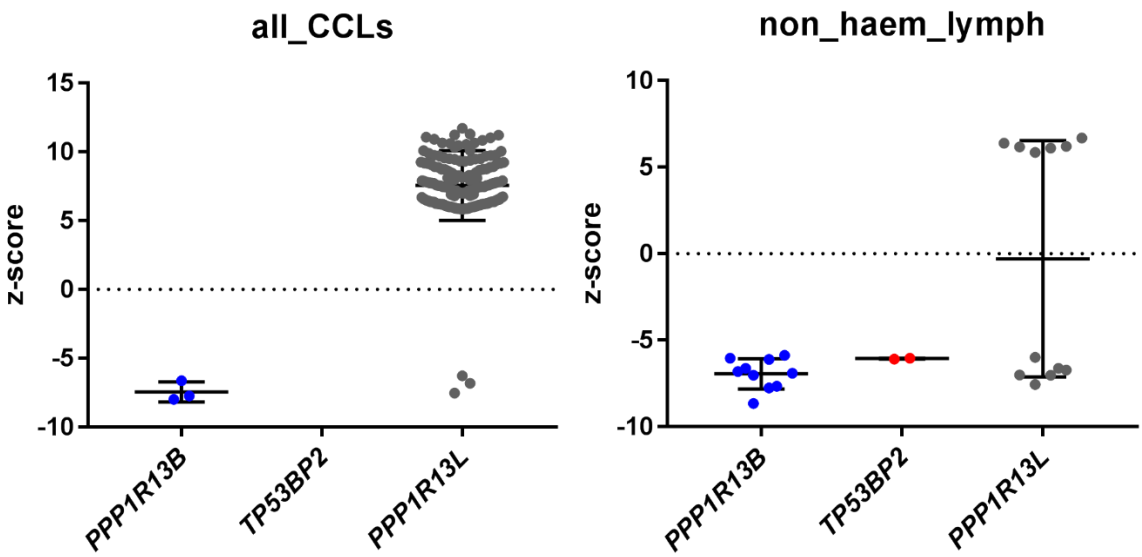


Figure 4.7 Associations between ASPP mRNA levels and cell line chemosensitivity. Only associations clearing the experiment-wide Bonferroni-corrected significance threshold of |z| > 5.83. Positive association indicates correlation with resistance, computed as area under the dose response curve. CCLs, cancer cell lines; non_haem_lymph, non-haematopoietic and lymphoid cell lines. Data retrieved from Rees et al, 2016 (379). 481 compounds were tested in this experiment.

Plotting z-scores for all drug associations shows that the distribution of *PPP1R13L* z-scores is positively skewed (**Figure 4.8**), indicating an overall positive correlation with drug resistance and negative correlation with chemosensitivity in cancer cell lines. The z-scores of *PPP1R13B* are centred close to 0, however notable outliers exist in the negative-z-score region as shown above. Interestingly the z-scores for *TP53BP2*'s association form a narrow

distribution centred just below 0, implying that mRNA levels of *TP53BP2* in cancer cell lines correlate poorly with chemosensitivity.

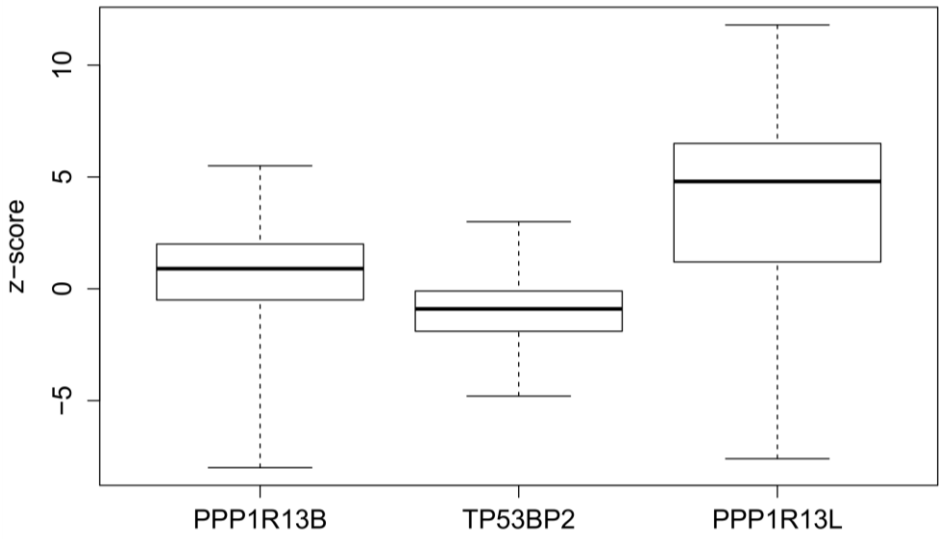


Figure 4.8 All z-scores for Pearson correlation of ASPP gene mRNA levels with chemosensitivity; for all CCLs. Data mined from the CTRP v2 portal. Boxplots shows median, quartiles and whiskers show minimum and maximum values.

4.1.8 Lineage specificity of cancer genes

The TSG or oncogene function of some genes is restricted to one or more cell lineage (e.g. *MLL2*); other multi-modal genes act as TSGs in one lineage and oncogenes in another (e.g. *NOTCH1*). Therefore, to gain a full understanding of the functional role of a given gene in cancer, the data on its dysregulation must be de-convolved by tumour type. One challenge in performing such analysis is the lack of comprehensive data on all tumour types – with the exception of TCGA, most cancer studies were limited to a small number of analyses, such as whole-exome sequencing. In addition, studies vary significantly in their scale, introducing a bias in confidence as a function of sample size, e.g. the differential expression of

a sarcoma driver gene may be more significant in lung cancer because more lung tumours and matched tissue controls have been analysed to date.

Nevertheless, data may be parsed by tissue type and assessed for concordance of evidence on gene function. As an example, data for *PPP1R13B* and the mouse orthologue *Ppp1r13b* are presented in **Table 4.7**. The data reveal a convincing role for ASPP1 as a TSG in colon cancer: in mice, 3 independent transposon screen identified *Ppp1r13b* as a hit; the gene is significantly downregulated in human colon tumours vs normal tissue ($p < 10^{-14}$), and 14 truncating mutations in human colon tumours have been reported, making up 27% of nonsynonymous mutations in this tissue, significantly higher than expected by random chance. These highly concordant data demonstrate *PPP1R13B* is a TSG in colon cancer.

Table 4.7 Evidence of gene cancer role by tissue type for *PPP1R13B.* LoH, loss of heterozygosity.

Tumour type	Hit in mouse transposon screen?	Downregulated in human tumours vs normal tissue?	Truncating mutations in human tumours?	Truncating muts with LoH?	Missense muts with LoH?	% of truncating mutations of all nonsilent somatic
Colon	Yes (3x) Takeda 2015; March 2011; Starr 2009	Yes, p=4.45E-15	Yes n=14			14/51=27%
Liver	Bard-Chapeau 2014 (HBV+)	No, p=n.s.	Yes	Yes, n=2	Yes, n=2	
Peripheral nerve sheath (neurofibrosarcoma)	Rahrmann 2013	N/A	N/A			
Pancreas	Perez Mancera 2012	Yes, p=0.092	Missense only			
Stomach carcinoma	Takeda 2016		Yes, 7		Yes, n=2	
Bladder	-	Yes, 0.003	Missense only			
HNSC	-	Yes, p=9.0E-4	Missense only			
Lung adenocarcinoma	-	Yes, p=6.0E-4	Yes	Yes	Yes	
Lung squamous	-	Yes, p=3.46E-9	Yes	Yes	Yes, n=2	
Leukemia	Yamashita 2015 – p53-/- potentiated	Yes				
Prostate		No, p=0.076			Yes	

Interestingly, loss of heterozygosity was not observed for nonsynonymous *PPP1R13B* mutations in colon cancer. This, along with the data presented above, suggests that *PPP1R13B* is not a recessive but dosage-dependent TSG in this tissue. Partial loss of ASPP1 function is sufficient to enhance tumorigenesis in the colon. In contrast, liver carcinomas are more likely to require full inactivation of *PPP1R13B*: 2 truncating and 2 missense mutations with LoH were observed but mRNA levels of *PPP1R13B* did not differ significantly between tumour and normal tissue. In accordance with the recessive model, it is possible the subset of liver tumours with biallelic inactivation of *PPP1R13B* benefit, while tumours carrying at least one wild type copy of the gene do not experience selection on ASPP1 levels. Taken together, these data suggest the classification of a TSG as strictly recessive versus dosage-dependent can vary by lineage.

4.1.9 Towards genome-wide classification of cancer genes

The data presented in this chapter provide unbiased, multi-platform, multi-source evidence that can be used to assess the likelihood of any gene being a cancer gene, as well as the likelihood of being a TSG versus an oncogene, or more broadly, being anti- or pro-tumorigenic. While each individual study has its caveats, concordance across multiple data types from multiple independent studies can provide conclusive evidence.

The ASPP genes represent a suitable example of the utility of this approach. Although an extensive body of evidence exists of the ASPP genes' involvement in cancer (see Chapter 1, section 1.4.6), none of the ASPP genes are included on the Sanger Cancer Gene Census (cancer.sanger.ac.uk/census/), Vogelstein et al.'s cancer gene list (151), the list of significantly mutated genes

called by MutSig2CV (162), or the Network of Cancer Genes 5.0 (200). Therefore, a significant gap exists between the experimental data supporting a causal role for ASPPs in cancer, and curated cancer genomic data-derived lists of cancer drivers. The analysis presented here aids in closing that gap. **Table 4.8** shows an example of how the different types of evidence can be interpreted. In this example, we establish thresholds for the classification of each gene as a TSG, oncogene (OG), or inconclusive (0). The number of TSG and OG points scored is tallied up and their concordance evaluated.

For example, *PPP1R13B* scored 6 TSG points and 0 OG points, equivalent to a 100% concordance rate. Given that this evidence includes multiple hits from independent mouse mutagenesis screens, and in light of the experimental evidence discussed in Chapter 1.4.6, there is little doubt that *PPP1R13B* is a tumour suppressor gene in mice and humans. Its relatively low absolute somatic mutation frequency suggests that *PPP1R13B* mutations might act as rare driver mutations, mini-driver mutations, and/or that the major mode of action of *PPP1R13B* in cancer is that of an Epi-driver gene, i.e. epigenetic dysregulation of ASPP1 function is sufficient for inactivation in most tumours. This is consistent with the observation of an extremely high concordance rate of *PPP1R13B* mRNA under-expression across human cancers. Perhaps down-regulation of ASPP1 sufficiently reduces the threshold required for ASPP1-dependent activation of apoptosis, and somatic mutation of *PPP1R13B* is required only in tumours where ASPP1 experiences unusually high levels of activation and/or where pathways downstream of ASPP1, such as the cell-intrinsic apoptosis pathway, are not already inactivated by another mechanism.

PPP1R13L, implicated as a candidate proto-oncogene by existing experimental data, shows hallmarks of an oncogene such as consistent upregulation in tumours vs normal tissue, and significant association of mRNA levels with drug resistance in cancer cell lines (**Table 4.8**). However, a close inspection reveals that *PPP1R13L* does not frequently experience copy number gain or amplification in cancer, and few highly recurrent missense mutations have been found in *PPP1R13L*. Taken together, *PPP1R13L* dysregulation fits the profile of an Epi-driver oncogene, i.e. a pro-tumorigenic gene which can be sufficiently co-opted to drive tumour growth in its wild-type state by epigenetic and post-translational mechanisms.

Finally, *TP53BP2* shows an ambiguous pattern of dysregulation in cancer, with evidence of both TSG function (>20% truncating somatic mutations, downregulation in some tumour types) and OG function (>25% somatic amplification rate, upregulation in some tumour types). Given that *TP53BP2* is significantly mutated in human pancreatic cancer and shows truncating mutations in this dataset (266); *Trp53bp2* is a haploinsufficient tumour suppressor in mice (104, 105); and ASPP2 is directly targeted by the oncoprotein CagA in the stomach (243, 244), it is likely that *TP53BP2* works as a TSG in some tissues and an OG in others.

Table 4.8 Multi-variable assessment and classification of potential cancer genes. ASPP genes used as an example. Threshold of 20% truncating mutations were used (based on Vogelstein et al.); concordance threshold of 66% was required for expression data. N/C, not conclusive; homdel, homozygous deletion; NT, normal tissue; TP, primary tumour tissue; cBio, cBioPortal; hetloss, heterozygous deletion; CTRP, Cancer Therapeutics Response Portal.

TSG/Oncogene classification	Current threshold	*PPP1R13B*	TSG/OG	*TP53BP2*	TSG/OG	*PPP1R13L*	TSG/OG	class
Causal experimental evidence								
ACHILLES2016: mean score across all cell line		0.479		-0.039		0.089		cell line
ACHILLES: % of studies –ve growth impact (with dependency)		4		19		17		cell line
No. of hits in mouse TSG transposon screens		7	TSG	0	0	0	0	mouse
Patient genomic evidence								
% truncating muts of all nonsilent – of cBio (their summary)	20%	24.6	TSG	15.6	0	11.3	0	mutations
% truncating muts of all nonsilent – curated	20%	20.3	TSG	19.0	0	11.2	0	mutations
No. of truncating muts with LoH	>2	4	TSG	3	TSG	0	0	Mut+CNV
% of truncating muts with LoH		8.5		5.4		0.0	N/C	Mut+CNV
No. of missense muts with LoH		15		1		8	N/C	Mut+CNV
% of missense muts with LoH		8.3		0.4	N/C	4.2	N/C	Mut+CNV
No. of missense muts with gain/amp		11		39		13		Mut+CNV
% of missense muts with gain/amp		6.1		16.8		6.9		Mut+CNV
% homdel (overall, cBio)		0.8		0.2		0.3		CNV
% of hetloss/homdel (overall, cBio)		18.3		4.9		13.7		CNV
% amp (overall, cBio)		0.8	0	25.7	OG	0.9	0	CNV
% of gain/amp		11.0		28.1	OG	14.4		CNV
mRNA TP vs NT: no. of studies sign underexpressed [TCGA]		10		5		1		mRNA
% of studies sign underexpressed [TCGA]	>66	100	TSG	71	TSG	8	0	mRNA
mRNA TP vs NT: no. of studies signif. Overexp. [TCGA]		0		2		11		mRNA
% of studies sign overerexpressed [TCGA]	>66	0	0	29	0	92	OG	mRNA
mRNA TP vs NT: no. of studies sign underexpressed [Oncomine]		19		7		6		mRNA
% of studies sign underexpressed [Oncomine]	>66	100	TSG					mRNA
mRNA TP vs NT: no. of studies signif. Overexp. [Oncomine]		0		7		12	OG	mRNA
% of studies signif. overexp. [Oncomine]	>66	0	0	50	0	67	OG	mRNA
Correlational experimental evidence								
CTRPv2: median z-score of Pearson corr	z>1.95	0.9	0	-0.9	0	4.8	OG	cell line
CTRPv2: no. of drugs z>0 at whole-exp signif threshold, i.e. z> 5.83		0		0		154		cell line
% of drugs z>0 at whole-exp signif threshold, i.e. z> 5.83	>66	0%		N/A		98%	OG	cell line
CTRPV2: no. of drugs z<0 at whole-exp signif threshold, i.e. z<-5.83		3		0		3		cell line
Concordance rate (%)			100		50		100	
Score			6		2+2		5	
Classification			TSG		N/C		OG	

In order to discover all cancer genes, the approach described in this section can be extended genome-wide. One possible strategy, demonstrated in **Table 4.8**, involves a points-based system with classification gates established with a training set of curated cancer genes. This approach is similar to clinical guidelines directing treatment based on empirically scored observations of the patient's phenotype. The data for each gene can be parsed dynamically from public APIs to ensure the classification can be updated as datasets grow.

The disadvantage of a points-based system is its limitation to the variables examined and the analysis applied to calculate scores. Scoring relies on intelligent decisions regarding the relative importance of different types of biological evidence. However, our understanding of the biology of cancer is incomplete. Therefore, a more powerful approach will utilise machine learning applied to a wide range of variables. The limitations of the machine learning approach are 1) the bias introduced by a specific class of curated cancer genes, and 2) the degree to which tumour data of different types and sources can be synchronised. However, we envisage that with the emergence of platforms such as the NCI Genomic Data Commons, the degree of harmonisation of cancer genomic and clinical data in the public domain will keep increasing. Future work will investigate the genome-wide application of this integrative approach to cancer gene classification.

Results section 4.2: Dissecting the mechanism of action of driver missense mutations

4.2.1 Genetic interaction between *TP53* and *PPP1R13B* mutations

In order to fully understand the biology of cancer, molecular mechanisms underlying tumorigenesis must be mapped. Two major challenges in the analysis of function of missense mutations in cancer are separating driver from passenger mutations, and deconvolving individual functions of a pleiotropic gene. *PPP1R13B* is an example of a pleiotropic gene: its protein product ASPP1 has multiple binding partners across distinct pathways, such as p53, p63, p73, YAP and PP1. It is unclear which, if any, of these interactions are targeted by missense mutations in cancer.

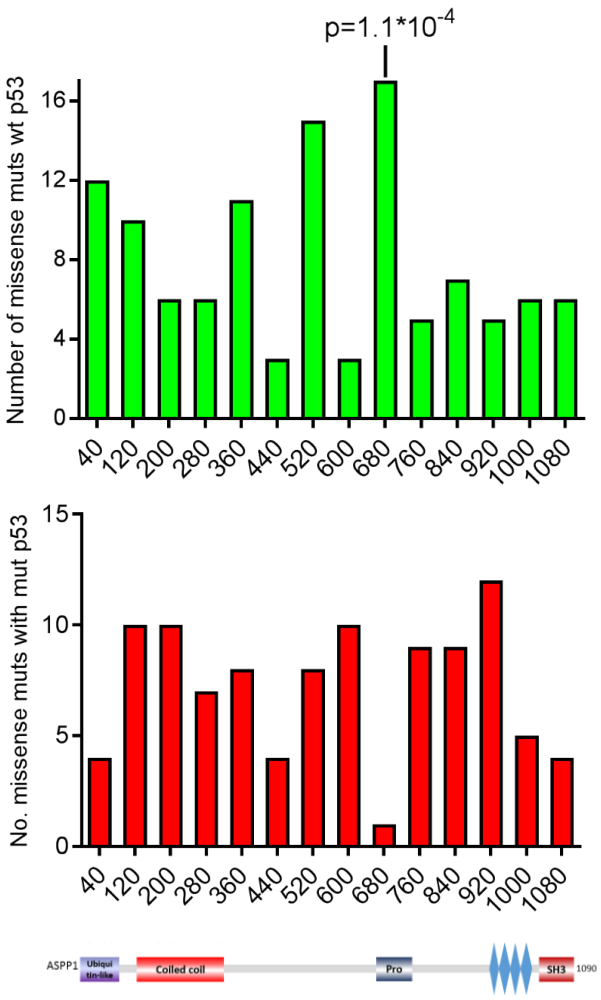


Figure 4.9 Frequency distribution of missense ASPP1 somatic mutations in wt TP53 (top) and mutant TP53 tumours (bottom) along ASPP1 protein sequence. Horizontal axis is the residue number at the centre of each bin. Mutations in the 640-720 region occur almost exclusively in wt-p53 tumours; Fisher's 2-tailed exact test comparing number of mutations in 640-720 bin vs total in other bins.

Because genetic interactions between somatic mutations can be lost in the background of passenger mutations, we analysed the domain distribution of missense mutations in ASPP1 in p53-wild type vs mutant tumours (**Figure 4.9**). Only mutations detected in tumours (not cell lines) were included in this analysis. Interestingly, we found one 80-residue region of ASPP1 that is mutated almost exclusively in wild-type p53 tumours ($p = 1.1x10^{-4}$, Fisher's exact two-tailed test). This demonstrates there is a genetic interaction between somatic mutations in *PPP1R13B* and *TP53*. This sequence region, residues 640-720, is also the most frequently mutated region of ASPP1 in wild type-p53 tumours. It includes residues in the N-terminal part of the proline-rich (Pro-rich) region and the preceding residues (**Figure 4.9**). These data suggest the Pro-rich region of ASPP1 directly or indirectly modulates ASPP1's interaction with p53 and that this mechanism is targeted by missense mutations in human cancer.

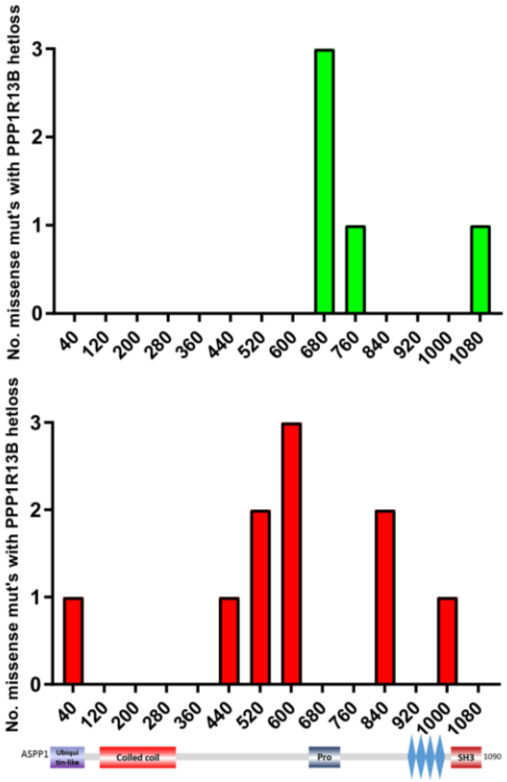


Fig. 4.10 Missense mutations in Pro-rich region of ASPP1 are enriched in tumours with loss of *PPP1R13B* heterozygosity (hetloss). Tumours carrying missense *PPP1R13B* mutations with loss of heterozygosity and wt *TP53* (top, green) vs mutant *TP53* (bottom, red).

We also observed that mutations in and preceding the Pro-rich region are enriched in tumours with loss of *PPP1R13B* heterozygosity (**Figure 4.10**). Missense mutations accompanied by LoH are a frequent mechanism of recessive TSG inactivation. This supports the hypothesis that the Pro-rich region and pre-Pro-rich-region sequence mutations have undergone positive selection in tumours and suggests they have an inactivating rather than activating role.

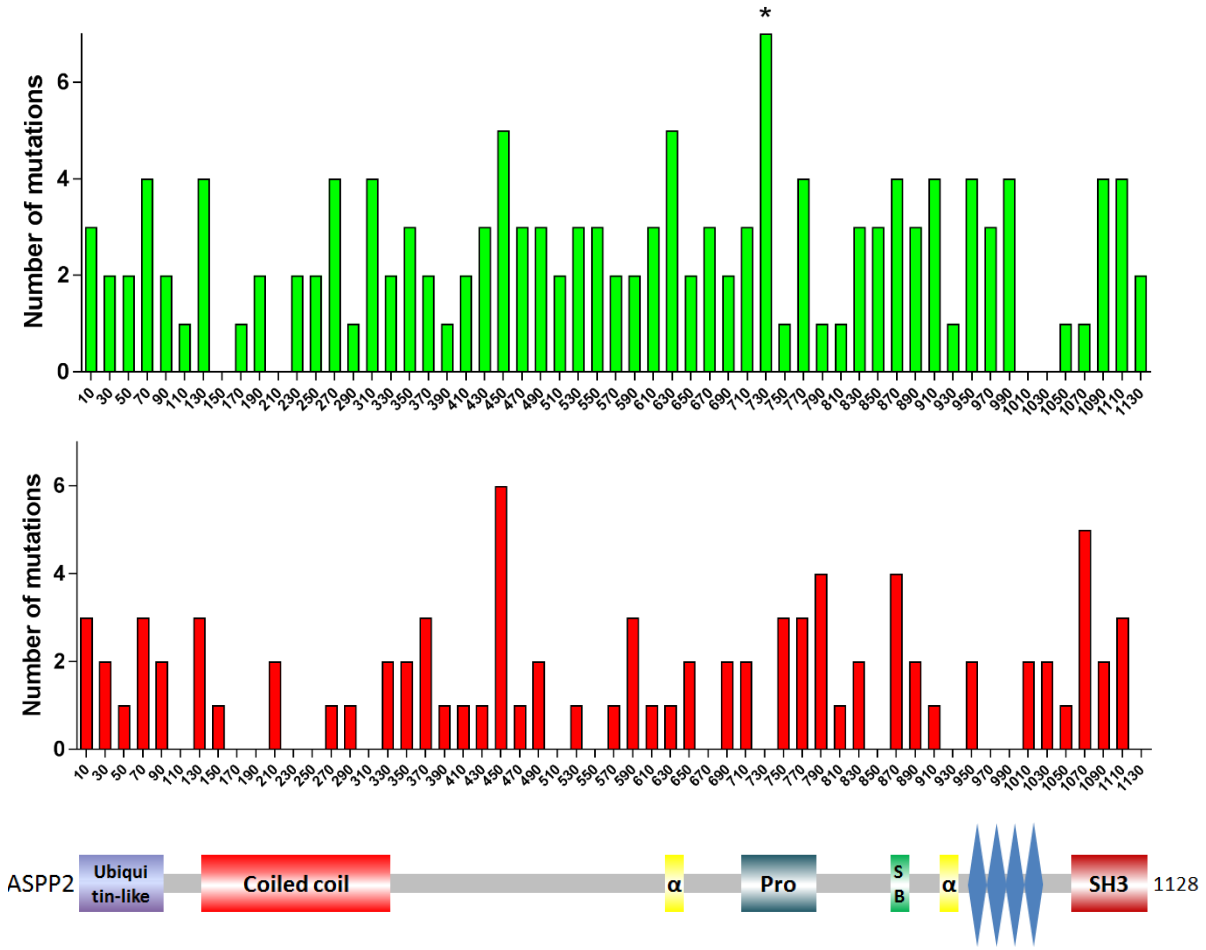


Figure 4.11 Somatic missense *TP53BP2* mutations in wt-p53 (top) and mutant-p53 tumours (bottom). The top mutated bin in wt-p53 tumours centred at residue 730 is not mutated in any mutant-p53 tumours. Data source: manually curated mutations from COSMIC v78, cBio v1.2.4 and ICGC r22. * p<0.05, Fisher's exact two-tailed test comparing bin size with sum of other bins in wt-p53 vs mutant-p53 tumours.

Mutations in *TP53BP2* show a similar and statistically significant, but more narrowly defined relationship with *TP53* status in the Pro-rich region (**Figure 4.11**). The specificity of missense mutations to wild type-p53 tumours is limited to the ASPP2 residues 720-740, whereas in the more C-terminal part of the Pro-rich region (abbreviated C-Pro), mutations are more frequent in mutant-p53 tumours. This suggests the more N-terminal part of the Pro-rich region (abbreviated N-Pro below) in ASPP1 and ASPP2 plays a role in controlling the interaction of each protein with p53. However, as in ASPP1, the N-Pro region of ASPP2 is also the

most frequently mutated region of ASPP2 sequence in wt-p53 tumours (**Figure 4.11**), highlighting an important role for this autoregulatory region of ASPP1 and ASPP2 in cancer.

4.2.2 Proline-rich region mutations in *PPP1R13B* and *TP53BP2* correlate with tumour *BAX* mRNA levels

ASPP1 and ASPP2 enhance p53's transcriptional activity in a promoter-specific manner, thereby increasing the transcription of pro-apoptotic genes *BAX*, *TP53I3* and *BBC3* at the expense of other p53-target genes such as *CDKN1A*. Therefore, we asked whether missense mutations interacting with *TP53* mutation status correlated with altered *BAX* mRNA levels in each tumour. Because absolute levels of *BAX* are tissue-dependent, we used mRNA z-scores comparing the *BAX* mRNA level in each tumour to the expression level of *BAX* in other tumours of that tissue type. To visualise the distribution of *BAX* levels as a function of ASPP1 mutations, we plotted the density of tumours with negative vs positive *BAX* z-scores along the ASPP1 sequence (**Figures 4.12 and 4.13**).

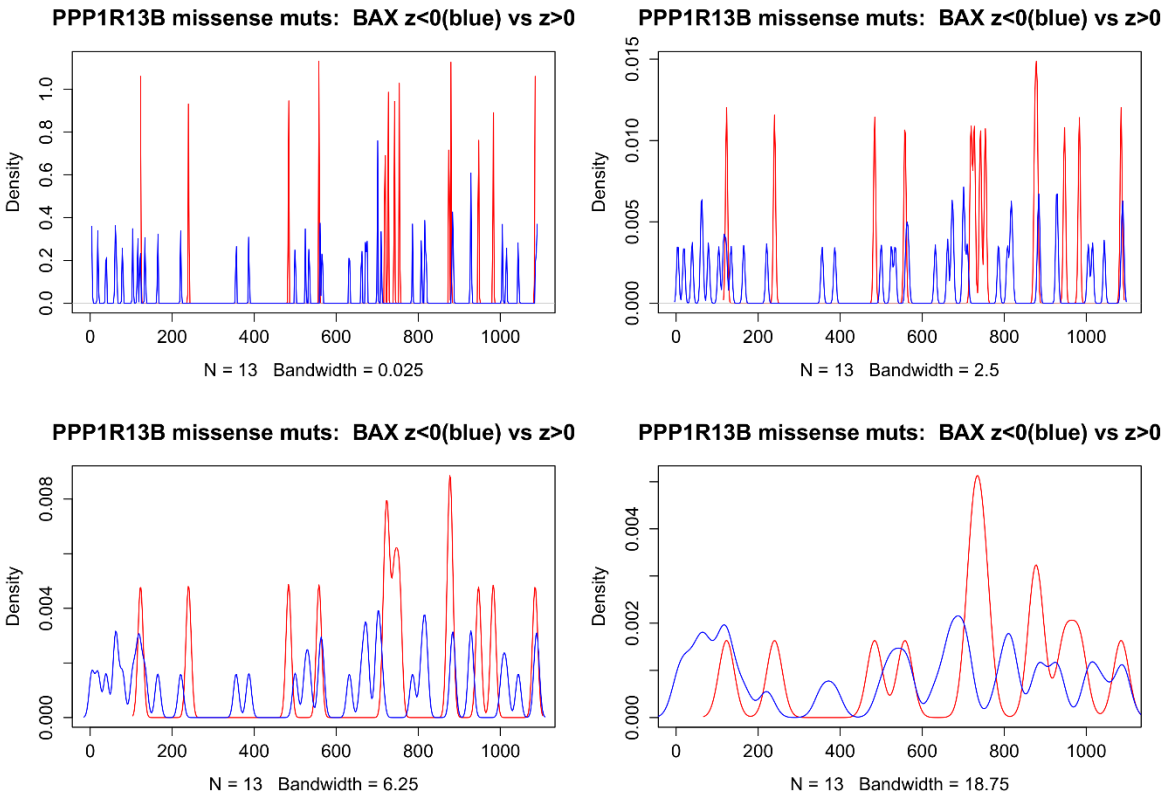


Figure 4.12 *BAX* z-scores vs *PPP1R13B* missense mutations using different density bandwidths.

As expected, tumours with mutations in the N-Pro and preceding region had lower levels of *BAX* than other tumours of the same tissue type. Tumours with negative *BAX* z-scores also clustered within the ubiquitin-like fold, suggesting a role for this domain in regulating ASPP1's interaction with p53 (**Figures 4.12 and 4.13**). Similar results were observed for N-Pro mutations in ASPP2 (**Figure 4.13**) with all ASPP2 N-Pro-mutant tumours showing negative z-scores of *BAX*. This result was encouraging because there was no overlap between patients with N-Pro mutations in ASPP1 and ASPP2, suggesting this correlation is shared by the two proteins and can be replicated.

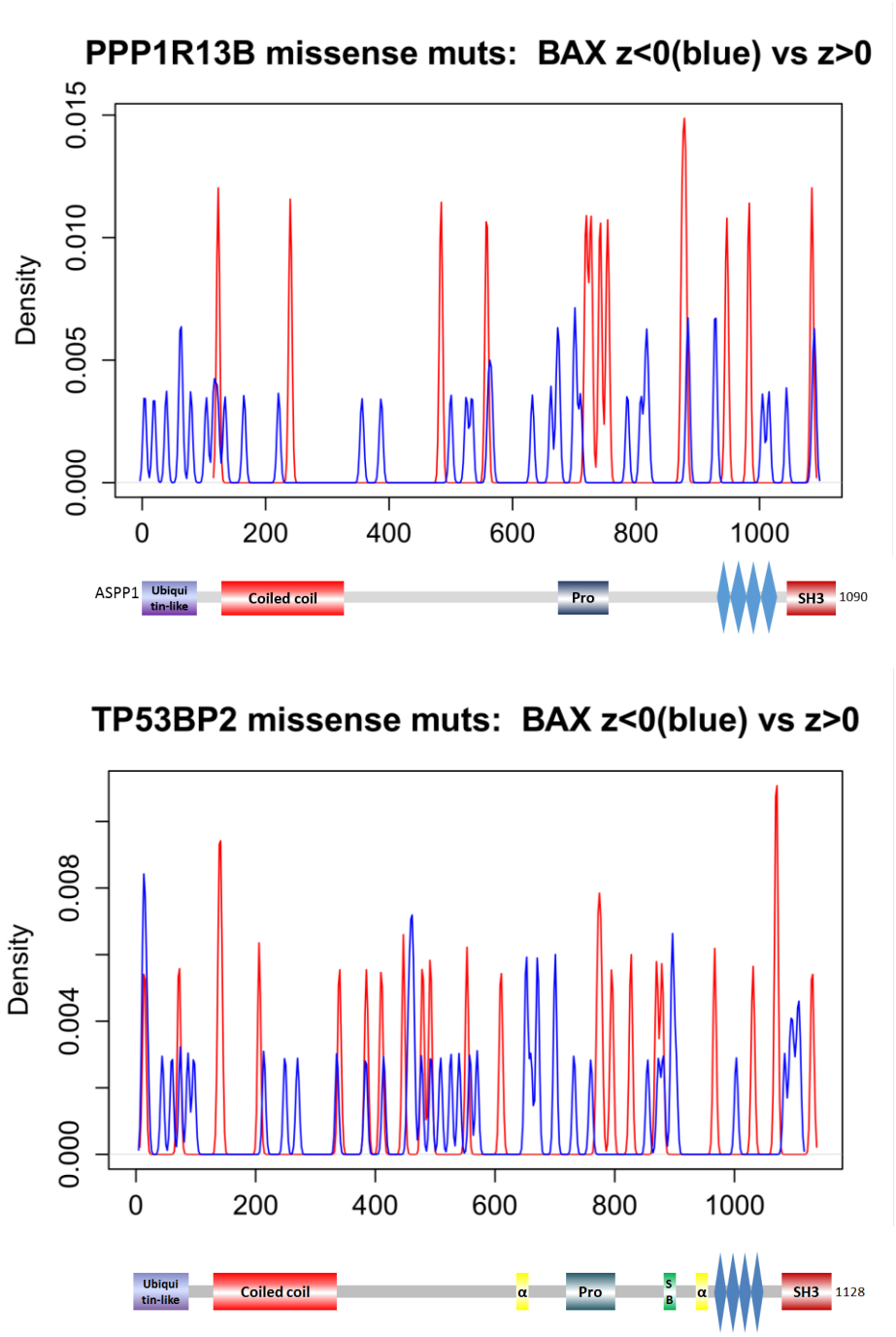


Figure 4.13 Density plot of tumours with above-average BAX mRNA levels (red) vs below-average (blue) within the tumour type, in mutant-ASPP1 (top) and mutant-ASPP2 (bottom) tumours, as a function of ASPP protein sequence. Pro, proline-rich region; SB, SH3-binding domain; SH3, SH3 domain; α, predicted α-helical sequence.

4.2.3 Dipolar correlation of Pro-rich region mutations with *BAX* levels

The distribution of *PPP1R13B*-mutant tumours with above-average *BAX* levels ($z > 0$) vs below-average levels ($z < 0$) shows a clear pattern within the

proline-rich region (**Figure 4.14**). Whereas in most domains tumours with positive and negative z-scores are mixed together, in the Pro-rich region a cluster of $z < 0$ tumours is adjacent to a cluster of $z > 0$ tumours with no visible overlap. Strikingly, this region shows a more significant degree of differentiation by *BAX* levels than the C-terminus of ASPP1 which binds p53 family members directly. Unexpectedly however, not all Pro-rich region mutant tumours have negative z-scores, instead, mutations in the C-Pro region are associated with increased levels of *BAX* for both ASPP1 and ASPP2-mutant tumours (**Figure 4.14**).

This result is surprising for two reasons. Firstly, ASPP1 Pro-rich region mutations came exclusively from wild-type-*TP53* tumours, whereas ASPP2 Pro-rich mutations were observed in both wild type-*TP53* and mutant-*TP53* tumours. Secondly, it is unclear why somatic mutations linked with increased mRNA levels of a pro-apoptotic gene such as *BAX* would be positively selected. Indeed, the observation of negative selection of co-occurrence of mutations in the Pro-region of ASPP1, and in *TP53* within the same tumour suggests ASPP1 protein with Pro-rich-region mutations can exhibit enhanced pro-apoptotic activity in mutant-*TP53* tumours. Therefore, why are these ASPP1 and ASPP2 mutations not selected out?

One possibility is that mutant ASPP1 and ASPP2 with C-Pro mutations (somatic mutations in the proline-rich region closer to the C-terminus) exhibit both pro- and anti-tumour effects, of which the pro-tumour effect becomes dominant to allow positive selection (limited to wt-p53 tumours for ASPP1). In a mutant-p53 tumour, the anti-tumour effect becomes dominant and the mutation is selected out. To investigate this possibility, we surveyed the relationship between ASPP1, ASPP2 mutations and mRNA levels of other p53 target genes. Interestingly,

levels of *CDKN1A* which encodes the cell cycle inhibitor p21 were reduced in tumours carrying C-Pro mutations in *TP53BP2* and *PPP1R13B* (**Figure 4.14**). A cluster or below-average-p21-level tumours was observed in the C-Pro region of both genes. This is unexpected as *BAX* and *CDKN1A* levels are generally positively correlated in tumours and indeed in the N-Pro-mutant tumours, both *BAX* and *CDKN1A* levels are below average.

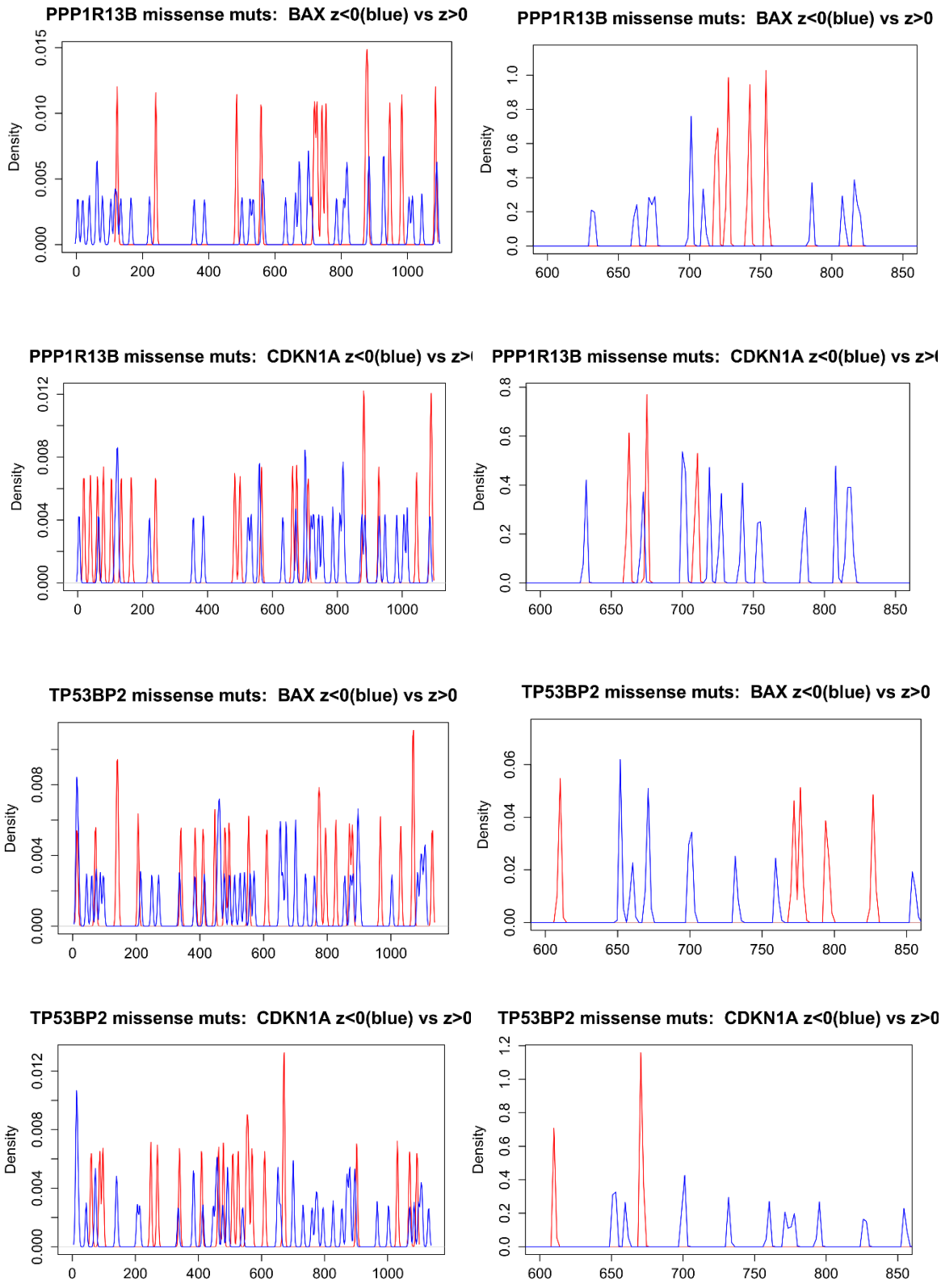
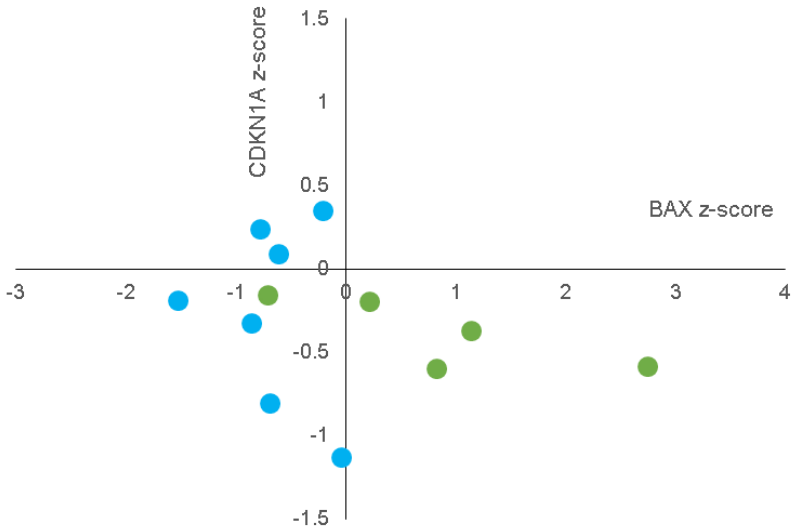


Figure 4.14 *CDKN1A* mRNA levels as a function of position of mutant *PPP1R13B* (top) and *TP53BP2* (bottom) residues with a focus on the Pro-rich region. Blue = z < 0; red = z > 0; left panels show the entire sequence of the proteins; right panels zoomed in on Pro-rich region. Only mutations indexed by cBio were included on this plot.

4.2.4 ASPP1 and ASPP2 mutations might redirect p53 family transcription factors to the *BAX* promoter to inhibit *CDKN1A* transcription

The observations that ASPP1 C-Pro missense mutations are negatively selected in mutant-*TP53* tumours but positively selected in wild-type-*TP53* tumours, and that these C-Pro tumours express above-average levels of *BAX* yet below-average levels of *CDKN1A*, present the interesting possibility that C-Pro ASPP1/2 mutations are positively selected based on their ability to divert the p53-family proteins from the *CDKN1A* promoter. If so, the levels of *BAX* in C-Pro ASPP mutation tumours would be increased at the expense of transcriptional activity at the *CDKN1A* promoter, thus decreasing *CDKN1A* mRNA levels. The relationship between *CKDN1A* and *BAX* z-scores in Pro-mutant tumours in both ASPP1 and ASPP2 support this hypothesis (**Figures 4.15 and 4.16**). There is an overall differentiation between N-Pro (blue) and C-Pro (green) mutant tumours with C-Pro mutations clustering in the *BAX*-positive *CDKN1A*-negative quadrant. The levels of *BAX* are significantly higher in C-Pro vs N-Pro mutant tumours (**Figure 4.16**) whereas *CDKN1A* levels are lower.



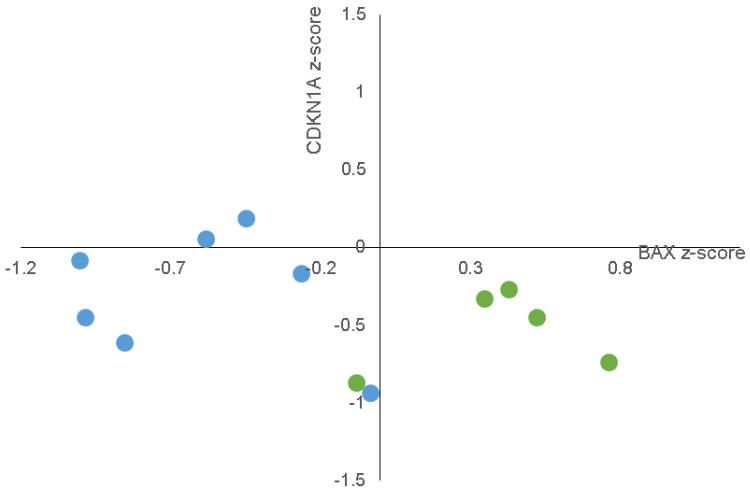


Figure 4.15 *CDKN1A* vs *BAX* z-scores in *PPP1R13B* (top) and *TP53BP2* (bottom) Pro-rich-region mutant tumours. Blue points = mutations in more N-terminal Pro-rich region (600-710 for ASPP1; 650-730 for ASPP2); green points = mutations in more C-terminal Pro-rich region (710-800 for ASPP1; 731-850 for ASPP2).

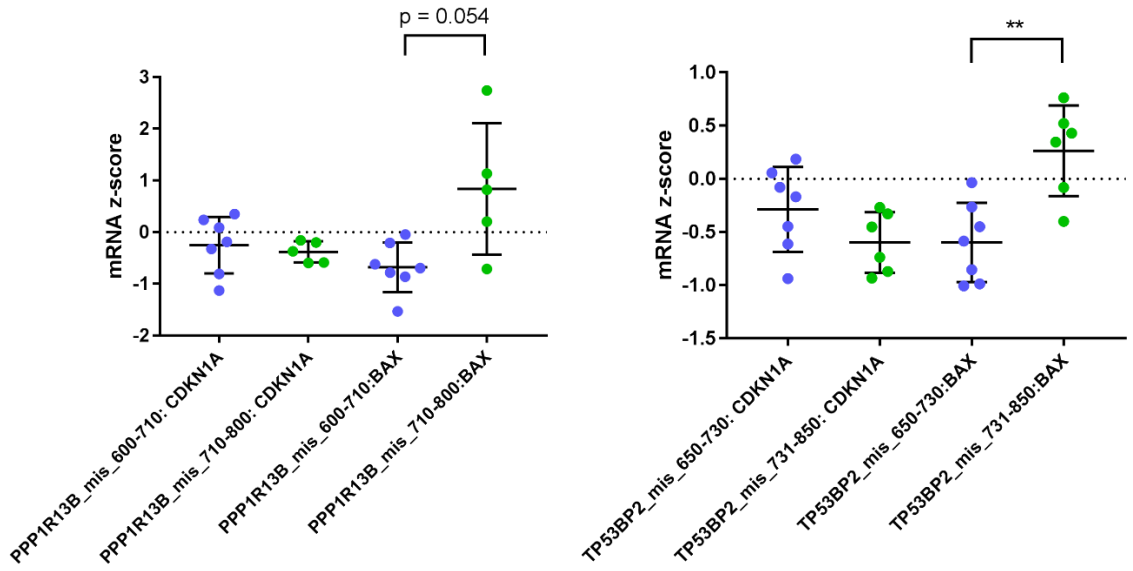


Figure 4.16 *BAX* and *CDKN1A* mRNA z-scores in *PPP1R13B*-mutant (left) and *TP53BP2*-mutant (right) tumours, focused on Pro-rich region mutants. Blue = proline rich segment closer to the N-terminus (N-Pro); green = Pro-rich segment closer to the C-terminus (C-Pro). Student's two-tailed t-test; ** $p < 0.01$.

An outstanding question is why do both N-Pro and C-Pro mutations in ASPP1 occur exclusively in wild type p53 tumours whereas C-Pro mutations in ASPP2 are associated with mutant p53? ASPP1 and ASPP2 both enhance p53-,

p63- and p73-dependent transcription of *BAX*. However, ASPP2 is able to amplify p53-dependent *BAX* transcription to significantly higher fold increases than ASPP1 (**Figure 4.17**) (107). The effect of ASPP1 and ASPP2 on p63- and p73-dependent *BAX* transcription is equivalent within the margin of error. This suggests that there is stronger negative selection against ASPP2 C-Pro mutations in wild type-p53 tumours because the resulting *BAX* levels would be higher than in ASPP1 C-Pro tumours. This is consistent with the fact that ASPP1 C-Pro mutant tumours show a more dramatic increase in *BAX* with a mean z-score of approx. 1 compared with 0.3 for ASPP2 C-Pro mutant tumours (**Figure 4.16**), suggesting the presence of wild type p53 in ASPP1 C-Pro mutant tumours results in a higher magnitude of *BAX* transactivation. Therefore, the net balance between an unfavourable *BAX* increase versus favourable p21 decrease is overall tumorigenic in wild-type p53 tumours with ASPP1 C-Pro mutations, whereas likely to be overall anti-tumorigenic in wild-type p53 tumours with ASPP2 C-Pro mutations that are able to activate p53-*BAX* more robustly (**Figure 4.17**).

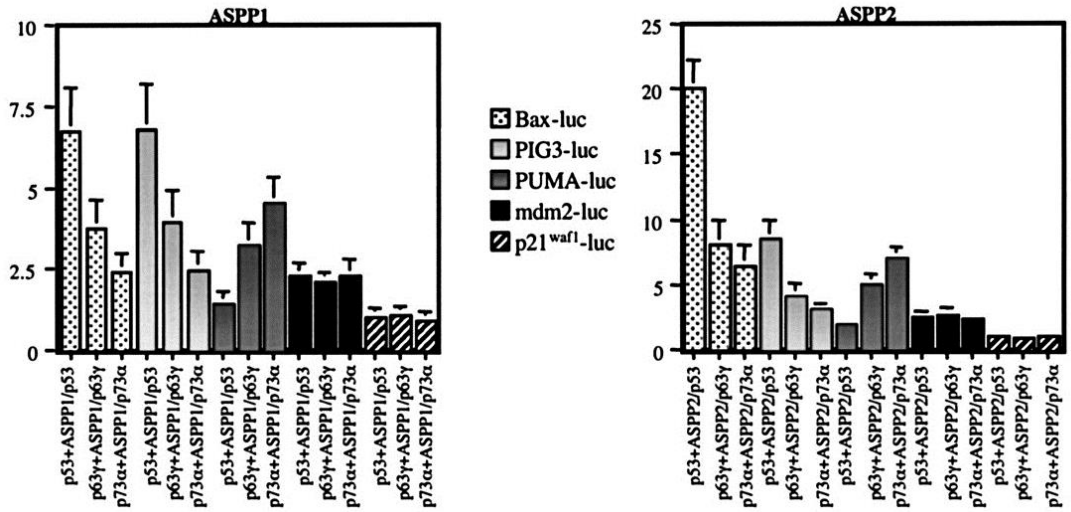
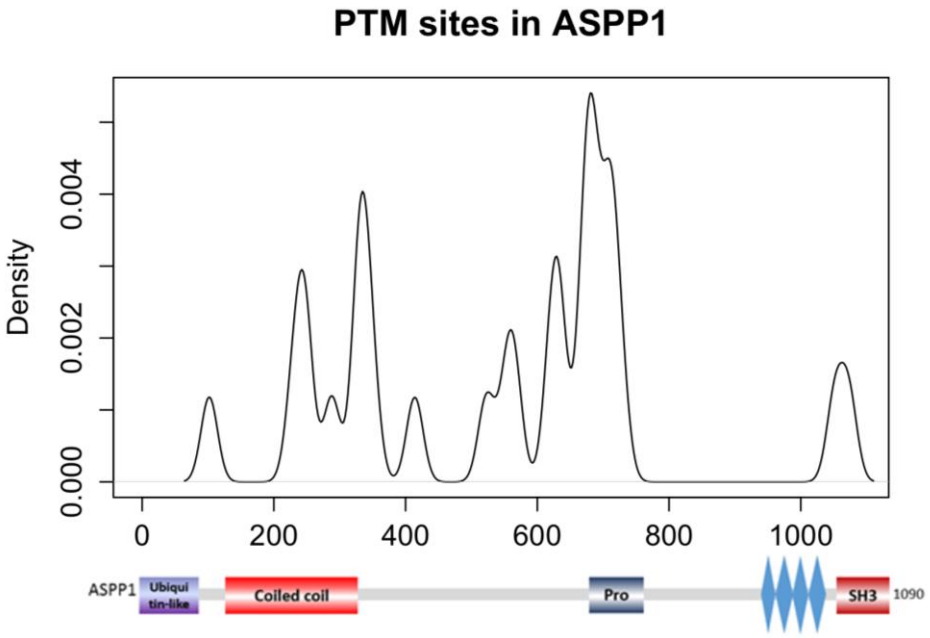


Figure 4.17 ASPP1 and ASPP2 enhance p53-family-dependent transcription of pro-apoptotic target genes. Reproduced from Bergamaschi et al (107) with permission.

Altogether, the results suggest that ASPP1 and ASPP2's ability to selectively recruit the p53-family proteins to pro-apoptotic promoters is hijacked by somatic mutations in order to maintain low levels of the cell cycle inhibitor p21 (C-Pro mutants), or inhibited in order to lower the levels of *BAX* (N-Pro mutants).

4.2.5 Pro-rich region is the top PTM regulatory domain in ASPP1

The observation of selection for missense mutations at the Pro-rich region in both ASPP1 and ASPP2 argues for an important role of this domain in regulating ASPP activity. An important way of regulating protein activity are post-translational modifications (PTMs). We asked whether the Pro-rich region played a role in regulating ASPP function by post-translational modifications. PTM data were retrieved from Phosphosite and the density of post-translationally modified sites plotted as a function of ASPP1 sequence (**Figure 4.18**).



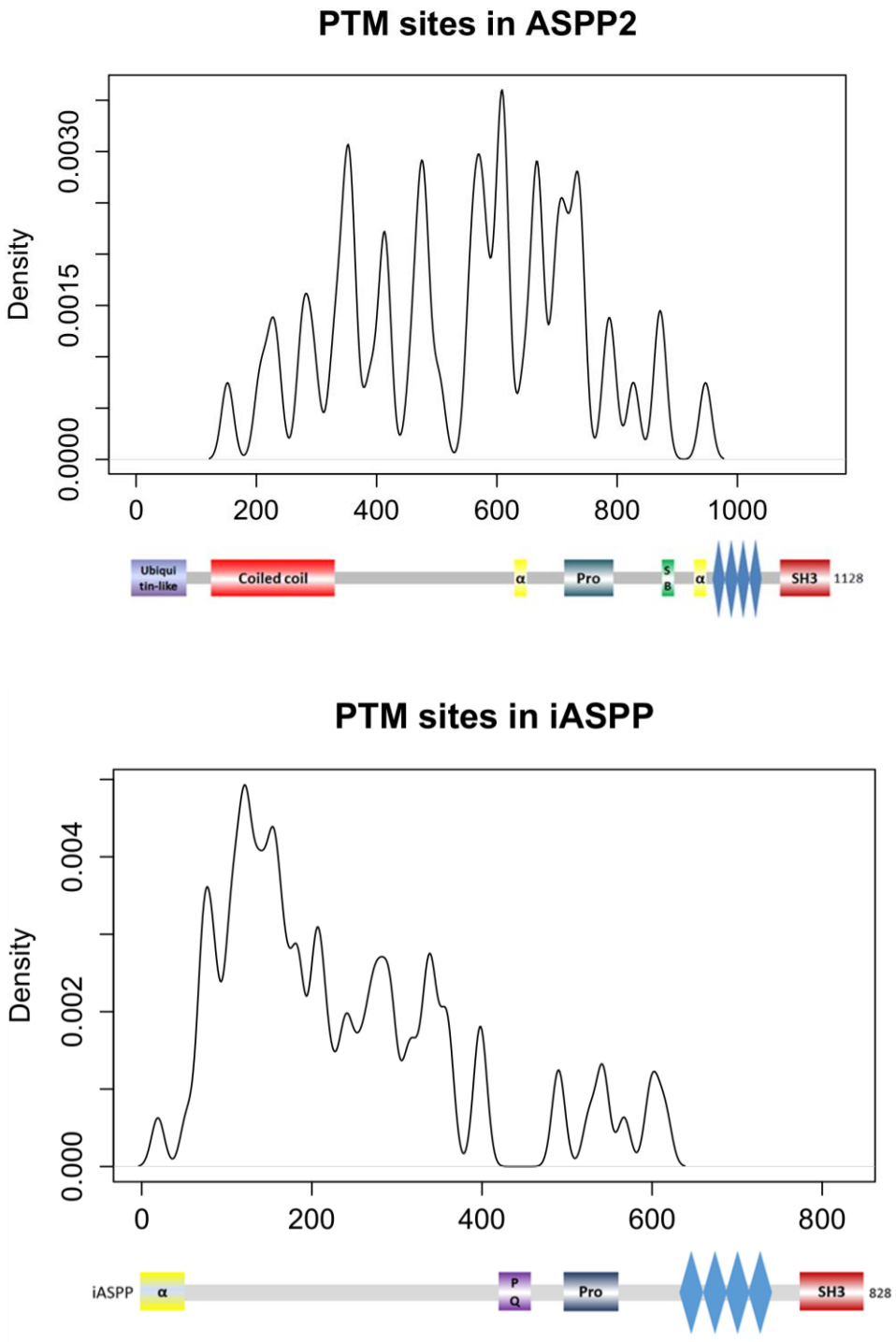


Figure 4.18 Density of distinct post-translational modification sites in ASPP1, ASPP2 and iASPP (top, middle, and bottom, respectively).

The pattern of PTM site distribution among the 3 ASPPs offers an explanation as to the enrichment of candidate driver mutations around and within

the Pro-rich region of ASPP1: this region is the most widely used PTM regulatory domain in the ASPP1 sequence, in contrast to ASPP2 and iASPP that show a broader distribution of PTM sites and a heavily N-terminus-skewed distribution, respectively. These results suggest the Pro-rich region is the dominant regulatory domain of ASPP1 and therefore is preferentially targeted by oncogenic missense mutations to modulate, rather than completely inactivate, ASPP1's function.

The distribution of PTM sites across the Pro-rich region is uneven. In ASPP1 and ASPP2, the N-Pro segment and the preceding less conserved sequence have significantly more PTM sites than the more C-terminal side (C-Pro). Moreover, some of these C-Pro phosphorylation sites are conserved between human and mouse, and widely phosphorylated across multiple tissues, such as ASPP2 S737.

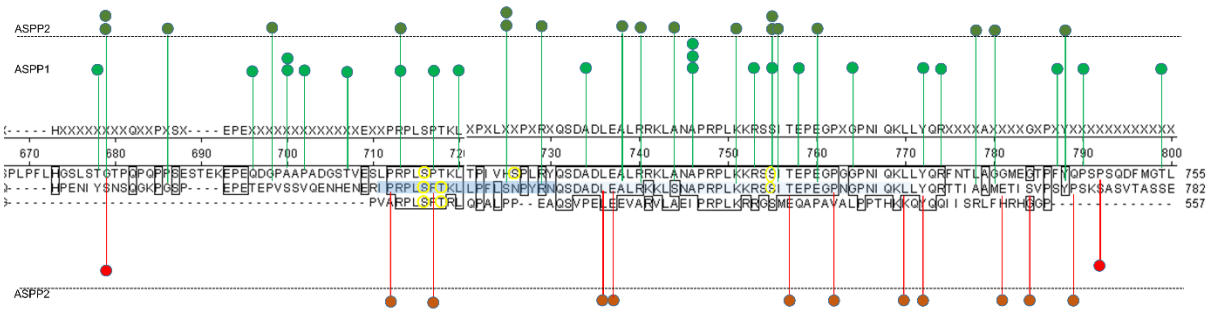


Figure 4.19 Alignment of Pro-rich region in ASPP family genes and the associated missense cancer mutations. Horizontal lines separate *TP53BP2* mutations (inside line) and *PPP1R13B* mutations (outside line). Green points indicate mutations in wild-type *TP53* tumours; red lines mark mutations in mutant-*TP53* tumours. Blue region marks the peptide interacting with Ank-SH3 domain in ASPP2: less transparent blue marks a stronger interaction, more transparent a weaker interaction. Yellow circles mark top 3 phosphorylation sites detected within the region in each protein, in human or mouse samples. The circles indicate the most frequently observed phosphosites in each protein, not an absolute frequency of phosphorylation.

Rotem et al. showed that the Pro-rich region of ASPP2 contains peptides which bind the C-terminus (Ank-SH3 domain) (253). 5 different peptides were able to bind the C-terminus, falling into 2 groups: 1) residues between 693-752 and 2) residues 893-912 (**Figure 4.19**). The strongest signal came from a peptide spanning residues 693-712. In turn, the Ank-SH3 peptides which interacted with the Pro region spanned residues 931-960 and 1083-1096. Interestingly, missense mutations in all of these regions, both in the Pro and Ank-SH3 regions, are uniformly associated with decreased tumour *BAX* levels (**Figure 4.20**). This suggests tumour missense mutations hijack ASPP2's autoregulatory interaction to decrease *BAX* transcription by inhibiting ASPP2's ability to enhance promoter activity of p53-family TFs.

Peptides shown to interact with the Ank-SH3 domain of ASPP2 end at L752. Interestingly, the most N-terminal Pro-rich region residue whose mutation is associated with increased *BAX* levels (C-Pro domain) is Y766. Therefore, the boundary between N-Pro and C-Pro based on the mutants' effects on *BAX* levels coincides with the boundary between the Ank-SH3-interacting domain (N-Pro) and the domain that does not interact (C-Pro) (**Figure 4.19**). This suggests ASPP2 N-Pro mutations modulate the directly interacting residues of N-Pro with Ank-SH3, whereas C-Pro mutations influence this binding indirectly.

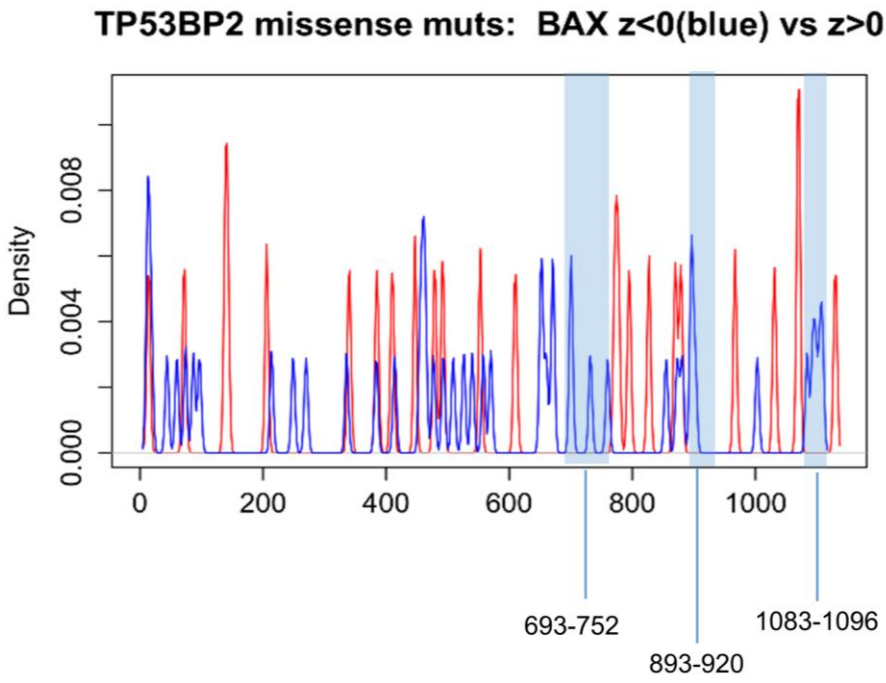


Figure 4.20 Cancer mutations of residues involved in self-binding are associated with decreased *BAX* levels in human tumours. Sequence regions involves in self-interaction in ASPP2 are marked and shaded in blue.

4.2.6 C-terminal mutations in *PPP1R13L* disrupt iASPP's autoregulatory interactions

iASPP is an inhibitor of p53 and contains autoregulatory sequences which are dysregulated by CDK1 phosphorylation in melanoma (257). Therefore we asked whether this autoregulatory mechanism might also be disrupted by missense mutations in cancer. The domain distribution of missense mutations does not show a clear domain-specific interaction with p53 status, however multiple regions are mutated more frequently in wild-type p53 tumours than in p53-mutant tumours (**Figure 4.21**). These regions include residues 80-120, 440-480, 600-640 and 760-818. Interestingly, regions 80-120 and 600-640 contain an identical peptide sequence, SPRKA, that binds the Ank-SH3 domain in the C-terminus and residues 760-818 are involved in this interaction (**Figure 4.21**). Moreover, multiple residues in these regions have been mutated recurrently in

human cancer: R86 (2 counts); G92 (3 counts); H107 (2 counts); S120 (2 counts); A623 (4 counts); G761 (2 counts); R812 (4 counts).

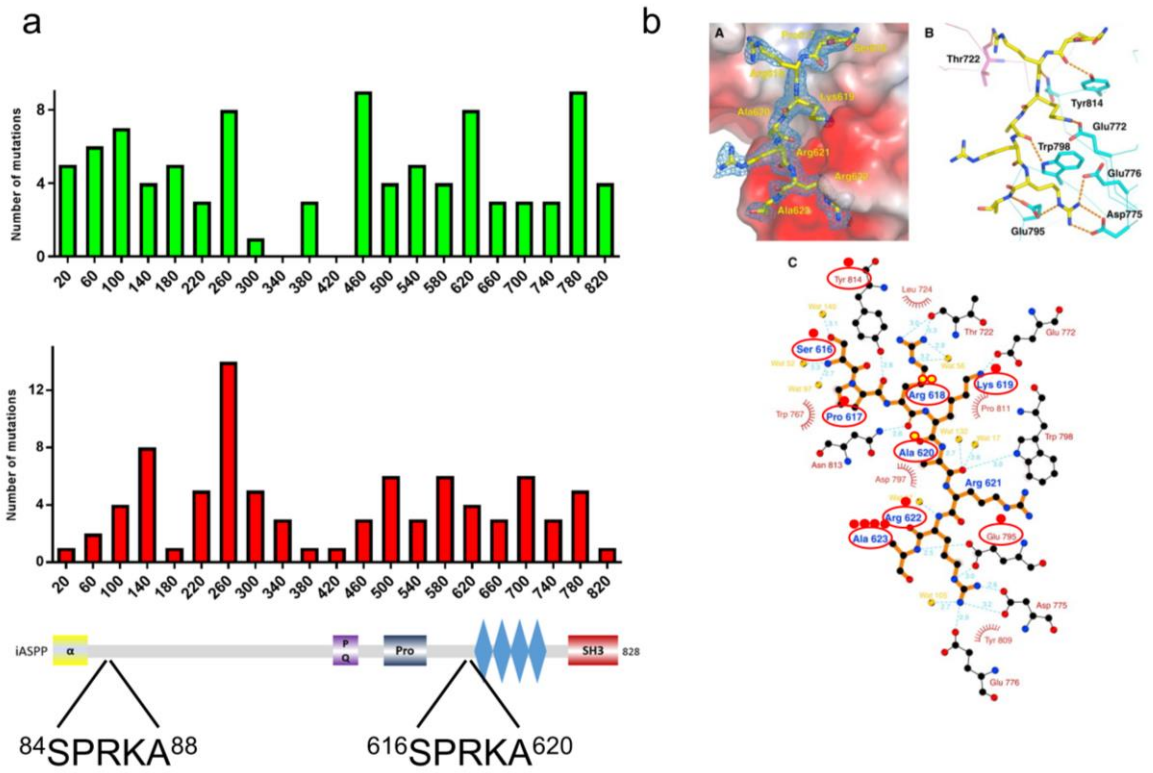


Figure 4.21 iASPP residues preferentially mutated in wild type-p53 tumours are involved in autoregulatory interaction. (a) Distribution of missense mutations in human wild type-p53 (green) vs mutant-p53 (red) tumours; position of the Ank-SH3-interacting peptides shown; (c) iASPP peptide binding Ank-SH3 domain in the C-terminus; reproduced from Robinson et al (224) with permission. Residues marked with red circles are mutated in cancer; dots mark the number of times each residue has been mutated; yellow-filled mutations are mutations in the N-terminal SPRKA region; red-filled mutations mark direct mutations in the residues depicted in the crystal structure.

To investigate the effect of the C-terminal mutation on the interaction with iASPP's N-terminus, we recombinantly expressed C-terminal iASPP fragments and tested their binding to an N-terminal iASPP fragment. Strikingly, each C-terminal mutant showed decreased binding to the wild type-N-terminal fragment of iASPP (**Figure 4.22**). We also observed that most of the mutant residues cluster around the cleft between the ankyrin repeats and the SH3 domain, a site involved in the interaction with the Pro-rich region, with no mutations targeting residues on

the opposite face (**Figure 4.22**). Taken together, the data show missense iASPP mutations enriched in wild type-p53 tumours disrupt iASPP's autoregulatory interaction. The probable function of these mutations is to increase iASPP's inhibition of p53 by making the C-terminus available for binding, which is known to occur in wild type iASPP in melanoma by excessive CDK1 phosphorylation of the N-terminus.

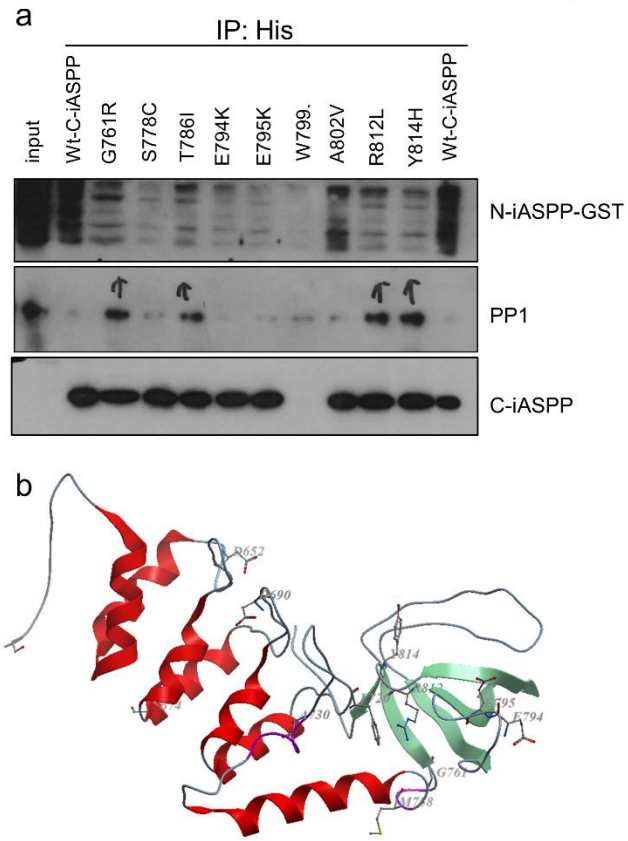


Figure 4.22 iASPP mutations in the C-terminus disrupt interaction with the N-terminus. (a) Immunoprecipitation of recombinantly expressed C-terminal iASPP mutants with N-terminal wild type iASPP fragment (N-GST) and PP1 (middle lane); bottom lane = loading control for C-terminal iASPP. (b) location of C-terminal mutants within the tertiary structure.

The mutation of distinct Pro-rich regions in iASPP, including the SPRKA motifs, alongside the domain conserved between all ASPPs, points to a differential regulation of iASPP function. While most PTM sites in ASPP1 are concentrated in

and prior to the Pro-rich region, in iASPP the majority of sites are located in the N-terminal Pro-rich sequence unique to iASPP (**Figure 4.18**). Therefore, the distribution of driver missense mutations in iASPP reflects the usage of distinct regulatory sequences which are not shared by ASPP1 and ASPP2. However, despite differences in usage of specific domains, the basic paradigm of an autoregulatory intrinsically disordered domain modulating the protein-protein interactions of the ordered C-terminus by self-binding or dimerisation is shared between all 3 ASPPs, and based on the data presented in this chapter, is also targeted by missense mutations in cancer in all 3 ASPP genes.

Discussion

The data presented in section 4.2 demonstrate that ASPP's control of p53/p63/p73 activity is not only observed *in vitro*, but that this interaction shapes cancer genomes in sporadic human tumours. Specifically, we identified a genetic interaction between mutations in the Pro-rich regions of ASPP1, ASPP2 and iASPP with mutations in *TP53* and showed that these missense mutations act by perturbing the autoregulatory interaction between Pro-rich regions and the Ank-SH3 C-terminal domain, to modulate p53/p63/p73-dependent transcription at the pro-apoptotic versus cell cycle inhibitor gene promoters. These mutations seem to be mini-drivers, with evidence of positive selection in cancer but unlikely to have cell-transforming abilities in isolation. Although the absolute number of mutations detected that are likely to act by this mechanism is low, this is only a subset of the ASPP Pro-rich mutations observed in human tumours, limited by the number of tumour samples for which both exome sequencing and RNA sequencing data

were available. Thus, the total number of mutations of this class is higher than the number of dots plotted in **Figure 4.15** (see **Appendix Tables S4-S6** for all manually curated mutations).

Importantly, this methodology can be generalised to discover other mini-driver mutations. The code used to analyse genetic interactions between domain-specific missense mutations in different genes, as well as code used for mutations versus mRNA level analysis, take advantage of public APIs and can be applied to any pair of genes with non-zero mutation rate in cancer. As proof of principle, we queried the interaction between *TP53BP2* and *KRAS* mutations, which revealed a significant bias for *TP53BP2* mutations in the ubiquitin-like domain and N-Pro region of ASPP2 (**Figure 4.23**). This figure was obtained using cBioPortal public API data, demonstrating the sufficient power of the current database release (v1.2.4) to detect low-frequency mutation interactions. This figure provides further insight into the mechanism of action of N-Pro missense mutations in ASPP2. Oncogenic Ras has been shown to cooperate with ASPP2 to promote p53-dependent transcription of pro-apoptotic genes (247). S698A mutation targeting a key conserved residue in the N-Pro region of ASPP2 enhances the efficiency of this process (255). Therefore, N-Pro mutations in wild type-p53 and mutant-*KRAS* tumours are likely to limit p53's pro-apoptotic activity by dysregulating the N-Pro region of ASPP2, perhaps by strengthening the autoregulatory interaction of N-Pro with Ank-SH3, thereby increasing the relative abundance of the closed form of ASPP2 and preventing it from binding p53.

Application of similar methodology will aid in the discovery and functional annotation of rare driver and mini-driver mutations in human cancer.

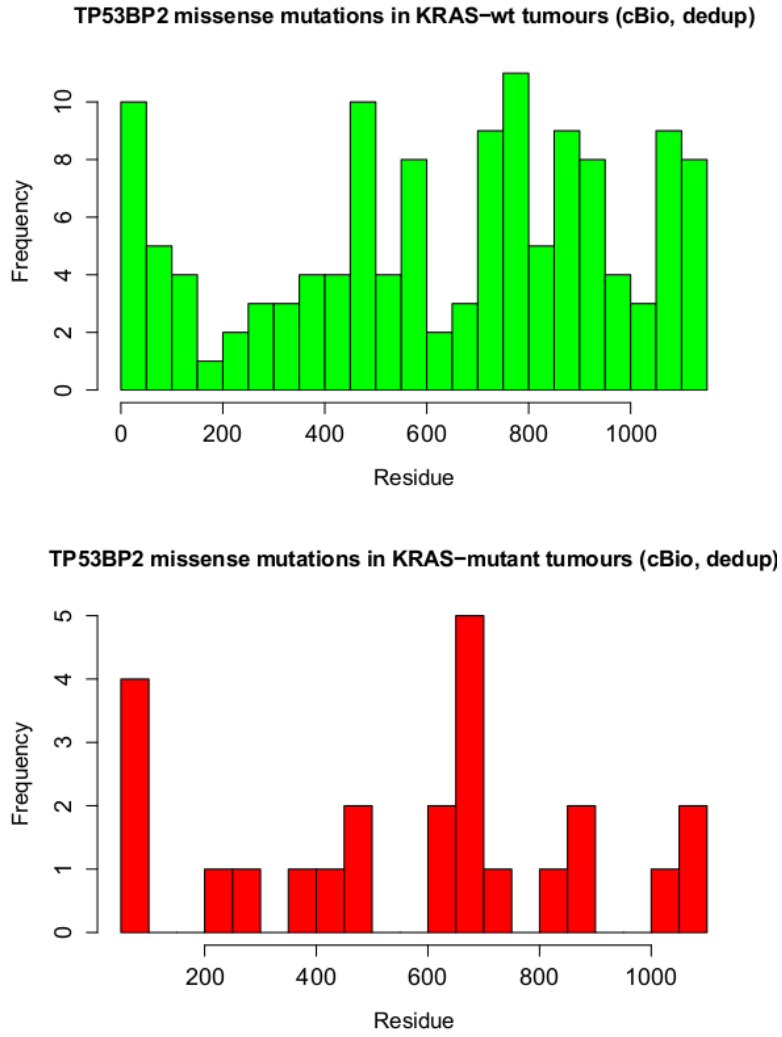


Figure 4.23 Missense *TP53BP2* mutations in wild type- (green) versus mutant-*KRAS* (red) tumours.

Materials and Methods

The results presented in this chapter are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

Mining somatic mutations in cancer

For manual curation of somatic mutations in ASPP genes, data were aggregated from COSMIC, cBioPortal, ICGC Data Portal and NCI Genomic Data Commons (GDC). The database versions used for final thesis figures were COSMIC v78; cBioPortal v1.2.4; ICGC r22 and GDC data release 2.0. Mutation data were parsed to synchronise classification with respect to mutation class, reference transcript, patient identifier, tissue source (primary tumour, metastasis, cell line etc), number of nonsynonymous mutations in sequenced sample, date mutation added and version of source database. Subsequently, duplicate mutations were deleted: double-indexed mutations, mutations from multiple biopsies of the same patient, mutations from different version of the same cell line etc. To control for contamination with legacy mutations that were have been removed from public databases, analyses were performed with all mutations and repeated with a set of mutations currently indexed by the source databases.

Copy number variants

Sample-level copy number data were retrieved from cBioPortal. The origin of TCGA data in cBioPortal is the Broad Firehose (https://gdac.broadinstitute.org/).

Mouse transposon screens

Data on mouse cancer transposon screens (insertional mutagenesis experiments) were mined from the Candidate Cancer Gene Database (http://ccgd-starrlab.oit.umn.edu/about.php) (380).

mRNA levels in tumour vs normal tissue

Sample-level TCGA data on mRNA levels derived from RNAseq were retrieved from the Firebrowse API (http://firebrowse.org/api-docs/) using scripts in the R programming language (352). Log2-transformed gene expression values were used to compare the normal vs tumour samples using a two-tailed Student's t-test. Multiple hypothesis correction was carried out by multiplying the raw p-value by the number of cancer studies tested. The R script is available in the Appendix and was tested in R version 3.3.1. Published microarray studies on tumour vs normal mRNA expression were retrieved from Oncomine (https://www.oncomine.org/) using the online graphical user interface (381). P values and fold changes reported by Oncomine were adopted.

sgRNA screen data

Gene-level data from the ACHILLES CRISPR screen were retrieved from the ACHILLES Data Portal (https://portals.broadinstitute.org/achilles). Following gene search in the Achilles v3.3.8 dataset, file entitled 'sgRNA level scores for [genename]' was downloaded. R programming language was used to plot density of the sgRNA level scores.

Chemosensitivity cell line data

Data on mRNA level correlations with drug sensitivity across cancer cell lines was retrieved from Rees et al (379). Data not included in the Supplementary information was mined through the Cancer Therapeutics Response Portal (http://portals.broadinstitute.org/ctrp/) or through the Cancer Target Discovery and Development (CTD2) Data Portal (https://ctd2.nci.nih.gov/dataPortal/). GraphPad Prism was used to produce some figures.

Genetic interaction between somatic mutations

Sample-level mutation data for a specific gene were retrieved from cBio using the R package CGDSR (https://cran.r-project.org/web/packages/cgdsr/index.html). The mutations were de-duplicated prior to plotting. The associated R script is available in the Appendix.

mRNA levels in mutation-carrying tumours

Sample-level mRNA z-scores were retrieved from cBio using the CGDSR package (see above). The density plots were produced using core function in R v3.3.1. The associated R script is given in the Appendix.

Post-translational modifications

PTM data were retrieved from PhosphoSite (http://www.phosphosite.org/) (382). The density plots were produced using core functions in R.

Cloning and protein expression

Site-directed mutagenesis was performed as described in Chapter 2. To recombinantly express C-terminal iASPP-His fragment protein (wild type and mutants), bacteria were transformed with the expression plasmid, incubated at 37℃ for 16h, then protein expression induced with IPTG and the cultures incubated at ambient temperature at 250 rpm for 16 h. Resulting cultures were spun, supernatant discarded and the cell pellet sonicated on ice for 10-15 min until homogeneous. N-terminal iASPP-GST fusion protein was expressed following the same protocol.

Crystal structure visualisation

Mutated iASPP residues were visualised using Molsoft ICM Browser (Molsoft; San Diego, CA).

CHAPTER 5: OVERCOMING T-CELL EXHAUSTION BY SELECTIVE EXPANSION OF FOLLICULAR CYTOTOXIC T CELLS

RESULTS AND DISCUSSION

5.1 Exhausted effector phenotype of virus-specific CD8+ T cells in chronic LCMV infection

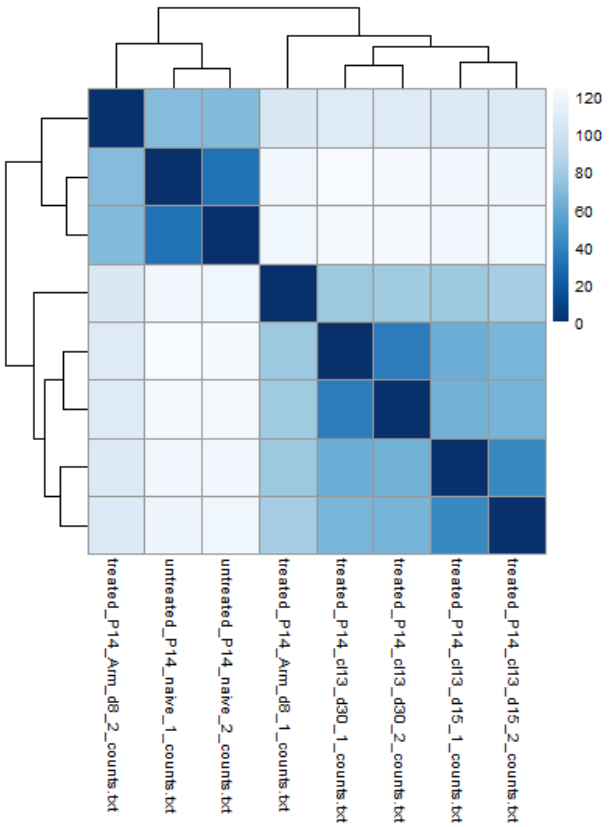


Figure 5.1 Sample similarity matrix for P14 cells based on gene expression
Hierarchical clustering follows similarity coefficients based on rld-transformed counts using DESeq2 package.

The signalling circuitry underlying dysfunctional T cell responses is still incompletely understood. Wherry et al. described the transcriptional profile of virus-specific CD8+ and CD4+ T cells in chronic LCMV infection (383, 384). However, these studies used microarrays as a means to determine gene expression levels, limiting the detection of certain genes. Moreover, tetramer reagent was used to isolate endogenous virus-specific T cells which might lead to minor contamination because tetramer staining often shows limited separation of

positive from negative cell populations. In order to gain a deeper understanding of the transcriptome of T cells specific for the immunodominant epitope, we used CD8+ T cells with a transgenic TCR specific for the gp33 epitope, which gives rise to the most abundant pool of LCMV-specific CD8+ T cells in B6 mice. A mouse line with a transgenic TCR specific gp33 was used as a source of antigen-specific CD8+ T cells (385). These mice, known as P14, were crossed with Thy1.1+ mice to allow for the specific identification and sorting of $P14^+$ $Thy1.1^+$ cells following adoptive transfer, as wild type C57BL/6 mice lack Thy1.1. Thy1.1+ P14 cells can be readily identified by flow cytometry and show a large separation from Thy1.1- populations, minimising contamination with other cell types.

To obtain 4 sample groups of P14 cells in uninfected, acutely infected (Arm 8 d. p. i.) and chronically infected (cl13 15 and 30 d. p. i.), wild type C57BL/6 mice were injected with 2,000 P14 Thy1.1+ cells and infected with vehicle, LCMV Armstrong or clone 13 the following day. At day 8 post infection for Armstrong and days 15 and 30 for clone 13, splenocytes were harvested, stimulated with gp33 peptide for 2 h, and then subjected to total T cell isolation and cell sorting to isolate P14 Thy1.1+ cells. RNA was isolated from these cells and duplicated from each sample group were subjected to next-generation sequencing. The sample groups were abbreviated P14_naive (uninfected mice), P14_Arm_d8 (mice infected with Armstrong, 8 d. p. i.), P14_cl13_d15 and P14_cl13_d30 (mice infected with clone 13, 15 and 30 d. p. i., respectively).

Transcriptomes of all 8 samples were first compared for similarity using DESeq2. Replicates of three of the four sample groups clustered together but replicates for P14_Arm_d8 clustered independently of each other, suggesting a biological difference in the course of infection (**Figure 5.1**). One likely source of

this variability is incomplete T cell priming by LCMV Armstrong in some individuals. Therefore, additional replicates will be required to increase the robustness of this study and work is in progress to expand the sample number for both groups.

Differential expression analysis of P14 in chronic vs acute infection showed extensive overlap with the earlier study by Wherry et al. (**Figure 5.2**) (383). *Pdcd1,* which encodes the inhibitory receptor PD1, was the most prominently upregulated gene on P14 cells at d30 of chronic cl13 infection (**Figure 5.2**), along with inhibitory receptors *Lag3* and *Havcr2* (Tim3), but also chemokines *Ccl3*, *Ccl4, Ccl9*; and the cytokine *Ifng*. These data are consistent with the notion that exhausted CD8+ T cells first undergo priming and effector differentiation before losing cytokine production and upregulating very high levels of inhibitory receptors, hallmarks of the T cell exhaustion state.

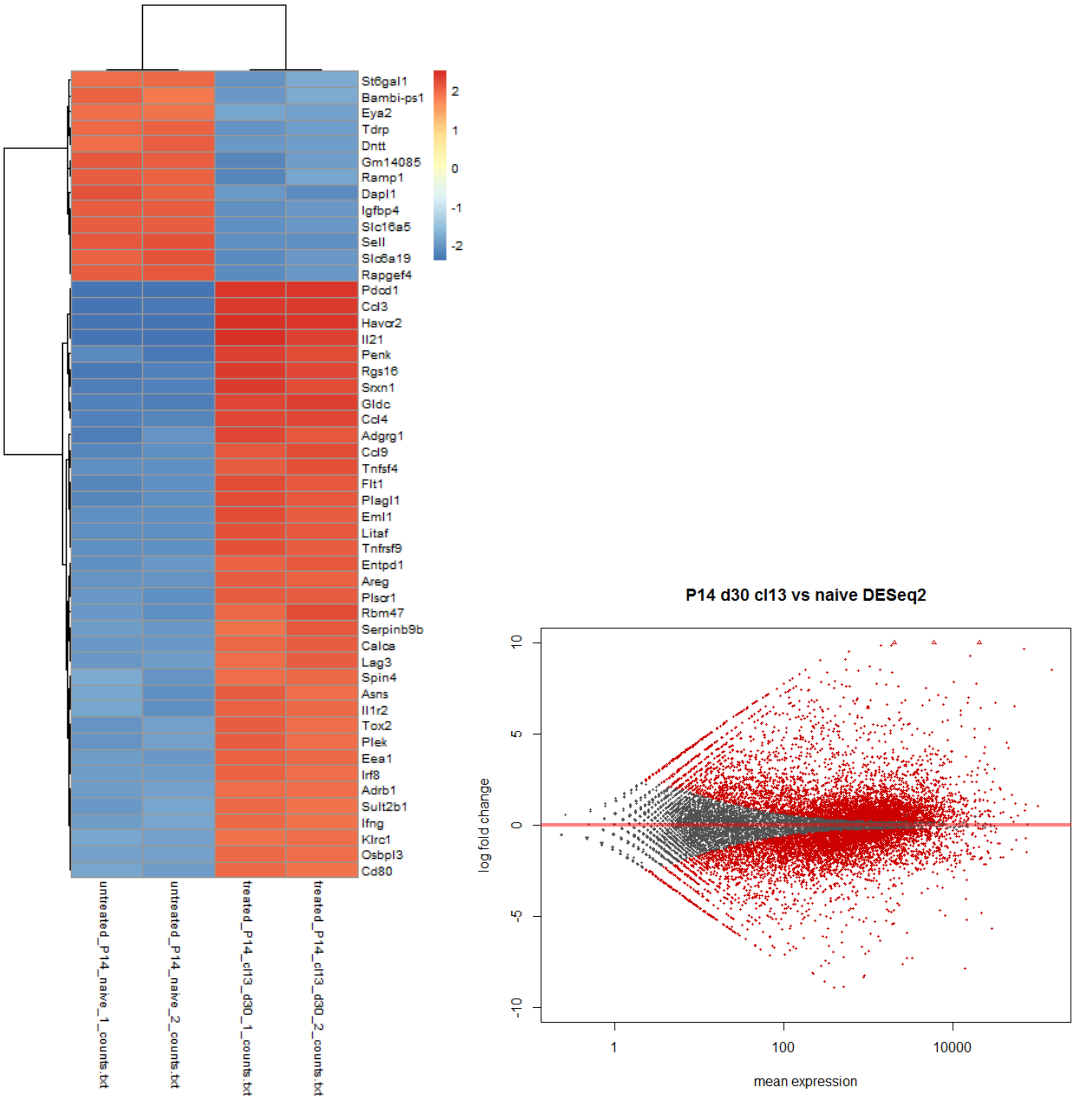


Figure 5.2 Differential expression analysis of P14 cells in mice infected with cl13 for 30 days vs uninfected mice. Left: heatmap of top differentially expressed genes by variance; right: log2fold change vs mean expression level for all non-zero expressed genes; significantly expressed genes marked in red.

5.2 Reduced absolute and relative numbers of terminally differentiated effector CD8+ T cells in chronically infected mice

To determine the phenotypic composition of the splenic CD8+ cell population in chronic vs acute infection, CD8+ cells from uninfected mice, mice infected with LCMV Arm for 8 days, and mice infected with LCMV cl13 for 30 days were sorted, stimulated with peptide of the 3 major immunodominant MHC class I epitopes (gp33, gp276 and np396), and their RNA isolated and sequenced. We

used deep sequencing (ca. 60 million reads per sample) to detect both major and minor populations of CD8+ cells.

As expected, the spleens of acutely infected mice contained significantly higher numbers of cells including CD8+ cells than uninfected or chronically infected spleens. RNA sequencing revealed that effector-associated genes dominated the transcriptional landscape of CD8+ cells in acute infection, with interferon gamma (*Ifng*) being the second most highly expressed gene by absolute level (the first one being *Eef1a1*). High levels of chemokines and effector molecules such as *Gzma* and *Gzmb* were also observed, consistent with an activated effector phenotype dominating the CD8+ population during acute infection.

Table 5.1 Sequencing reads obtained and mapping efficiency for total CD8+ cell samples (sequenced using paired-end 2x150 bp reads).

Sample	No. of reads	% mapped uniquely (mm10)
N1	57,282,531	86,68
N2	56,760,138	87.51
Arm1	64,322,597	89.79
Arm2	63,714,021	89.81
Cl13-1	71,119,014	89.83
Cl13-2	63,096,875	83.75
Mean	62,715,863	88.14

Similar to the antigen-specific P14 cells during chronic infection, differential expression analysis of CD8+ cells in chronically infected vs uninfected mice identified *Pdcd1*, as the most significantly upregulated gene in chronic infection (adjusted $p<10^{-250}$). This is consistent with the fact that although PD1 levels increase upon T cell activation, very high sustained levels of PD1 are a hallmark of T cell exhaustion (**Figure 5.3**). Other upregulated genes include interferon-

induced GTPases *Iigp1*, *Gbp2b*, *Gbp2*, *Gbp3*, *Igtp*, *Iigp1* and others, which are known to dominate the transcriptome of cells stimulated with interferon gamma; chemokines *Cxcl10*, *Ccl4*; effector molecules *Gzma*, *Gzmb*, *Gzmk*; inhibitory receptor *Tigit*; inhibitory ligand *Cd274* (encodes PD-L1); enzyme *Odc1* (ornithine decarboxylase, a gene associated with cell transformation); pro-apoptotic genes *Bcl2l11*, *Fasl* and transcription factors *Ar*, *Plagl1*, *Irf4*, *Zhx2* and *Zeb1*. These results show that CD8+ cells in chronically infected mice contain effector-differentiated T cells but that those cells express inhibitory receptors and pro-apoptotic genes, indicative of T cell exhaustion.

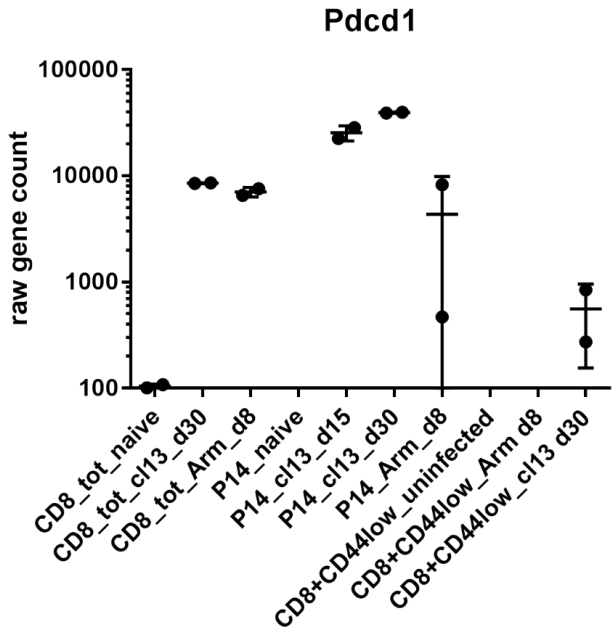


Figure 5.3 *Pdcd1* mRNA levels across CD8+ total, virus-specific and CD8+ CD44low cells in uninfected, acutely infected and chronically infected mice.

To compare the CD8+ cell compartments between chronically and acutely infected mice, we calculated differential gene expression between cl13 d30 and Arm d8 samples. Surprisingly, *Pdcd1* was not differentially expressed between these two groups (**Figure 5.3**). Given that virus-specific exhausted CD8+ T cells

express very high levels of *Pdcd1*, this result suggests the relative concentration of effector CD8+ T cells in the chronically infected spleen is significantly reduced compared with the acutely infected mice. Indeed, the second most significantly upregulated gene in chronically vs acutely infected mouse-derived CD8+ cells was *Sell*, encoding L-selectin, a marker of naïve and memory CD8+ T cells. Therefore, chronically infected mice not only possess reduced numbers of splenocytes compared to acutely infected and uninfected mice, but they also show a lower percentage of effector-differentiated CD8+ T cells within the splenic CD8+ compartment.

Significantly downregulated genes in chronic vs acute infection include chemokines *Ccl1*, *Ccl4*; effector molecules *Gzma*, *Gzmb, Gzmc*; cytokines *Ifng*, *Il2*; regulators of T-cell activation *Cd44*, *Cd5*, *Cd80*, *Cd52*, *Ctla4*. This is consistent with a decrease in the percentage of effector CD8+ T cells in chronic infection and the onset of exhaustion.

5.3 Distinct transcriptional response to IFN-γ in effector vs naïve CD8+ T cells

The observed increase in interferon-induced GTPases in chronic infection despite the lower levels of *Ifng* presents a potential paradox. In mice, cell-autonomous defence GTPases of the GBP and IRG subfamilies dominate the transcriptional response to IFN-γ, accounting for over 30% of the genes induced by IFN-γ in mouse primary cells (386). IFN-γ response is mediated by the IFNGR receptors 1 and 2, encoded by *Ifngr1* and *Ifngr2* in mice. Whereas most mammalian cell lineages express these receptors, effector-differentiated CD8+ T cells lack *Ifngr2* transcripts (**Figure 5.4**). This suggests IFN-γ cannot signal

through the canonical heterodimer of IFNGR1 and IFNGR2 which is classically required for response to IFN-γ (387). IFNGR2, sometimes called IFNGR-β, is also absent in T_H1-polarised CD4+ T cells which are insensitive to the anti-proliferative effect of IFN-γ on T_H2-polarised CD4+ T cells (388). Therefore, the lack of robust transcription of IFN-γ-induced GTPases in effector and exhausted CD8+ T cells might reflect the low-to-no levels of IFNGR2 in these cells. In contrast, naïve CD8+ T cells express *Ifngr1* and *Ifngr2* and upregulated IFN-γ-induced GTPases upon viral infection. Given that spleens in chronically infected mice contain a higher percentage of naïve CD8+ T cells than spleens of acutely infected mice, the observed IFN-γ target gene transcripts might originate from the naïve CD8+ T cells exposed to IFN-γ which is being secreted by the remaining population of incompletely exhausted effector T cells. Alternatively, the observation of these interferon stimulated genes might reflect a shift in STAT signalling in the process of T cell exhaustion. In CD4+ T cells, absence of transcription factor T-bet causes an aberrant overlap between genes induced by type I IFN and IFN-γ (389). Indeed, GTPase genes including *Gbp2* and *Iigp1* can also be induced by IFN-β and mRNA levels of T-bet were reduced in cl13-infected versus Arm-infected mice (**Figure 5.4**). It is possible that the lack of a repressor such as T-bet in exhausted CD8+ T cells leads to the expression of type I IFN targets through activation of IFNGR.

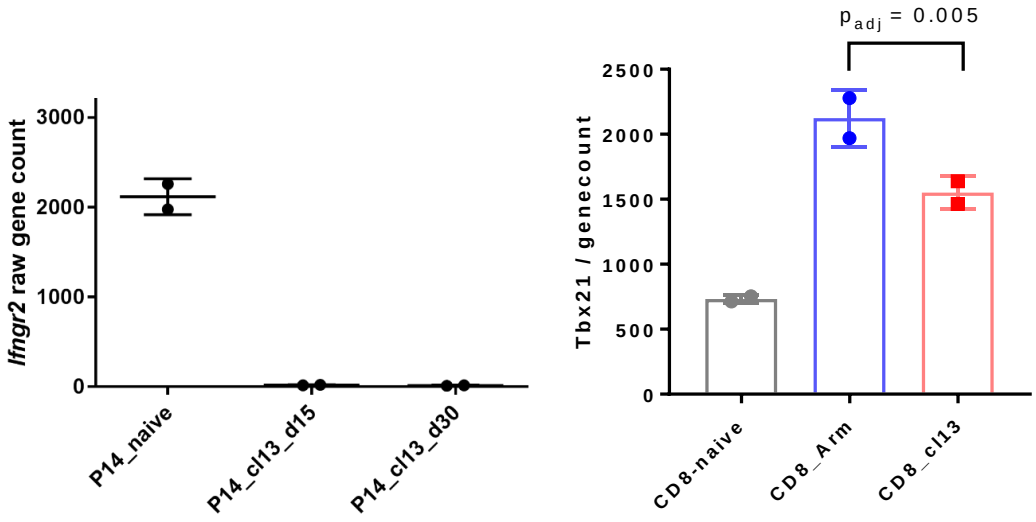


Figure 5.4 Lack of *Ifngr2* mRNA expression in antigen-experienced CD8+ T cells and reduced expression of *Tbx21* in CD8+ T cells from chronically infected mice.

5.4 Non-coding transcription at the immunoglobulin locus in a subtype of CD8+ cells enriched in chronic infection

Following the transcriptomic characterisation of LCMV gp33-specific (P14) and total CD8+ T cells in uninfected, acutely infected and chronically infected mice, we asked how the transcriptional landscape of naive CD44^{low} CD8+ T cells differed between chronic and acute infection. Sample clustering produced 4 distinct clusters with all CD8+ CD44^{low} cells forming one cluster, infected CD8+ total samples forming another cluster, all but one P14 sample from infected mice forming a third cluster, and naïve P14 with P14_Arm_d8_2 forming a small cluster between the other P14 samples and CD8+ total samples. This is consistent with the fact that CD8+ total samples contain both CD8+ CD44^{low} cells and effector cells which are almost invariably CD44-high.

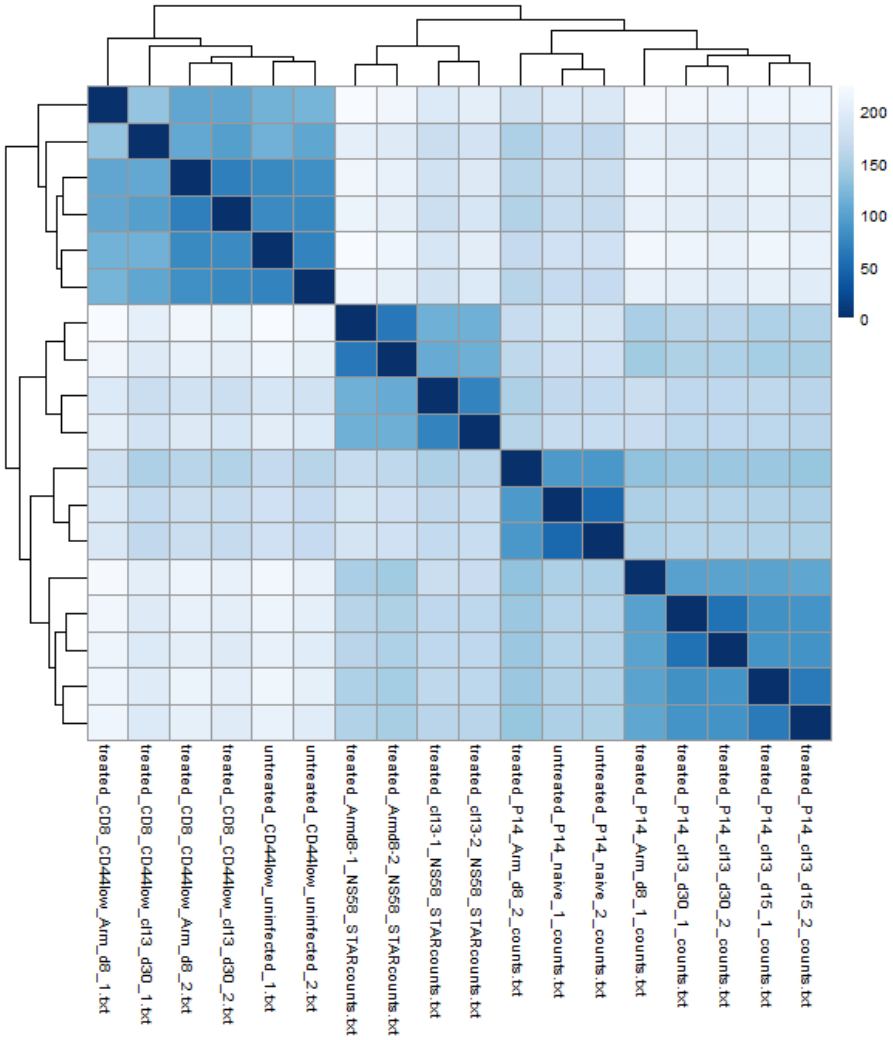


Figure 5.5 Transcriptomes of CD44low cells, P14 cells and CD8total samples cluster by group (CD8tot naïve not shown). Sample distance matrix computed from gene counts in DESeq2.

Hierarchical clustering of samples based on the 50 top-variance genes surprisingly revealed a cluster of immunoglobulin genes as those that differentiate CD8+ total cells in chronically infected mice (cl13 d30) from P14 cells and CD8+ $CD44^{low}$ cells regardless of infection status (**Figure 5.6**). These immunoglobulin genes feature both constant and variable fragments of the heavy chain with some minor contribution from the kappa light chain.

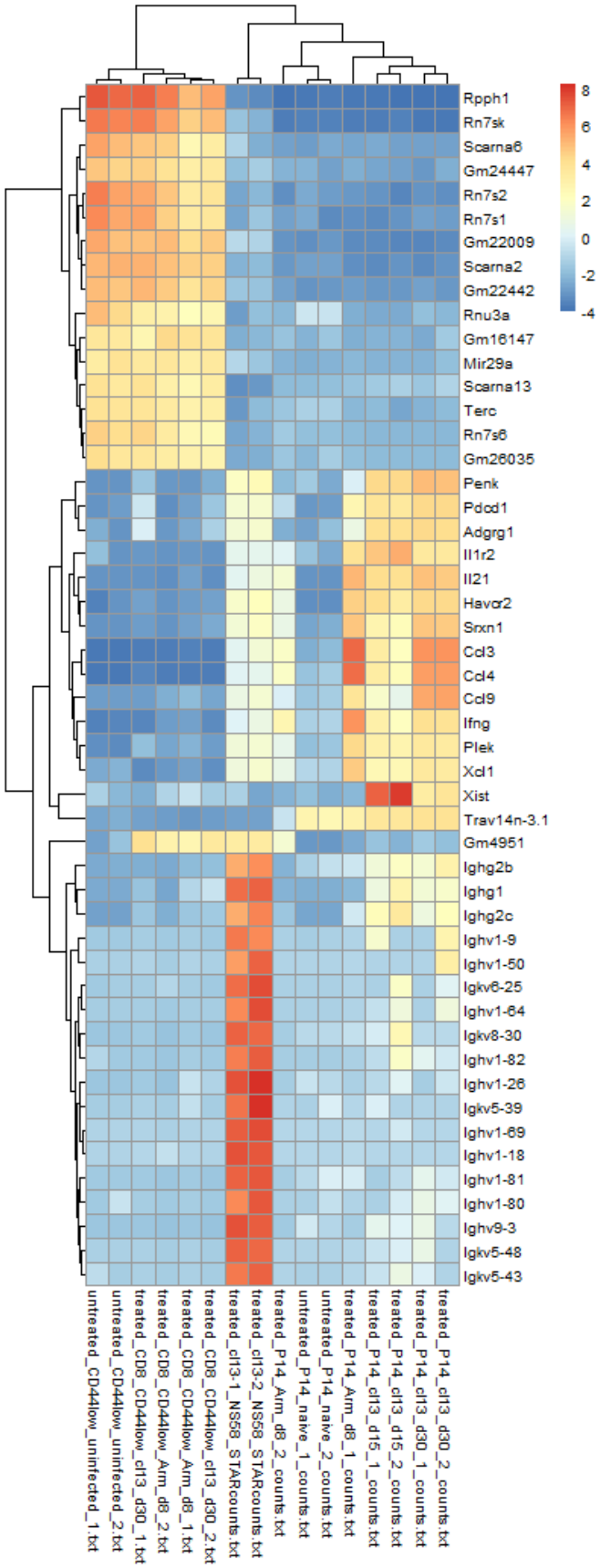


Figure 5.6 A cluster of immunoglobulin genes distinguishes total CD8+ cells in chronically infected mice from other samples.

It is known that sterile transcription from non-coding loci can occur within the immunoglobulin cluster in mice but this is thought to be limited to B cells prior to undergoing recombination in the bone marrow (390, 391). Mature B cells carry recombined Ig loci and express protein-coding antibody genes. Therefore, one way to rule out B-cell contamination as the source of these transcripts is to determine the presence of sterile transcripts including transcription from intergenic regions that are removed in the process of recombination.

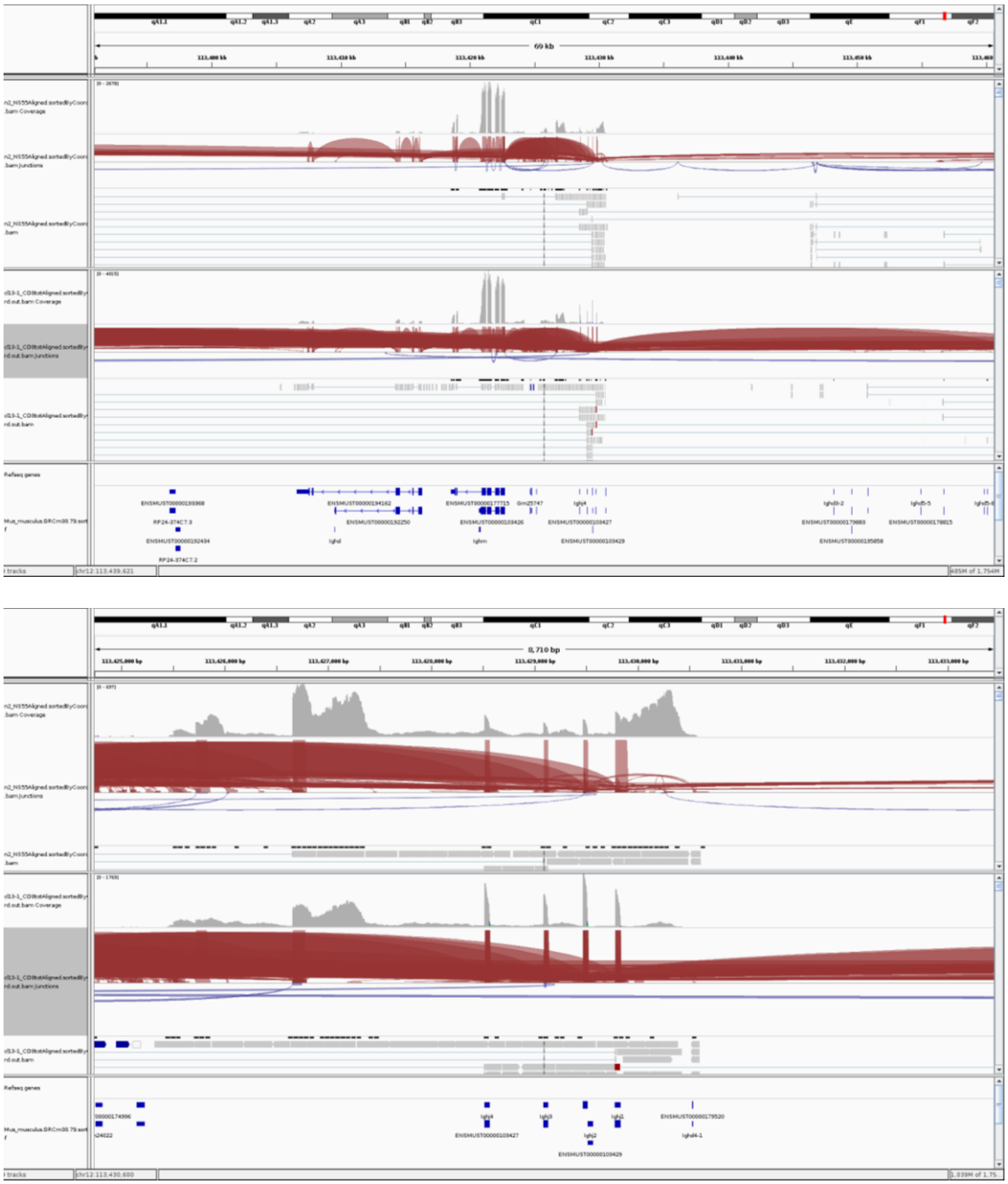


Figure 5.7 Presence of sterile heavy chain immunoglobulin transcripts in CD8+ splenocytes. Top: mapped reads in naïve (top track) and cl13_d30 CD8+ cells; view of the constant mu exons, constant delta exons, J segments and several D segments. Bottom: close-up view of the J segments and the surrounding non-coding regions. The region distal to the J1 segment is lost during somatic recombination in B cells.

Examining mapped reads surrounding the J segments of the heavy chain locus revealed extensive transcription in the distal region prior to the J1 segment. This non-coding region is lost during D to J recombination in B cell development which occurs on both alleles. Therefore, splenic B cells cannot express transcripts originating in this region and the observed reads serve as a marker of sterile transcription on un-recombined heavy chain immunoglobulin locus. Whether transcripts originating from recombined *Igh* loci exist in CD8+ T cells remains to be determined and will require more stringent negative controls as protein-coding *Igh* transcripts can originate from splenic B cells. These results demonstrate that CD8+ splenic cells exhibit sterile transcription at the heavy chain locus and the relative abundance of these reads within the CD8+ pool increases in chronic infection. Whether the observed sterile transcription of immunoglobulin intron sequence plays a functional role in T cells remains to be determined.

5.5 Follicular cytotoxic T cells (T_{FC}) selectively expand following IFNAR blockade

In conjunction with the presence of sterile immunoglobulin transcripts, we observed an upregulation of genes traditionally associated with follicular helper CD4+ T cells (T_{FH}) and germinal centre responses such as *Bcl6* and *Zeb1* (**Figure 5.8**). Interestingly, a novel subtype of CD8+ T cells characterised by the

expression of T_{FH}-associated markers *Tcf7*, *Cxcr5* and *Bcl6* has recently been discovered and named follicular cytotoxic T cells (T_{FC}) (214, 392, 393). Examination of our data revealed that genes which distinguish the CXCR5^{+} T_{FC} population from CXCR5^{-} CD8+ T cells are all upregulated in CD8+ T cells in chronically infected mice when compared to acutely infected mice (**Figure 5.8**). This suggests chronic LCMV infection is associated with an increase in the relative number of T_{FC} within the CD8+ compartment.

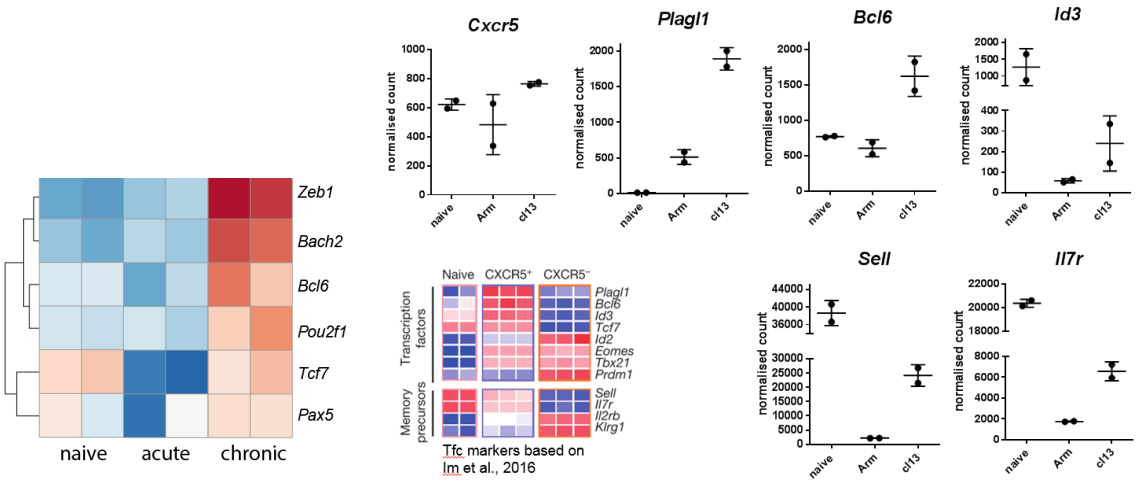


Figure 5.8 Upregulation of T_{FC}-associated genes in CD8+ T cells from chronically infected mice.

Given that antibody-mediated IFNAR blockade significantly accelerates clearance of LCMV clone 13 and is associated with increased germinal centre responses (394, 395), we wondered how T_{FC} are affected by IFNAR blockade. Wild type mice were treated with anti-IFNAR1 antibody or isotype control, followed by clone 13 infection and adoptive transfer of 2,000 virus-specific Thy1.1^{+} P14 cells. At day 10 post infection, Thy1.1+ P14 cells were sorted, and their RNA subjected to RNA-sequencing. In a principal component analysis the top two principal components separated naïve, isotype-treated and anti-IFNAR1 treated

cells into distinct clusters (**Figure 5.9A**). Interestingly, the variance separating naïve P14s from the two infected groups largely followed principal component 1 (PC1), a component orthogonal to the bulk of variance separating isotype and anti-IFNAR1-treated cells (parallel to PC2). Consistently, marker genes of T cell activation such as *Cd44* were comparably expressed in the infected P14 groups compared to naive (**Figure 5.10**). This suggests P14 cells were activated in both anti-IFNAR1 and vehicle treated mice but developed distinct gene expression patterns.

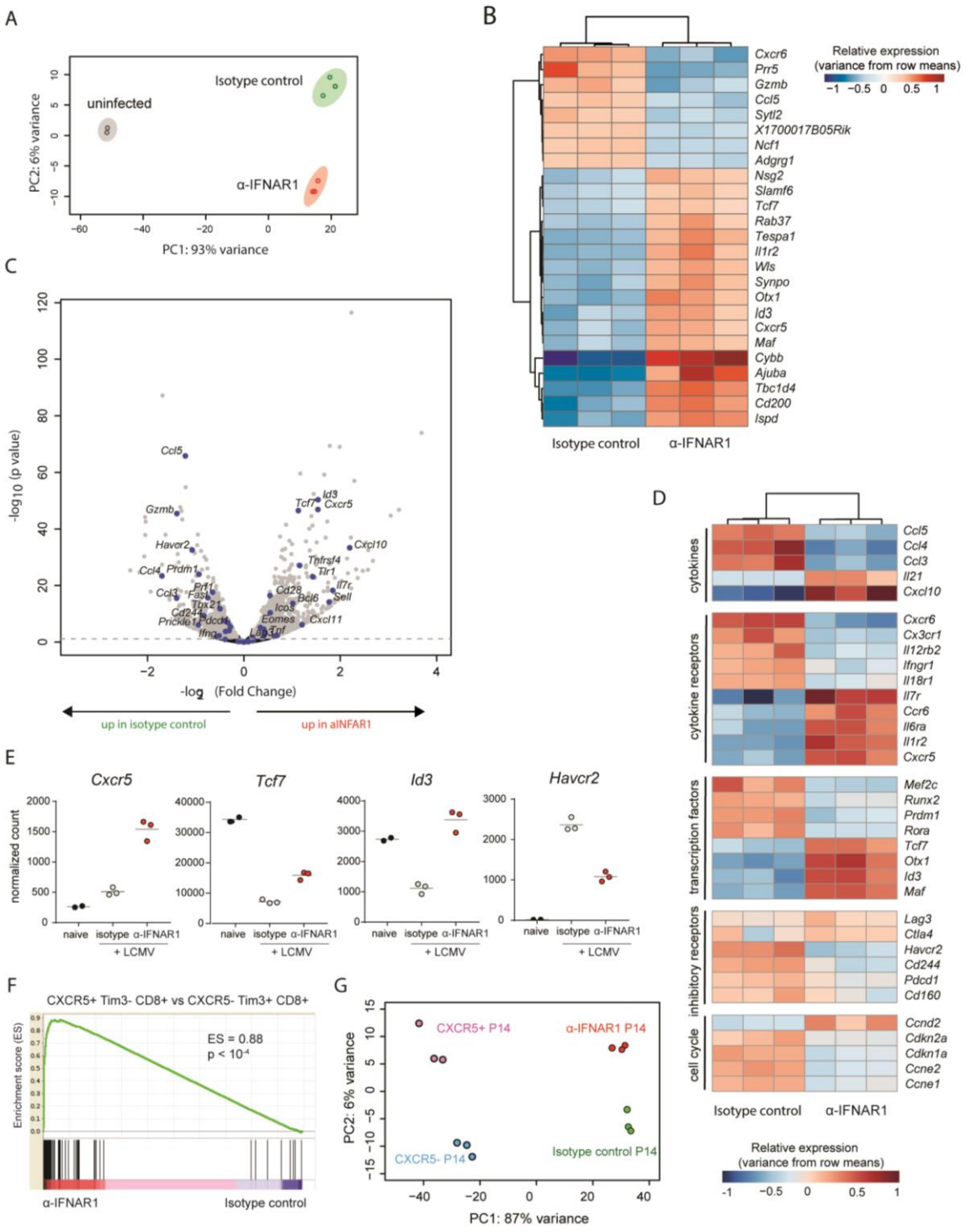


Figure 5.9 Blockade of IFNAR significantly perturbs the transcriptome of virus-specific CD8+ T cells toward the T_{FC} phenotype. P14 cells were transferred into isotype or aIFNAR1 treated mice, followed by LCMV CL-13 infection the next day. P14 cells were then sorted at day 10 post infection and subjected to RNA-seq analysis. (A) Principal components analysis of naïve, isotype-treated, or aIFNAR1-treated P14 cells separates P14 cells by treatment group. (B) Relative gene expression of top 25 genes with the highest variance between isotype- or aIFNAR1-treated P14 cells. (C) Volcano plot showing statistical significance (p value) versus fold increase of gene expression comparing aIFNAR1-treated P14 cells to isotype-treated control. Left half: downregulated

with aIFNAR1 treatment; Right half: upregulated with aIFNAR1 treatment. Genes with important function in T cells were highlighted. (D) Heat map of differentially expressed genes with key functions in CD8 T cells separated by functional class. (E) Normalized counts of *Cxcr5*, *Tcf7*, *Id3*, and *Havcr2* in P14 cells from uninfected, and LCMV Cl-13 infected plus isotype control- or aIFNAR1-treated groups. (F) Gene set enrichment analysis demonstrating a significant enrichment of T_{FC} gene signature within P14 of the aIFNAR1-treated group but not the isotype-treated controls. T_{FC} gene set was constructed from publicly available dataset (GSE84105) by taking top 100 genes most significantly upregulated in CXCR5^{+} Tim3^{-} CD8+ cells versus CXCR5^{-} Tim3^{+} CD8+ T cells. Statistical significance was determined using DESeq2 and GSEA (see Materials and Methods).

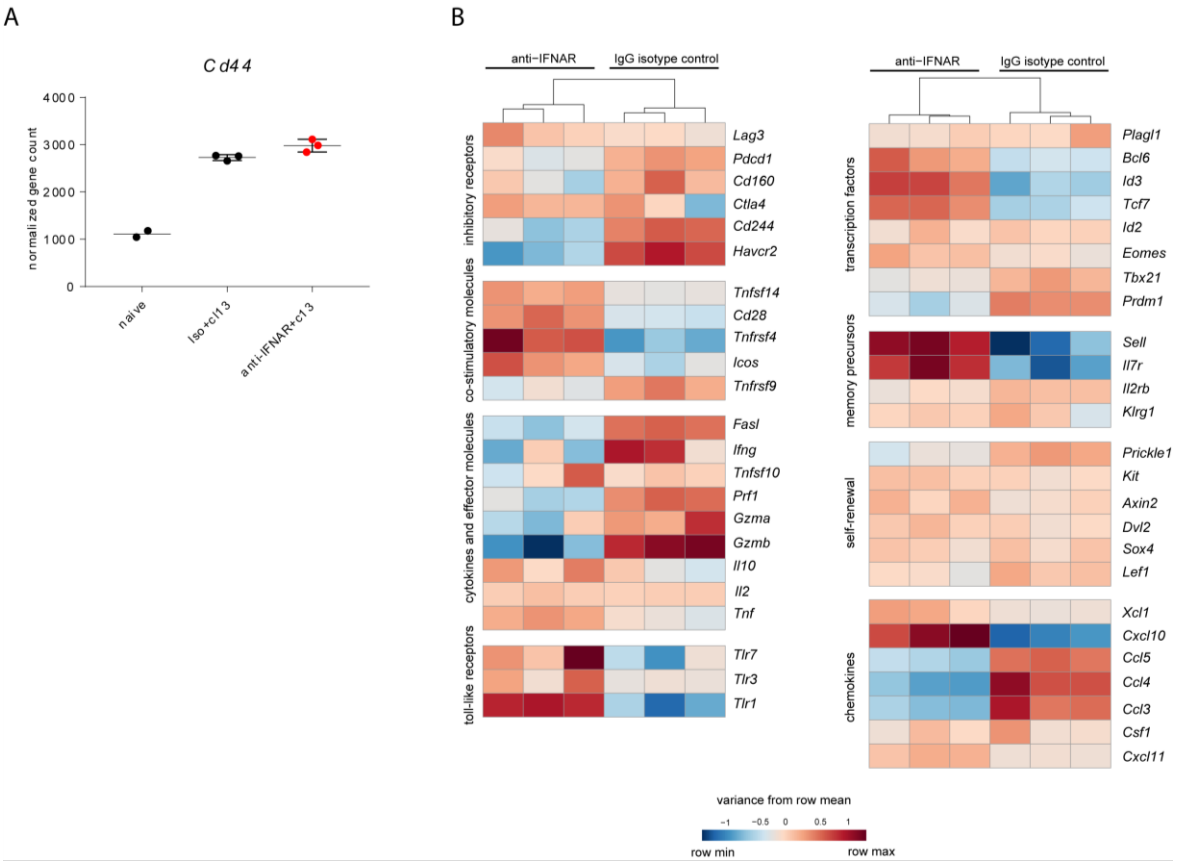


Figure 5.10 Increased T_{FC} transcriptional signature in P14 cells from aIFNAR1-treated mice. (A) Normalized gene count of *Cd44* in P14 cells from uninfected, LCMV-infected isotype-treated and LCMV-infected aIFNAR1-treated mice. (B) Heatmap of genes with important functions in T cells in aIFNAR1 versus isotype control-treated mice infected with LCMV Cl-13. Gene panel was adapted from Im et al. (214).

Differential expression analysis of P14s in anti-IFNAR1 versus isotype control-treated mice revealed 3565 differentially expressed genes (FDR<0.05). P14 cells in anti-IFNAR1 treated mice expressed increased relative levels of surface markers *Sell*, *Fas* and *Il7r*, molecules associated with naïve and memory

CD8+ T cells, while expressing lower levels of effector molecules such as *Ccl3*, *Ccl4*, *Ccl5* and *Gzmb* (**Figure 5.9C,D**). Strikingly, previously identified T_{FC} marker genes *Cxcr5*, *Tcf7* and *Id3* were among the 15 most significantly upregulated genes in P14 cells from anti-IFNAR treated versus isotype-treated mice (**Figure 5.9C-E**).

Quantifying the effect of anti-IFNAR1 versus isotype control treatment on the T_{FC} expression signature, the top 100 genes most significantly upregulated in virus-specific $CXCR5^+$ $Tim3^-$ cells compared to $CXCR5^-$ $Tim3^+$ cells were significantly enriched in the anti-IFNAR1 treated group (**Figure 5.9F**). Leong et al. recently reported the transcriptomes of $CXCR5^+$ and $CXCR5^-$ P14 cells from mice infected with the Docile clone of LCMV (396). Remapping the raw data together with our P14 RNA-seq data using a common pipeline and subsequent principal component analysis revealed that the majority of transcriptional variance separating $CXCR5^+$ from $CXCR5^-$ P14 cells is identical to the variance separating anti-IFNAR1 from isotype control-treated mice-derived P14 cells (**Figure 5.9G**). FACS experiments confirmed that T_{FC} are indeed significantly expanded during cl13 infection following IFNAR1 blockade (Huang, Zak and Pratumchai et al., manuscript in preparation). Taken together, these data show that T_{FC} cells share a characteristic transcriptome signature that can be readily identified across independent studies, and that IFNAR1 blockade during LCMV infection significantly increases the relative number of T_{FC} cells within the virus-specific CD8+ pool.

5.6 Unexpected upregulation of type I Interferon-stimulated genes in P14 cells following IFNAR1 blockade

Type I interferon is known to induce a gene expression programme relying on the STAT and IRF family transcription factors (397, 398). Intriguingly, we observed a significant upregulation of type I interferon stimulated genes (ISGs) in mice treated with anti-IFNAR1 compared with the isotype-treated controls (**Figure 5.11**). Gene set enrichment analysis demonstrated a strong upregulation of the Gene Ontology set Response to Type I Interferon after anti-IFNAR1 blockade versus isotype controls whereas no enrichment of this set was observed in genes upregulated in isotype controls versus anti-IFNAR1 treated mice (**Figure 5.11A**). This is unlikely to be caused by resurgence of IFNAR activity following antibody treatment because virus-specific CD4+ T cells expressed significantly lower levels of ISGs in anti-IFNAR1 treated mice compared to controls as expected (**Figure 5.12**).

IFNAR is not the only receptor able to activate the JAK-STAT pathway; cytokine receptors can also signal through JAK-STAT (399). Therefore, it is possible that alternative signals converging on the JAK-STAT axis give rise to the observed transcription of genes that are also induced by IFNAR activation. For example, IFN-γ receptor (IFNGR) activation induces IFN-β-stimulated genes in CD4+ T cells in the absence of T-bet (389). We asked whether T-bet deficiency in the presence of IFN-γ might be responsible for the observed ISG expression. Although levels *Tbx21*, the gene encoding T-bet, were significantly reduced in P14 cells from anti-IFNAR versus isotype-treated mice, the gene is still expressed at moderate levels (**Figure 5.12**). Moreover, expression of *Ifngr1* in P14 cells was reduced even more significantly than expression of *Tbx21* (**Figure 5.12**).

Therefore, IFNGR activity is unlikely to be solely responsible for the upregulation of ISGs following INFAR blockade.

It is possible that the upregulation of type I IFN target genes is partly caused by the increased ratio of TFC to effector CD8+ T cells. Indeed, we observed that CXCR5+ TIM3- CD8+ T cells express higher levels of type I IFN stimulated genes than CXCR5- TIM3+ CD8+ T cells (**Figure 5.11C**). This effect was replicated in CXCR5+ vs CXCR5- P14 cells (**Figure 5.11D-E**). However, the entire P14 cell compartment is not composed merely of CXCR5+ cells. Moreover, the mice were treated with anti-IFNAR antibody, suggesting type I IFN signalling should be attenuated. Therefore, other factors are likely to contribute to the observed upregulation of type I IFN stimulated genes.

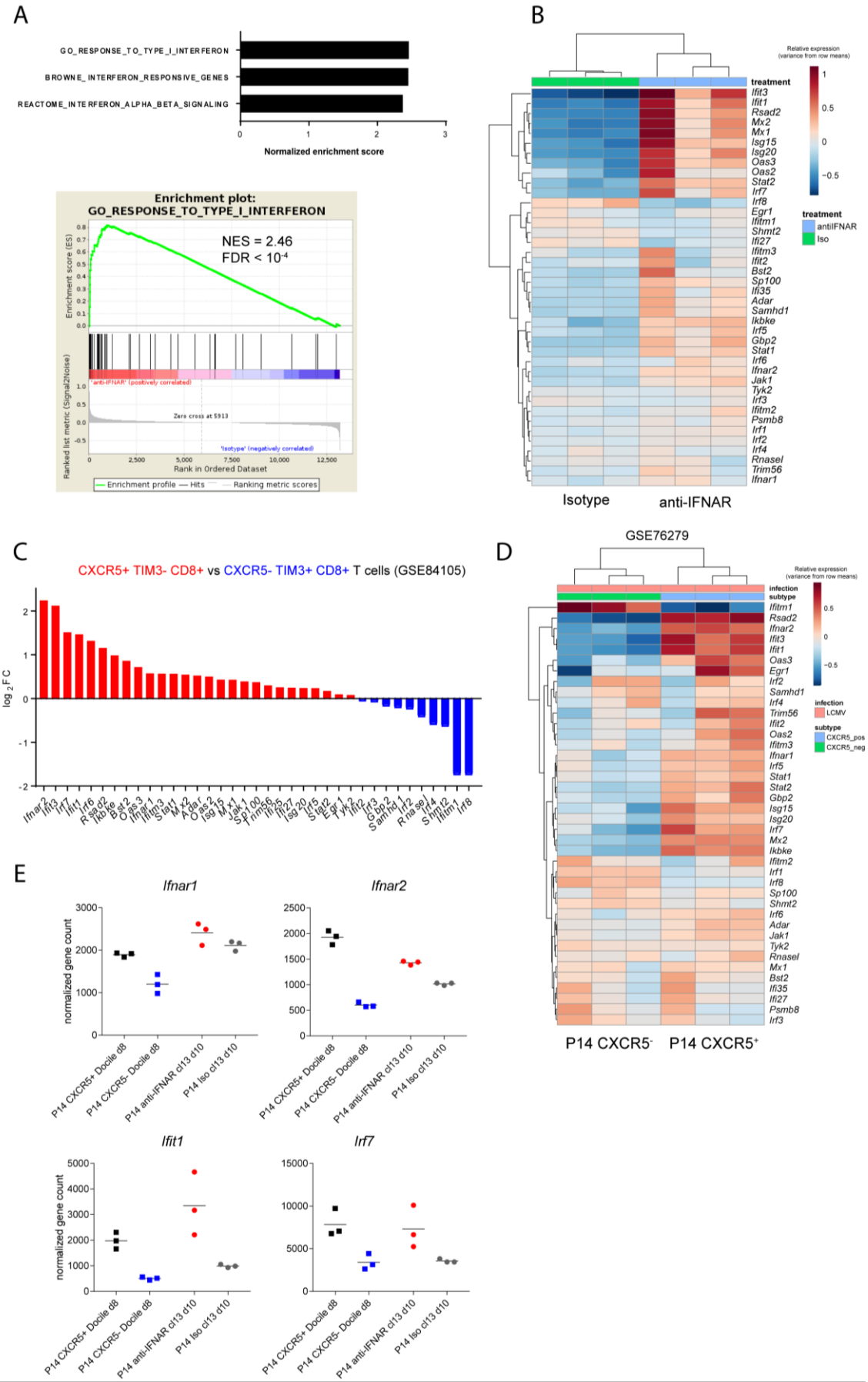

A
GO_RESPONSE_TO_TYPE_I_INTERFERON
BROWNE_INTERFERON_RESPONSIVE_GENES
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
Normalized enrichment score
Enrichment plot:
GO_RESPONSE_TO_TYPE_I_INTERFERON
NES = 2.46
FDR < 10-4
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'anti-IFNAR' (positively correlated)
Zero cross at 5913
'Isotype' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
B
Relative expression
(variance from row means)
treatment
Ifit3
Ifit1
Rsad2
Mx2
Mx1
Isg15
Isg20
Oas3
Oas2
Stat2
Irf7
Irf8
Egr1
Ifitm1
Shmt2
Ifi27
Ifitm3
Ifit2
Bst2
Sp100
Ifi35
Adar
Samhd1
Ikbke
Irf5
Gbp2
Stat1
Irf6
Ifnar2
Jak1
Tyk2
Irf3
Ifitm2
Psmb8
Irf1
Irf2
Irf4
Rnasel
Trim56
Ifnar1
antiIFNAR
Iso
Isotype
anti-IFNAR
C
CXCR5+ TIM3- CD8+ vs CXCR5- TIM3+ CD8+ T cells (GSE84105)
log2FC
D
GSE76279
infection
subtype
LCMV
CXCR5_pos
CXCR5_neg
P14 CXCR5-
P14 CXCR5+
E
Ifnar1
Ifnar2
Ifit1
Irf7
normalized gene count
P14 CXCR5+ Docile d8
P14 CXCR5- Docile d8
P14 anti-IFNAR cl13 d10
P14 Iso cl13 d10

Figure 5.11. T_{FC} phenotype is associated with increased expression of type 1 IFN stimulated genes. (A) Significant enrichment of type 1 IFN response genes in P14 cells from aIFNAR1-treated mice versus isotype-treated controls. Analysis was performed using GSEA. (B) Gene expression heatmap of Gene Ontology 'Response to type 1 interferon' gene set in P14 cells from aIFNAR1 versus isotype control treated mice. (C) Fold change of genes in the Gene Ontology 'Response to type 1 interferon' gene set in CXCR5$^+$ Tim3$^-$ CD8+ versus CXCR5$^-$ Tim3$^+$ CD8+ T cells (GSE84105) calculated using GEO2R. (D) Gene expression heatmap of Gene Ontology 'Response to type 1 interferon' gene set in CXCR5+ versus CXCR5- P14 cells from mice infected with LCMV Docile (GSE76279). (E) Normalized gene counts of type 1 IFN receptor chains (*Ifnar1*, *Ifnar2*) and example type 1 IFN stimulated genes (*Ifit1*, *Irf7*) in CXCR5$^+$ P14, CXCR5$^-$ P14 (GSE76279) and P14 cells from aIFNAR1-treated and isotype control treated mice. Read counts were normalized using a single pipeline containing both datasets.

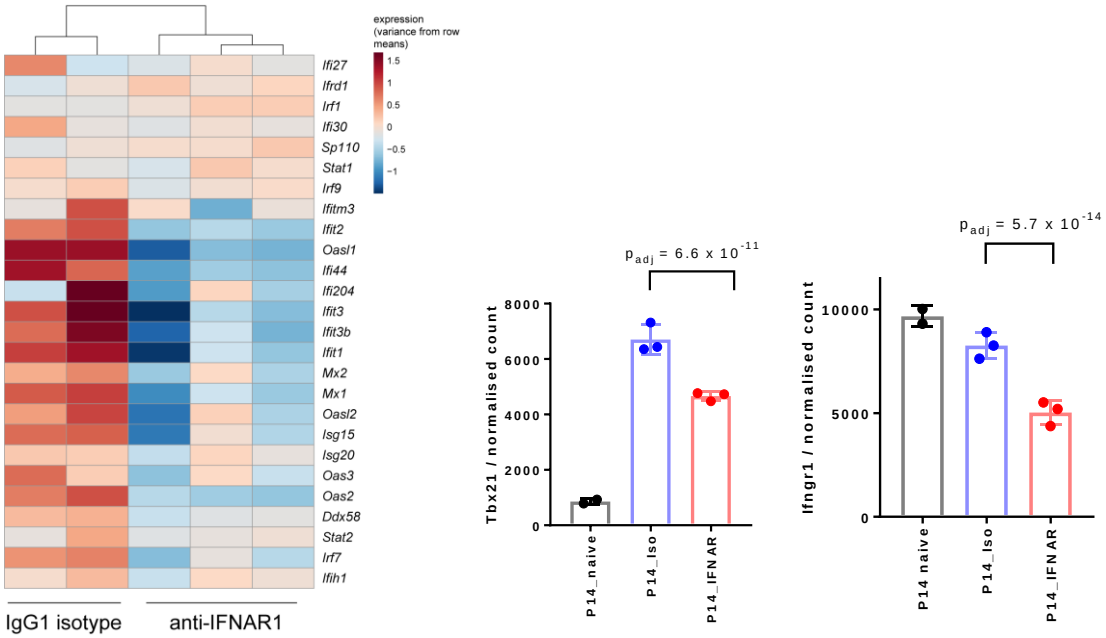


Figure 5.12. Reduced expression of type I IFN stimulated genes in virus-specific CD4+ T cells during cl13 infection following IFNAR blockade (left). *Tbx21* and *Ifngr1* levels in P14 cells following IFNAR blockade (right).

5.7 IL-27 induces transcription of canonical type I interferon target genes via STAT1

Two other cytokines are also known to activate the JAK-STAT pathway: IL-6 and IL-27. These cytokines bind distinct receptors which however share a common subunit, gp130 (400). IL-6 levels in serum of cl13-infected mice are significantly increased by IFNAR1 blockade (394). It is unknown how levels of IL-

27 are affected by IFNAR1 blockade. We set out to investigate whether IL-6 and/or IL-27 signalling on virus-specific CD8+ T cells after IFNAR blockade might contribute to the observed upregulation of ISGs.

The distinct transcriptional responses elicited by IL-6 versus IL-27, and their utilisation of STAT1 versus STAT3, were extensively characterised by Hirahara et al. (401). Making use of their publicly available RNA-seq data, we compared genes upregulated by IL-6 and/or IL-27 treatment with genes upregulated in mice treated with anti-IFNAR versus isotype-treated controls. To ensure uniformity, raw sequencing reads were remapped and differential expression analysis performed using a common pipeline for all datasets. 39% of genes induced by both IL-27 and IL-6 at 6 h post-treatment were also significantly upregulated by anti-IFNAR versus isotype. These genes included the cytokine *Il21* and transcription factor *Maf*, which are both important for T_{FH} function (402), but overall no Gene Ontology terms were enriched at a statistically significant level. Strikingly, a strong bias was observed in the overlap between IL-6-only-induced versus IL-27-only-induced genes with anti-IFNAR-upregulated genes: 20% of IL-6-only-induced genes at 6 h post-treatment were also significantly upregulated after anti-IFNAR versus isotype whereas the overlap was 50% for IL-27-only-induced genes and anti-IFNAR upregulated genes ($p = 4.0x10^{-7}$, Fisher's exact two-tailed test) (**Figure 5.13**). This bias towards IL-27- versus IL6-induced genes was also observed for genes upregulated at 24 h after treatment ($p = 4.0x10^{-5}$, Fisher's exact two-tailed test) (**Figure 5.13**). Therefore, although genes induced by both IL-6 and IL-27 were upregulated in anti-IFNAR- versus isotype-treated mice, IL-27-specific target genes were significantly enriched in the anti-IFNAR-induced transcriptome compared to IL6-specific targets.

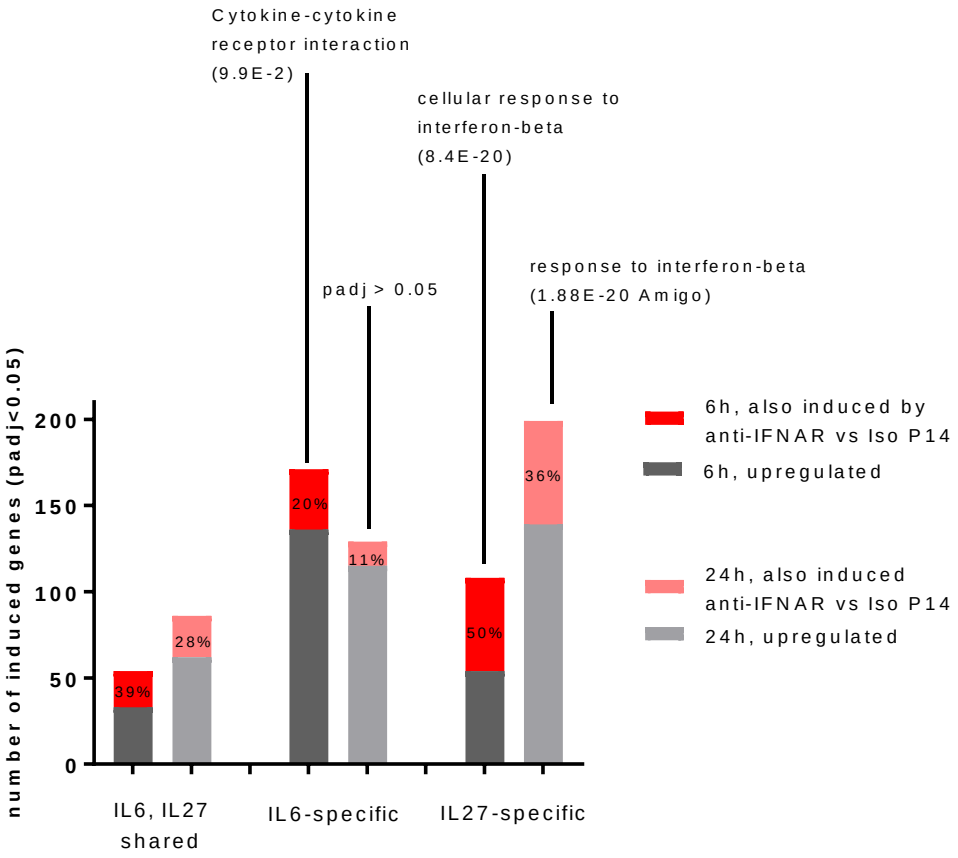


Figure 5.13. Comparison or the overlap between IL-6 and IL-27-induced genes with genes upregulated by anti-IFNAR vs isotype treatment. Publicly available data (GSE65621) were used to obtain genes induced by IL-6 and/or IL-27 compared to vehicle-treated cells, and compared to differentially expressed genes from anti-IFNAR vs isotype-treated mice.

Enrichment analysis of genes induced by both IL-27 and anti-IFNAR treatment showed a highly significant overlap with type I IFN target genes: the top enriched GO term was 'cellular response to interferon-beta (Bonferroni-corrected p = 8.4×10^{-20}). The enrichment of these type I IFN target genes remained highly significant for genes induced at 24 h post-IL-27-treatment (**Figure 5.13**). In addition, mRNA levels of the IL-27 receptor subunit WSX1 that is not shared with the IL-6 receptor and is encoded by the *Il27ra* gene, were significantly increased following IFNAR blockade (**Figure 5.14**). Expression of *Il6ra* was also increased but remained at much lower absolute levels than *Il27ra*. The striking overlap

between IL-27 and type I IFN induced genes stems from their activation of STAT1. Whereas IL-6 preferentially induces STAT3 phosphorylation, the response to IL-27 is skewed towards STAT1, a transcription factor also activated by IFNAR (401). These results lend credence to the hypothesis that IL-27R activation in the presence of anti-IFNAR antibody results in the transcription of canonical type I IFN stimulated genes, offering to resolve the apparent paradox.

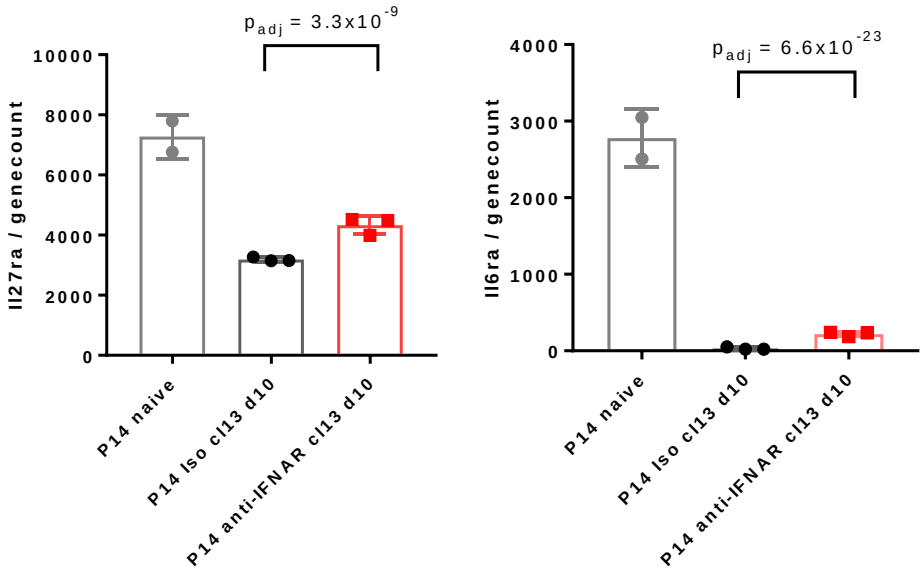


Figure 5.14. mRNA levels of IL-27 and IL-6 receptors are increased following IFNAR blockade. Normalised counts shown, mean +- s.d. p-values were calculated using DESeq2 package, differential expression analysis function and adjusted for multiple hypothesis testing.

5.8 IL-27R is essential for the selective expansion of T_{FC} following IFNAR blockade

Having determined IL-27R as a candidate receptor that is over-activated on P14 cells following IFNAR blockade, we wondered whether this activation had functional significance during LCMV cl13 infection. Wild-type and IL-27R-deficient mice (*Il27ra*$^{-/-}$) were treated with anti-IFNAR1 antibody or IgG1 isotype control and

infected with LCMV cl13. At day 9 post infection, virus-specific splenic T cells were examined using a tetramer stain and FACS. The percentage of CD4+ T_{FH} cells of total CD4+ T cells increased in a similar manner in both wild type and *Il27ra*$^{-/-}$ mice after anti-IFNAR treatment (**Figure 5.15A-C**). Consistently with previous studies, germinal centre B cells also exhibited a relative increase following IFNAR blockade and this effect was replicated in *Il27ra*$^{-/-}$ mice (**Figure 5.15D-F**). Remarkably, although the percentage of virus-specific T_{FC} increased in wild type mice following IFNAR blockade, this increase was completely abolished in *Il27ra*$^{-/-}$ mice (**Figure 5.15G-I**). Taken together, these data demonstrate that IL-27R is required for the selective expansion of virus-specific T_{FC} and this requirement holds even in the presence of expanded germinal centres and T_{FH} cells. (Experiments in Section 5.8 were performed in collaboration with Isaraphorn Pratumchai and Zhe Huang; Huang, Zak and Pratumchai et al., manuscript in preparation).

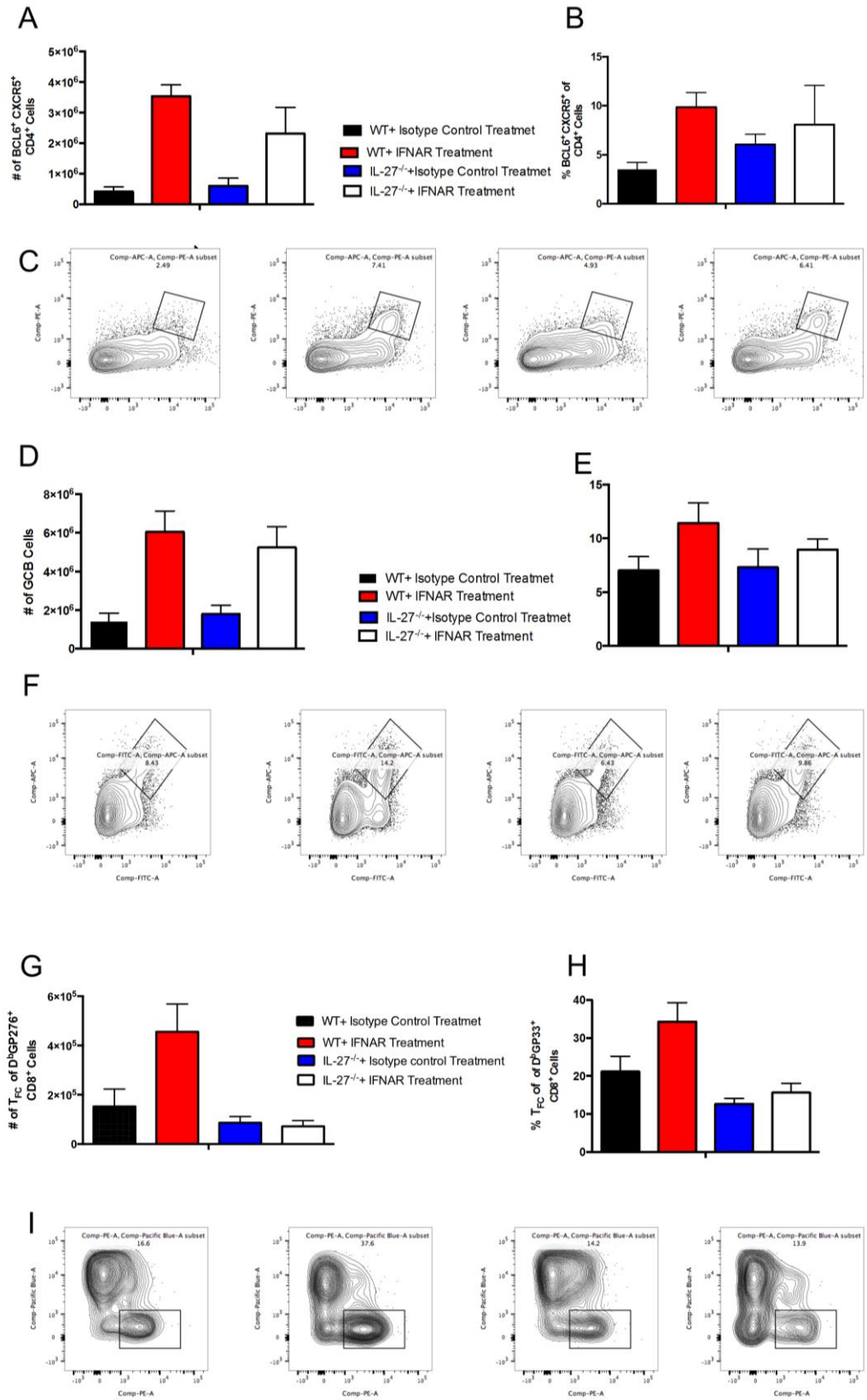


Figure 5.15 *Il27ra*$^{-/-}$ mice do not exhibit expansion of T_{FC} following IFNAR blockade. Panels A, D, G show the number of T_{FH}, GCB and T_{FC} cells in the spleen; panels B, E, H show the percentage of T_{FH}, GCB and T_{FC} cells of their parent population as specified. Experiments performed in collaboration with Isaraphorn Pratumchai.

Discussion

The introduction of checkpoint inhibitors in cancer therapy brought unprecedented clinical responses. These therapies, including antibodies targeting PD-1, PD-L1 and CTLA-4, rely in part on the ability to reactivate tumour-specific CD8+ T cells that had lapsed into the hypofunctional state of T cell exhaustion. Two factors that still limit checkpoint inhibitor responses are epigenetically-imprinted durability of T cell exhaustion, and scarcity of tumour-specific T cell clones which may not have the ability to self-renew (403, 404). The recent identification of T_{FC} as the cell type responsible for the proliferative burst of CD8+ effector cells following anti-PD-1 treatment (214) opened doors to strategies towards improving the efficacy of checkpoint inhibitors.

Here we examined the transcriptional landscape of total and virus-specific CD8+ T cells during the exhaustion-associated phase of persistent viral infection. Previously identified markers of exhaustion were upregulated on virus-specific P14 cells during chronic verses acute infection. However, we also observed upregulation of T_{FC} markers such as *Bcl6* in total CD8+ T cells, suggesting the percentage of T_{FC} of total CD8+ T cells increases during persistent versus acute infection. Treatment with anti-IFNAR1 antibody prior to LCMV cl13 infection significantly increased the expression of T_{FC} markers in virus-specific P14 cells and FACS experiments confirmed the selective expansion of both the total number and percentage of TFC among virus-specific CD8+ T cells. Intriguingly, type I IFN stimulated genes were also upregulated following IFNAR blockade. We found a combination of factors as the likely cause: increased *Ifnar1*, *Ifnar2* and type I IFN target genes on CXCR5+ versus CXCR5- P14 T cells; reduced levels of T-bet following IFNAR blockade enabling activation of type I IFN targets by IFN-γ, and

induction of STAT1 by IL-27. Given that type I IFN and IL-27 both activate STAT1 and share a significant number of downstream target genes, blockade of IFNAR likely acts to amplify the magnitude of IL-27 downstream signalling even in the presence of a stable amount of IL-27 ligand.

Experiments with IL-27R-deficient mice demonstrated the essential role of IL-27R in the selective expansion of T_{FC} following IFNAR blockade. Wu et al. previously showed that IFNAR$^{-/-}$ P14 cells form a higher percentage of T_{FC} during LCMV infection (405). Therefore, it is possible that IFNAR blockade leads to prioritisation of IL-27-dependent STAT1 target genes which positively regulated T_{FC} proliferation and/or survival. If so, activation of IL-27 signalling in the presence of attenuated IFN signalling might potentiate the efficacy of anti-PD-1 by expanding the T_{FC} population, thereby increasing the tumour-specific CD8+ T cell pool. Future experiments are warranted to address whether modulation of the IL-27/IFN balance can enhance anti-PD-1 responses.

Materials and Methods

LCMV virus stocks

LCMV Armstrong and LCMV clone 13 (406) were propagated in BHK cells. BHK cells were infected with virus and incubated for 60 h, then virus harvested from the medium. Viral stocks were aliquoted and kept at -80℃.

Antibody treatments

Mice were treated with 1 mg of anti-IFNAR1 antibody intraperitoneally (clone MAR1-5A3) or a mouse IgG1 isotype control (clone MOPC21) one day prior to virus infection. For PD-L1 blockade, isotype or aIFNAR1 treated LCMV Cl-13 infected mice received 200 mg isotype or anti–PD-L1 antibody every 3 days starting on day 15 post infection with a total of 5 treatments, then T cells and viral titers were analyzed at day 35.

Mouse LCMV infections

Wild-type C57BL/6 mice were injected intraperitoneally with LCMV Armstrong 200 µl at 10^6 pfu/ml, or intravenously using the tail vein with LCMV clone 13 at $2x10^6$ pfu/ml in a total volume of 200 µl.

Splenocyte isolation and counting

Mice were sacrificed using a lethal dose of isoflurane and mouse spleens dissected. The spleens were injected with collagenase and incubated at 37℃ for

30 min, then mechanically disrupted using a cell strainer, RPMI and a syringe. The resulting cell suspension was subjected to red blood cell (RBC) lysis, spun at 1500 rpm for 10 min and the supernatant removed. The RBC-free splenocytes were resuspended in RPMI and counted using the BioRad automated TC20 cell counter and slides.

T cell stimulation and isolation

Splenocytes were seeded in a 96-well plate at a concentration of 1 million per well in 200 µl of T cell stimulation medium. Gp33 peptide was added at a final concentration of 20 µM in medium to each well and the cells incubated at 37°C for 2 h. Subsequently the cells were pooled, spun at 1,500 rpm for 10 min and the supernatant discarded. The cell pellet was subjected to total T cell isolation using the Mouse Total T cell Isolation kit (Stem Cell Technologies) following the manufacturer's instructions.

Cell staining and sorting

Total mouse T cells were resuspended in stain buffer (PBS with 1% FBS) and appropriate antibodies added. The suspension was incubated at 4°C for 30 min, spun at 1,500 rpm for 10 min, supernatant discarded and cells resuspended in sorting buffer (PBS, 1% FBS, 1 mM EDTA). The cells were subjected to sorting at the Flow Cytometry Core Facility and collected in sort buffer.

RNA isolation

RNA from total CD8+ cells was isolated using the RNeasy Mini Kit (Qiagen; Hilden, Germany) with a DNase digest step using the RNase-Free DNase Set (Qiagen; Hilden, Germany) according to the manufacturer's instructions. RNA from P14 cells was isolated using the PicoPure RNA Isolation Kit (Thermo Fisher; Waltham, MA) with a DNase digest step using the RNase-Free Dnase Set (Qiagen; Hilden, Germany) according to the manufacturer's instructions. The RNA integrity was assessed by the Bioanalyzer (Agilent; Santa Clara, CA, USA).

RNA sequencing

For RNA from total CD8+ cells with satisfactory RIN scores (>8.5), libraries were constructed using the polyA-capture kit (NEB) and subjected to 2x150 bp sequencing on the NextSeq 500 sequencer (Illumina; San Diego, CA) at a mean depth of ca. 62.7 million reads per sample. For RNA from P14 cells with satisfactory RIN scores, libraries were constructed using the SMARTer v2 Universal Low Input RNA kit (Clontech; Mountain View, CA) and subjected to 2x150 bp sequencing on the NextSeq 500 sequencer (Illumina; San Diego, CA) at a mean depth of 25 M. For RNA from P14 cells with anti-IFNAR1 or isotype treatment amplified cDNA was prepared using SMART-Seq v4 Ultra Low Input RNA Kit (Clontech, Mountain View, CA), sheared using Covaris tubes (Covaris, Woburn, MA) and Illumina libraries constructed using the NEBNext Ultra DNA Library Prep Kit (NEB, Ipswich, MA) and sequenced using the NextSeq 500 system (Illumina, San Diego, CA) at a depth of 20-25 million single-end 75 bp reads per sample. Illumina's bcl2fastq Conversion Software (version 2.17.1.14) was used to de-multiplex reads and convert BCL files generated by the Illumina NextSeq 500 to FASTQ.

Read mapping and differential expression analysis

Quality of raw RNA-sequencing reads was checked using FastQC version 0.11.3 (Barbaham Institute, Cambridge, UK; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped to the mouse genome (GRCm38 assembly with Ensembl v87 transcriptome annotations) and gene counts calculated using STAR version 2.5.2a (407) using the following parameters: --quantMode GeneCounts –outSAMtype None –sjdbGTfile /Ensembl/Mus_musculus.GRCm38.87.gtf –sjdbOverhand 74. Statistical normalization and differential expression analysis were performed using the R package DESeq2 version 1.14.1 (408).

Principal component analysis and expression data visualisation

Principal component analysis was performed using R function prcomp, and the top 500 genes with the highest variance between samples were used. R version 3.3.2 was used for statistical analyses (352). Heatmaps were generated using R package pheatmap version 1.0.8 (409). Volcano plots were generated using custom code in the R language. GSEA version 2.2.3 was used for gene set enrichment analysis (410). Gene expression signature for CXCR5+ Tim3- CD8+ versus CXCR5- Tim3+ CD8+ T cells was retrieved from the publicly available dataset (GSE84105) by comparing CXCR5+ Tim3- to CXCR5- Tim3+ samples with GEO2R and calculating top 100 CXCR5+ specific genes by statistical significance (411). Raw RNA-sequencing data for CXCR5+ and CXCR5- P14 cells (GSE76279) were re-analyzed using the pipeline described above.

THESIS CONCLUSIONS

The goal of this thesis was to identify and functionally assess candidate pathogenic variants in cancer and congenital disorders. We have identified novel pathogenic deletions causing the 1q41q42 microdeletion syndrome; a candidate pathogenic 1q41q42 deletion in encephalocele; candidate pathogenic 1q41q42 microduplications, *de novo* single-nucleotide variants in childhood-onset schizophrenia and autism spectrum disorder; missense variants in neural tube defects, and candidate driver somatic mutations in cancer. The candidate pathogenic variants identified in this thesis are summarised in **Table 6.1**.

Table 6.1 Summary of disease-associated variants identified. Where appropriate, ACMG evidence codes and classification are given.

Variant	Phenotype	*De novo* / inherited	ACMG evidence codes (SNV)	ACMG classification (congenital disorders)	Notes
1q41q42 deletions	1q41q42 microdeletion syndrome; encephalocele	Most *de novo*	-	Pathogenic	Brain MRI consistent with *TP53BP2* dosage effect
1q41q42 duplication 1:223820857-224076362dup	Microcephaly; developmental delay, seizures, vision problems	Inherited (maternal)	-	Uncertain significance	Brain MRI consistent with *TP53BP2* dosage effect
1q41q42 duplication 1:222694279-224034322dup	Microcephaly, developmental delay, craniofacial, dental abnormalities, abnormal behaviour	Inherited (maternal)	-	Uncertain significance	Distal breakpoint within *TP53BP2*
PPP1R13B R72Q	Childhood-onset schizophrenia	*De novo*	PS2, PM2	Likely pathogenic	
PPP1R13B 14:104216291 AC>A	ASD	*De novo*	PS2, PM2	Likely pathogenic	
PPP1R13B G881R	ASD	*De novo*	PS2, PP3	Uncertain significance	
TP53 G279E	Spina bifida	Inherited	PS3, PM2, PP3	Likely pathogenic	
TP53 E358V	Spina bifida	Inherited / n/a	PS4, PM2	Likely pathogenic	Co-segregation with disease; 2 unrelated cases
TP73 V190F	Craniorachischisis	N/A	PS3, PM2, PP3	Likely pathogenic	

PARD3 D783G	Craniorachischisis	N/A	PM2, PS3?	Likely pathogenic (?)	Enrichment of aPKCBD variants in cases
PARD3 P913Q	Craniorachischisis	N/A	PM2, PP3, PS3?	Likely pathogenic (?)	Enrichment of aPKCBD variants in cases
PPP1R13B I322fs + LoH; X955_splice + LoH	Liver carcinoma	Somatic	-	Likely pathogenic	*Ppp1r13b* implicated in mouse liver carcinoma screen
PPP1R13B R81. + LoH; E64. + LoH	Non-small cell lung carcinoma	Somatic	-	Likely pathogenic	
TP53BP2 S769. + LoH	Bladder carcinoma	Somatic	-	Likely pathogenic	
PPP1R13B truncating mutations	Colorectal carcinoma	Somatic	-	Pathogenic	*PPP1R13B* significantly downregulated in colon cancer; *Ppp1r13b* implicated in 3 independent colorectal cancer screens in mice
TP53BP2 K513fs + LoH	Prostate carcinoma	Somatic	-	Likely pathogenic	
TP53BP2 P898fs + LoH	Lung squamous cell carcinoma	Somatic	-	Likely pathogenic	ASPP2-deficient mice get spontaneous lung carcinomas
PPP1R13B Pro-rich region missense mutations	Solid tumours	Somatic	-	Likely pathogenic	Enriched in wild-type-p53 tumours, associated with *BAX* and p21 levels
TP53BP2 Pro-rich region missense mutations	Solid tumours	Somatic	-	Likely pathogenic	Associated with *BAX* and p21 levels
PPP1R13L autoregulatory domains missense mutations	Solid tumours	Somatic	-	Likely pathogenic	Mutations disrupt self-binding; associated with *BAX* levels

Using a mouse model of ASPP2 deficiency, we showed that *TP53BP2* haploinsufficiency is likely to be responsible for structural brain abnormalities, and potentially encephalocele risk in 1q41q42 microdeletion patients, and that other phenotypes such as ventricular septal defect, cleft palate, and urogenital abnormalities also overlap between ASPP2-deficient mice and humans. Our

observation of microcephaly, developmental delay and seizures associated with 1q41q42 microduplications suggests *TP53BP2* is also triplosensitive in humans, albeit with incomplete penetrance. Analyses of spatiotemporal gene expression in human brain development, brain MRIs of *TP53BP2* CNV patients, and phenotypes of ASPP2-deficient mice suggest a function for ASPP2 in promoting the differentiation of striatum from the ganglionic eminences. *TP53BP2* deletion, diploid and duplication patients show an association between *TP53BP2* copy number, and caudate nucleus and lateral ventricle volumes. Dosage sensitivity of *TP53BP2* may also lead to abnormalities in ganglionic eminence-derived GABAergic interneuron populations in the cortex, which might underlie the observed epilepsy in 2/3 *TP53BP2* duplication patients.

Future studies will pinpoint the most critical developmental processes which are sensitive to *TP53BP2* dosage. Given that ASPP2 plays a role in multiple steps of CNS development and ASPP2 deficiency is associated with multiple distinct CNS phenotypes in mice, it is possible *TP53BP2* deletions confer both NTD risk and brain abnormality risk. Larger sample sizes are needed to deconvolve the contributions of haploinsufficiencies of individual genes towards the 1q41q42 microdeletion syndrome; current evidence supports a role for *DISP1*, *TP53BP2*, *FBXO28*, but other genes such as *CAPN2* might also a play a role. Importantly, our approach can be generalised to discover other dosage-sensitive genes in brain development. By integrating patient CNV and MRI data with population genomic statistics (such as ExAC), and patterns of spatiotemporal gene expression in brain development, it is possible to identify dosage-sensitive gene candidates and link them to the most critical stages in brain development.

Three *de novo PPP1R13B* mutations have been identified in childhood-onset schizophrenia and autism spectrum disorder. Based on the absence of 2/3 mutations from public databases, *PPP1R13B*'s pattern of expression in neurons, its spatiotemporal co-expression with synaptic genes, and the observation of ASPP1 protein in synaptic fractions, these mutations are biologically plausible candidate pathogenic variants. The newly generated ASPP1-deficient mice will provide a useful experimental system in which to assess the role of ASPP1 in brain development.

In a cohort of 340 neural tube defect cases, we identified a burden of rare missense variants in the apoptotic regulators *TP53* and *TP73*, including two loss-of-function variants (p53 G279E; p73 V190F) and one recurrent variant in *TP53* significantly enriched over the general population at a single-variant level (p53 E358V). This study also found a case-specific enrichment of missense variants in the aPKC-binding domain of *PARD3*, which were experimentally shown to disrupt Par3's binding to aPKC. The collective burden of these rare variants might point to a single step in neural tube closure: apical constriction. Par3 is known to control this process via aPKC-dependent control of ROCK. Recently, apoptotic cells have been shown to drive apical constriction by pulling myosin II cables (412). If this process also controls apical constriction in neural tube closure, the enrichment of both *PARD3* and *TP53* pathway missense variants in NTD cases might point to genetic disruption of apical constriction as a risk factor for human NTDs. This hypothesis is consistent with the implication of *PARD3* CNVs in NTD risk, and the discovery of 2 independent *de novo* truncating mutations in *SHROOM3*, another regulator of ROCK-dependent apical constriction.

An unbiased integrative analysis of cancer clarified the role of ASPP genes in human tumorigenesis. *PPP1R13B* shows highly concordant hallmarks of a tumour suppressor gene, with a causal role supported by multiple mouse mutagenesis screens and the recurrence of truncating variants in multiple cancer types, most prominently colorectal cancer. Based on the unequivocal multi-source evidence for ASPP1's tumour suppressive role in colorectal cancer, truncating *PPP1R13B* mutations can be considered pathogenic. *PPP1R13L* shows hallmarks of an Epi-driver oncogene, being broadly upregulated across cancer types, and high iASPP levels correlating with cancer drug resistance. *TP53BP2* has context-dependent behaviour suggesting pleiotropic function which may be pro- or anti-tumorigenic depending on the tissue type and molecular makeup of each tumour. This bipolar nature of *TP53BP2*'s role is in contrast to the molecular functions uncovered for ASPP2 so far, which are virtually all tumour suppressive. The genomic data suggest ASPP2 can play tumour promoting roles, or at least be compatible with tumour growth in certain tissues. Investigating these oncogenic modes of ASPP2's action at the molecular level will help bridge the gap between the current mechanistic understanding of ASPP2 and the genomic view of *TP53BP2* in human tumours.

Our mechanistic analysis of missense mutations in cancer uncovered a common theme of dysregulation: mutation of the autoregulatory Pro-rich regions in order to modulate p53 family-dependent transcriptional activity. This example serves as proof of principle that rare somatic mutations can be positively selected, and the methodology employed in discovering them can be scaled up to a genome-wide search. With the rise of harmonised cancer genomic data

repositories, cloud computing and public APIs, such analyses can be applied to discover a class of rare driver and mini-driver mutations which are otherwise overlooked by gene-level burden-type tests such as MutSig2CV.

Overall, the data presented in this thesis expand the genomic catalogue of human disease, and inform our understanding of the genetic architectures of congenital diseases and cancer. High-penetrance CNVs identified in syndromic developmental delay cases are consistent with a known role for *de novo* and rare CNVs in these disorders. In contrast, the diverse polygenic burden of rare missense variants found in our NTD cohort highlights the complex genetic makeup of this disease and suggests many incompletely penetrant rare variants, both SNVs and CNVs, may comprise the genetic risk of NTDs. In cancer, our results support a role for rare somatic mutations that are likely to be highly context-dependent, for example ASPP2 N-Pro mutations observed in *KRAS*-mutant wild type-*TP53* tumours. Such context-dependent mutations shift the view of somatic cancer genomes from Mendelian-like, with dominant oncogenic alleles and recessive tumour suppressor mutations in a sea of passengers, to a more complex architecture which includes a significant contribution from rare alleles of moderate to small effect.

REPRINT PERMISSIONS

Permissions for re-use were obtained through the RightsLink Copyright Clearance Center and/or the publisher in accordance with the publisher's copyright regulations.

Figure 4 from Robinson et al. (224) was reproduced with permission granted by license number 3956190083952 on 25 September 2016.

Figure 1 and Table 1 from Castro-Giner et al. (175) were reproduced with license number 3956880166858 granted on 27 September 2016.

Figure 5 from Lawrence et al. (162) was reproduced with license number 3957431075723 granted on 27 September 2016.

Tables 3-5 and Figure 1 from Richards et al. (65) were reproduced with permission granted by license no. 3958131471608 on 29 September 2016.

Figure 5 from the UK10K study (20) was reproduced in accordance with the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence.

Figure 2 from Hindorff et al. was reproduced with license number 3958750496258 obtained on 30 September 2016.

Figure 1 from Sullivan et al. (137) was reproduced with license number 3960461445736 obtained on 1 October 2016.

Figure 1 from Sala et al. (215) was reproduced with license number 3960920067455 granted on 2 October 2016.

Figure 1 from Song et al. (228) was reproduced with license number 3960990308154 granted on 2 October 2016.

Figure 2 from Marin and Rubinstein (259) was reproduced with license number 3961120636674 granted on 2 October 2016.

Figure 2 from Macara (360) was reproduced with license number 3962050275924 granted on 4 October 2016.

Permission was not required by the publisher for the reproduction of Figure 15 from Rossier et al. (1).

APPENDIX

Code

Code 1. Differential expression between tumour and normal tissue.

```
# Script to compute tumour vs normal expression for a given gene from Firebrowse
                RSEM log2 expression data
# Done in batches of 12 to enable pages of boxplots to be exported in PDF
# NB: script has not been configured to handle missing values in certain datasets,
                requires manual override.

genename <- readline(prompt="Enter official gene symbol:")
print(paste("Plotting tumour vs normal expression for", genename, sep=" "))

FC <- list()
log2fc <- list()
tcga_names <- c("BLCA", "BRCA", "CESC", "CHOL", "COAD", "COADREAD", "ESCA",
                "GBM", "GBMLGG", "HNSC", "KICH", "KIPAN", "KIRC", "KIRP", "LIHC",
                "LUAD", "LUSC", "PAAD", "PCPG", "PRAD", "READ", "SARC", "STAD",
                "THCA", "THYM", "UCEC")
pdf(paste(genename, "plot_TPvsNT_1.pdf", sep="_"), width=6, heigh=12)
par(mfrow=c(4,3))

for (i in 1:12){
 mrna_NT <-
                read.csv(url(paste("http://firebrowse.org/api/v1/Samples/mRNASeq?format
                =csv&gene=", genename, "&cohort=", tcga_names[[i]],
                "&sample_type=NT&protocol=RSEM&page=1&page_size=2000&sort_by=
                cohort", sep="")))
 mrna_TP <-
                read.csv(url(paste("http://firebrowse.org/api/v1/Samples/mRNASeq?format
                =csv&gene=", genename, "&cohort=", tcga_names[[i]],
                "&sample_type=TP&protocol=RSEM&page=1&page_size=2000&sort_by=
                cohort", sep="")))
 boxplot(mrna_NT[,3], mrna_TP[,3], names=c("NT", "TP"), col=c("blue", "red"),
                main=tcga_names[[i]], ylab=paste(genename, "expression_log2 (RSEM)"),
                sub=paste("p =",format(t.test(mrna_NT[,3], mrna_TP[,3])$p.value,
                scientific=TRUE), sep=" "))
 print(paste(tcga_names[[i]],"p =",format(t.test(mrna_NT[,3], mrna_TP[,3])$p.value,
                scientific=TRUE), sep=" "))
 FC[[i]] <- mean(mrna_TP[mrna_TP$cohort==tcga_names[[i]],3]) -
                mean(mrna_NT[mrna_NT$cohort==tcga_names[[i]],3])
 print(FC[[i]])
 }
dev.off()

pdf(paste(genename, "plot_TPvsNT_2.pdf", sep="_"), width=6, heigh=12)
par(mfrow=c(4,3))
```

```
for (i in 13:24){
  mrna_NT <-
                    read.csv(url(paste("http://firebrowse.org/api/v1/Samples/mRNASeq?format
                    =csv&gene=", genename, "&cohort=", tcga_names[[i]],
                    "&sample_type=NT&protocol=RSEM&page=1&page_size=2000&sort_by=
                    cohort", sep="")))
  mrna_TP <-
                    read.csv(url(paste("http://firebrowse.org/api/v1/Samples/mRNASeq?format
                    =csv&gene=", genename, "&cohort=", tcga_names[[i]],
                    "&sample_type=TP&protocol=RSEM&page=1&page_size=2000&sort_by=
                    cohort", sep="")))
  boxplot(mrna_NT[,3], mrna_TP[,3], names=c("NT", "TP"), col=c("blue", "red"),
                    main=tcga_names[[i]], ylab=paste(genename, "expression_log2 (RSEM)"),
                    sub=paste("p =",format(t.test(mrna_NT[,3], mrna_TP[,3])$p.value,
                    scientific=TRUE), sep=" "))
  print(paste(tcga_names[[i]],"p =",format(t.test(mrna_NT[,3], mrna_TP[,3])$p.value,
                    scientific=TRUE), sep=" "))
  FC[[i]] <- mean(mrna_TP[mrna_TP$cohort==tcga_names[[i]],3]) -
                    mean(mrna_NT[mrna_NT$cohort==tcga_names[[i]],3])
  print(FC[[i]])
}
dev.off()

pdf(paste(genename, "plot_TPvsNT_3.pdf", sep="_"), width=6, heigh=12)
par(mfrow=c(4,3))

for (i in 26:26){
  mrna_NT <-
                    read.csv(url(paste("http://firebrowse.org/api/v1/Samples/mRNASeq?format
                    =csv&gene=", genename, "&cohort=", tcga_names[[i]],
                    "&sample_type=NT&protocol=RSEM&page=1&page_size=2000&sort_by=
                    cohort", sep="")))
  mrna_TP <-
                    read.csv(url(paste("http://firebrowse.org/api/v1/Samples/mRNASeq?format
                    =csv&gene=", genename, "&cohort=", tcga_names[[i]],
                    "&sample_type=TP&protocol=RSEM&page=1&page_size=2000&sort_by=
                    cohort", sep="")))
  boxplot(mrna_NT[,3], mrna_TP[,3], names=c("NT", "TP"), col=c("blue", "red"),
                    main=tcga_names[[i]], ylab=paste(genename, "expression_log2 (RSEM)"),
                    sub=paste("p =",format(t.test(mrna_NT[,3], mrna_TP[,3])$p.value,
                    scientific=TRUE), sep=" "))
  print(paste(tcga_names[[i]],"p =",format(t.test(mrna_NT[,3], mrna_TP[,3])$p.value,
                    scientific=TRUE), sep=" "))
  FC[[i]] <- mean(mrna_TP[mrna_TP$cohort==tcga_names[[i]],3]) -
                    mean(mrna_NT[mrna_NT$cohort==tcga_names[[i]],3])
  print(FC[[i]])
}
dev.off()
```

Code 2. Plotting the distribution of missense mutations in tumours wild type vs mutant for a given gene.

```
#Script for genetic interaction testing on cBio data only


library("cgdsr")


# Start the clock

ptm <- proc.time()


genename <- readline(prompt="Enter official gene symbol:")

print(paste("Extracting cBio mutations for", genename, sep=" "))


genename2 <- readline(prompt="Which gene to test mutation interaction with? Enter symbol:")

print(paste("Extracting mutational status for", genename2, sep=" "))


cgds_studies <- getCancerStudies(mycgds)


#This _mutations names are called genetic_profile_id by cBio

cancer_studies_mutations <- lapply(1:nrow(cgds_studies), function(i){paste(cgds_studies[i,1], "mutations", sep="_")})

#This _sequenced names are called case_list_id by cBio

cancer_studies_sequenced <- lapply(1:nrow(cgds_studies), function(i){paste(cgds_studies[i,1], "sequenced", sep="_")})


# To-do: Modify script to run analysis for a list of 'gene2s'

mutations_gene1 <- lapply(1:nrow(cgds_studies), function(i){getMutationData(mycgds, cancer_studies_sequenced[[i]], cancer_studies_mutations[[i]], genename)})

mutations_gene1_df <- do.call(rbind, lapply(mutations_gene1, data.frame, stringsAsFactors=FALSE))

print(paste("Mutations for", genename, "extracted", sep =" "))


mutations_gene2 <- lapply(1:nrow(cgds_studies), function(i){getMutationData(mycgds, cancer_studies_sequenced[[i]], cancer_studies_mutations[[i]], genename2)})
```

```
mutations_gene2_df <- do.call(rbind, lapply(mutations_gene2, data.frame,
stringsAsFactors=FALSE))

print(paste("Mutations for", genename2, "extracted", sep =" "))


# Calculate how many gene1-mutant samples are mutant for gene2 and how many wt

comut_list <- mutations_gene1_df$case_id %in% mutations_gene2_df$case_id

a <- table(comut_list)

print(paste(genename, "-mutant but ", genename2, "-wt samples: ", a[[1]], sep=""))

print(paste(genename, "-mutant and ", genename2, "-mutant samples: ", a[[2]], sep=""))

print("Warning: does include duplicates")


#This expression returns a vector of logical values

# mutations_gene1_df$case_id %in% mutations_gene2_df$case_id


dataframe_status <- do.call(rbind, lapply(mutations_gene1_df$case_id %in%
mutations_gene2_df$case_id, data.frame, stringsAsFactors=FALSE))

colnames(dataframe_status) <- "gene2_mutated"

mutations_gene1_df <- cbind(mutations_gene1_df, dataframe_status)


#This command extracts a number from a string. Then add the residue number as a
column to the mutation list

Residue <- matrix(gsub("[^0-9]", "",mutations_gene1_df[,8]),
nrow=nrow(mutations_gene1_df))

Residue <- as.numeric(Residue)

mutations_gene1_df <- cbind(mutations_gene1_df, Residue, deparse.level=0)


#Plotting the frequency distribution for gene2-wt vs mutant tumours.

  par(mfrow=c(2,1))


hist(mutations_gene1_df[mutations_gene1_df$mutation_type=="Missense_Mutation"&mu
tations_gene1_df$gene2_mutated==FALSE,24], breaks=40, xlab="Residue",
main=paste(genename, " missense mutations in ", genename2, "-wt tumours (cBio)",
sep=""), col="green", cex.main=0.9, cex.lab=0.9)


hist(mutations_gene1_df[mutations_gene1_df$mutation_type=="Missense_Mutation"&mu
tations_gene1_df$gene2_mutated!=FALSE,24], breaks=40, xlab="Residue",
```

```
main=paste(genename, " missense mutations in ", genename2, "-mutant tumours (cBio)", sep=""), col="red", cex.main=0.9, cex.lab=0.9)


# Plotting the same thing after de-duplicating all case IDs (warning: some samples have more than 1 mutation in the same gene)

 mutations_gene1_dedup <- mutations_gene1_df[!duplicated(mutations_gene1_df$case_id),]

 mutations_gene2_dedup <- mutations_gene2_df[!duplicated(mutations_gene2_df$case_id),]


hist(mutations_gene1_dedup[mutations_gene1_dedup$mutation_type=="Missense_Mutation"&mutations_gene1_dedup$gene2_mutated==FALSE,24], breaks=40, xlab="Residue", main=paste(genename, " missense mutations in ", genename2, "-wt tumours (cBio, dedup)", sep=""), col="green", cex.main=0.9, cex.lab=0.9)


hist(mutations_gene1_dedup[mutations_gene1_dedup$mutation_type=="Missense_Mutation"&mutations_gene1_dedup$gene2_mutated!=FALSE,24], breaks=40, xlab="Residue", main=paste(genename, " missense mutations in ", genename2, "-mutant tumours (cBio, dedup)", sep=""), col="red", cex.main=0.9, cex.lab=0.9)

 # Density plots to compare distributions


plot(density(mutations_gene1_df[mutations_gene1_df$mutation_type=="Missense_Mutation"&mutations_gene1_df$gene2_mutated==FALSE,"Residue"], width=25), col="green", main=paste(genename, "mutations in ", genename2, "-wt vs -mut tumours"))


lines(density(mutations_gene1_df[mutations_gene1_df$mutation_type=="Missense_Mutation"&mutations_gene1_df$gene2_mutated==TRUE,"Residue"], width=25), col="red", main=paste(genename, "mutations in ", genename2, "-wt vs -mut tumours"))

 # Density plots of de-duplicated data

 par(mfrow=c(2,2))

 for (i in c(25,50,75,100)){


plot(density(mutations_gene1_dedup[mutations_gene1_dedup$mutation_type=="Missense_Mutation"&mutations_gene1_dedup$gene2_mutated==FALSE,"Residue"], width=i), col="green", main=paste(genename, "mutations in ", genename2, "-wt vs -mut tumours (dedup)"))


lines(density(mutations_gene1_dedup[mutations_gene1_dedup$mutation_type=="Missense_Mutation"&mutations_gene1_dedup$gene2_mutated==TRUE,"Residue"], width=i), col="red", main=paste(genename, "mutations in ", genename2, "-wt vs -mut tumours"))

 }
```

```
#Now we change the column name so that the final data frame has a record of which gene is gene2

  colnames(dataframe_status) <- paste(genename2, "mutated", sep="_")

  mutations_gene1_df <- cbind(mutations_gene1_df, dataframe_status)

  mutations_gene1_df <- mutations_gene1_df[,-23]


  write.csv(as.data.frame(mutations_gene1_df), file=paste(genename, "mutations_cBio.csv", sep="_"))

  #Try to assign bins using cut

  #mutations_gene1_df$bins <- with(mutations_gene1_df, cut(residue, breaks = seq(from = 0, to = 1200, by=50)))


# We want to identify the top bin in the gene2-mutated set and calculate a Fisher test of bin vs other bins combined in co-mut vs not

  # One tool for this is the hist$breaks function:


hist(mutations_gene1_df[mutations_gene1_df$mutation_type=="Missense_Mutation","Residue"], breaks=seq(from = 0, to = 1150, by=50))$counts

  # Find the maximum mutation count in the top bin in the co-mut subset

  count_mut <- hist(mutations_gene1_df[mutations_gene1_df$mutation_type=="Missense_Mutation"&mutations_gene1_df$gene2_mutated==TRUE,"Residue"], breaks=seq(from = 0, to = 1150, by=50))$counts

  count_gene1mut <- hist(mutations_gene1_df[mutations_gene1_df$mutation_type=="Missense_Mutation"&mutations_gene1_df$gene2_mutated==FALSE,"Residue"], breaks=seq(from = 0, to = 1150, by=50))$counts

  print(paste("Number of mutations in top co-mutated bin:", max(count_mut)))

  print(paste("Index of top co-mutated bin:", which.max(count_mut)))

  index_of_top_bin <- which.max(count_mut)

  sum_comut <- sum(count_mut)

  sum_gene1mut <- sum(count_gene1mut)

  # Fisher test to compare top bin vs all other bins

  fisher.test(matrix(c(count_mut[index_of_top_bin], sum_comut - count_mut[index_of_top_bin], count_gene1mut[index_of_top_bin], sum_gene1mut - count_gene1mut[index_of_top_bin]), nrow=2))
```

```
  print(paste("Fisher test p-value for top bin count vs other bins: ",
fisher.test(matrix(c(count_mut[index_of_top_bin], sum_comut -
count_mut[index_of_top_bin], count_gene1mut[index_of_top_bin], sum_gene1mut -
count_gene1mut[index_of_top_bin]), nrow=2))$p))
```

Code 3. mRNA z-scores for a given gene in a list of samples, e.g. tumours with a mutation in a specific gene.

```
# Computing z-scores for a subset of samples across multiple cancer based on e.g.
mutation status

# First we need to make a large data frame of z-scores collated from multiple tumour
types using cbind()

# For example an individual study df can be queried by z_rna_df[which(rownames(z_rna)
== "TP53"),"TCGA-QG-A5YW"]

# Generalised to z_rna_df[which(rownames(z_rna_df) == genename), samples[[i]]] to
iterate over i


   # Download the z-scores for TP only from cBio, bind them for a specific gene

   library("cgdsr")

   #define mycgds

   mycgds = CGDS("http://www.cbioportal.org/public-portal/")


   #Select gene to plot data for

   zgenename <- readline(prompt="Enter official gene symbol:")

   print(zgenename)


   #creating a matrix of TCGA study names and the associated RNASeq v2 sets

   tcga_names <- matrix(c("acc", "blca", "brca", "cesc", "chol", "coadread", "dlbc", "esca",
"gbm", "hnsc", "kich", "kirc", "kirp", "lihc", "luad", "lusc", "meso", "ov", "paad", "prad",
"pcpg", "sarc", "skcm", "stad", "tgct", "thym", "thca", "ucs", "ucec", "uvm"), nrow=30)

   tcga_names_v2 <- matrix(paste(tcga_names, "tcga_rna_seq_v2_mrna", sep="_"))

   tcga_names_v2_z <- matrix(paste(tcga_names,
"tcga_rna_seq_v2_mrna_median_Zscores", sep="_"))

   tcga_v2_names <- cbind(tcga_names, tcga_names_v2, tcga_names_v2_z,
deparse.level=0)

   tcga_v2_names_df <- data.frame(tcga_v2_names)


   #Loop through all TCGA studies to extract RNASeq V2 z-score data from cBio
```

```
  for (i in c(1:nrow(tcga_names))){
    rpkm <- getProfileData(mycgds, zgenename, tcga_v2_names_df[i,3],
tcga_v2_names_df[i,2])
    rpkm <- cbind(rpkm, matrix(tcga_v2_names_df[i,1], nrow=nrow(rpkm)))
    colnames(rpkm) <- c("mRNA_TP_Zscore", "study")
    assign(paste(i,"RPKM_v2", sep="_"), rpkm)
  }


  #combine individual studies' dataframes into one
  z_rbind <- rbind(`1_RPKM_v2`, `2_RPKM_v2`, `3_RPKM_v2`, `4_RPKM_v2`,
`5_RPKM_v2`, `6_RPKM_v2`, `7_RPKM_v2`, `8_RPKM_v2`, `9_RPKM_v2`,
`10_RPKM_v2`, `11_RPKM_v2`, `12_RPKM_v2`, `13_RPKM_v2`, `14_RPKM_v2`,
`15_RPKM_v2`, `16_RPKM_v2`, `17_RPKM_v2`, `18_RPKM_v2`, `19_RPKM_v2`,
`20_RPKM_v2`, `21_RPKM_v2`, `22_RPKM_v2`, `23_RPKM_v2`, `24_RPKM_v2`,
`25_RPKM_v2`, `26_RPKM_v2`, `27_RPKM_v2`, `28_RPKM_v2`, `29_RPKM_v2`,
`30_RPKM_v2`, deparse.level=0)
  plot_list <- lapply(1:nrow(tcga_names),
function(i){z_rbind[z_rbind$study==tcga_names[i],1]})
  rownames(z_rbind) <- gsub("\\.", "-", rownames(z_rbind))
  print("gene expression z-scores retrieved")


  #To search a list of names from a cBio-downloaded list of mutations and print z-scores
corresponding to custom gene
  #3rd column of mutations_gene1_df is the patient barcode
  mutations_gene1_df[3,3] %in% rownames(z_rbind)


  list1 <- matrix(nrow=nrow(mutations_gene1_df), ncol=5)
  for (i in 1:nrow(mutations_gene1_df)){
    if (mutations_gene1_df[i,3] %in% rownames(z_rbind) == "TRUE"){
      print(paste("Mutation:", mutations_gene1_df[i,8], zgenename, "z-score: ",
z_rbind[mutations_gene1_df[i,3],1]))
      list1[i,1] <- paste(mutations_gene1_df[i,8])
      list1[i,2] <- mutations_gene1_df[i,24]
      list1[i,3] <- paste(mutations_gene1_df[i,6])
      list1[i,4] <- zgenename
      list1[i,5] <- paste(z_rbind[mutations_gene1_df[i,3],1])
```

```
      colnames(list1) <- c("Mutation", "Mutation_type", "Residue", zgenename,
paste(zgenename, "z-score"))
     }
   }
   write.csv(list1, file=paste(zgenename, "z_scores_nodedup", genename, ".csv"))

   #Retrieving z-scores for de-duplicated samples only
   list1 <- matrix(nrow=nrow(mutations_gene1_dedup), ncol=5)
   for (i in 1:nrow(mutations_gene1_dedup)){
     if (mutations_gene1_dedup[i,3] %in% rownames(z_rbind) == "TRUE"){
       print(paste("Mutation:", mutations_gene1_dedup[i,8], zgenename, "z-score: ",
z_rbind[mutations_gene1_dedup[i,3],1]))
       list1[i,1] <- paste(mutations_gene1_dedup[i,8])
       list1[i,2] <- mutations_gene1_dedup[i,24]
       list1[i,3] <- paste(mutations_gene1_dedup[i,6])
       list1[i,4] <- zgenename
       list1[i,5] <- paste(z_rbind[mutations_gene1_dedup[i,3],1])
       colnames(list1) <- c("Mutation", "Residue", "Mutation_type", zgenename,
paste(zgenename, "z-score"))
       }
     }

       write.csv(list1, file=paste(zgenename, "z_scores_dedup", genename, "_muts.csv"))

       #Converting the dataframe to numeric for plotting. Include missense only - nonsense
mutations mess up residue number.
       list1_df <- as.data.frame(list1)
       list1_df <- list1_df[list1_df$Mutation_type=="Missense_Mutation",]
       list1_df3 <- list1_df[complete.cases(list1_df),c(2,5)]
       list1_numeric <- matrix(nrow=nrow(list1_df3), ncol=2)
       list1_numeric[,1] <- as.numeric(levels(list1_df3[,1]))[list1_df3[,1]]
       list1_numeric[,2] <- as.numeric(levels(list1_df3[,2]))[list1_df3[,2]]
       list1_numericdf <- as.data.frame(list1_numeric)
       colnames(list1_numericdf) <- c("residue", "z")
```

```
library("ggplot2")
library("plyr")
list1_numericdf$bins <- with(list1_numericdf, cut(residue, breaks = seq(from = 0, to = 1200, by=50)))
df.plot <- ddply(list1_numericdf[!duplicated(list1_numericdf),], .(bins), summarise, avg.z = mean(z))
qplot(bins, avg.z, data = df.plot, ylab = paste(zgenename, "avg z-score"), xlab = "PPP1R13B sequence bin")


# Plotting density for different bin widths
par(mfrow=c(2,2))
for (i in c(0.1,10,25,75)){
  plot(density(list1_numericdf[list1_numericdf$z>0,1], width=i), col="red", main=paste(genename, "missense muts: ", zgenename, "z<0(blue) vs z>0"), xlim=c(0,1135))
  lines(density(list1_numericdf[list1_numericdf$z<0,1], width=i), col="blue", main=paste(genename, "missense muts: ", zgenename, "z<0(blue) vs z>0"))
}
```

ACMG Guidelines on variant classification

Appendix Table S1. ACMG Guideline criteria for classifying pathogenic variants.

Reproduced from Richards et al. (65) with permission.

Evidence of pathogenicity	Category
Very strong	PVS1 null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g., *GFAP*, *MYH7*) • Use caution interpreting LOF variants at the extreme 3′ end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case–control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case–control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.
Moderate	PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation PM2 Absent from controls (or at extremely low frequency if recessive) (Table 6) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing. PM3 For recessive disorders, detected in *trans* with a pathogenic variant Note: This requires testing of parents (or offspring) to determine phase. PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level. PM6 Assumed de novo, but without confirmation of paternity and maternity
Supporting	PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant. PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

LOF, loss of function; OR, odds ratio.

Appendix Table S2. ACMG Guideline criteria for classifying benign variants.

Reproduced from Richards et al. (65) with permission.

Evidence of benign impact	Category
Stand-alone	BA1 Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
Strong	BS1 Allele frequency is greater than expected for disorder (see **Table 6**)
	BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age
	BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing
	BS4 Lack of segregation in affected members of a family
	Caveat: The presence of phenocopies for common phenotypes (i.e., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.
Supporting	BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease
	BP2 Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in *cis* with a pathogenic variant in any inheritance pattern
	BP3 In-frame deletions/insertions in a repetitive region without a known function
	BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)
	Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.
	BP5 Variant found in a case with an alternate molecular basis for disease
	BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation
	BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

Appendix Table S3. ACMG Guideline rules for combining criteria to classify sequence variants.

Reproduced from Richards et al. (65) with permission.

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

Supplementary data

Appendix Table S4 Manually curated somatic mutations in *PPP1R13B.*

Version snapshot 5 September 2016, updated to cBio v1.2.4, COSMIC v78, ICGC r23 plus legacy mutations.

Mutation	Residue	Cancer type	Patient/sample number	ASPP1 copy number	Tumour/ cell line	TP53 status
M4L	4	Breast invasive carcinoma	TCGA-A7-A26H	Diploid	T	wt
E19K	19	Lung adenocarcinoma	TCGA-93-8067-01	Gain	T	E285K
P21S	21	Cutaneous melanoma	TCGA-HR-A2OH	N/A	T	wt
P21S	21	Cutaneous melanoma	IPC-298 [CL]		CL	R213.
P24L	24	Stomach adenocarcinoma	PGM38	N/A	T	wt
T27A	27	Colorectal carcinoma	coadread_dfci_2016_115875	N/A	T	wt
R29.	29	Colon adenocarcinoma	587336	N/A	T	R175H
G39E	39	Bladder urothelial carcinoma	TCGA-FD-A3B3	Gain	T	wt
H44R	44	Colon adenocarcinoma	TCGA-A6-3809-01		T	wt
N52splice	52	Melanoma	TCGA-BF-A5EQ-01	N/A	T	wt
R54H	54	Colorectal carcinoma	coadread_dfci_2016_2379	N/A	T	wt
P55_H65del	55-65	Stomach adenocarcinoma	TCGA-CG-4304	Diploid	T	wt
M61I	61	Lung squamous cell carcinoma	TCGA-18-3414	Diploid	T	R273L
E64G	64	Colorectal carcinoma	coadread_dfci_2016_3451	N/A	T	wt
E64K	64	Cutaneous melanoma	TCGA-BF-A3DM	N/A	T	wt
E64K	64	Colorectal carcinoma	coadread_dfci_2016_3611	N/A	T	wt
E64.	64	Lung squamous cell carcinoma	TCGA-63-5131	ShallowDel	T	R158L
E74K	74	Liver cancer	HCC15T		T	wt
E75K	75	Colon carcinoma	SNU-407 [CL]		CL	wt
F79L	79	Lung adenocarcinoma	TCGA-86-8672	ShallowDel	T	F341Efs*7
L80F	80	CSCC	CSCC-40-T		T	G266R, S362fs
R81.	81	Lung adenocarcinoma	TCGA-55-1592-01A	ShallowDel	T	G244C
R81.	81	Esophageal adenocarcinoma	TCGA-L5-A4OP	Diploid	T	S240C
E83K	83	Lymphoid	4166706		T	wt
Q92H	92	Colon carcinoma	HCC2998 [CL]	Gain	CL	R213.
93splice	93	Cutaneous melanoma	SK-MEL-30 [CL]		CL	T284fs
Q96.	96	Endometrial carcinoma	TCGA-D1-A177		T	wt
E99.	99	Lung small cell carcinoma	COR-L311 [CL]		CL	
R101Q	101	Liver carcinoma	9210_T		T	wt
R101Q	101	Endometrial carcinoma	TCGA-BS-A0TC		T	wt

R101.	101	Colon adenocarcinoma	TCGA-AA-3973	Diploid	T	wt
R101.	101	Lung adenocarcinoma (LUAD-S00488)	LUAD-S00488	Diploid	T	wt
R101Q	101	Colon adenocarcinoma	TCGA-AA-A010	Diploid	T	R175C
Q103H	103	Esophageal carcinoma	ESCC_55	N/A	T	wt
R104I	104	Rectum adenocarcinoma	TCGA-EI-6917		T	R213.
R104T	104	Head and neck SCC	TCGA-CV-6961	Diploid	T	wt
E116A	116	Melanoma	TCGA-D9-A6EC	Gain	T	R213.
E116K	116	breast ductal carcinoma	PD24209a		T	wt
188splice	118	Neuroblastoma	CHP-134 [CL]		CL	wt
P122Q	122	Invasive breast cancer	TCGA-A2-A0EM	Diploid	T	wt
P123H	123	Prostate cancer	TCGA-XK-AAIW-01		T	P82L, A74T
T127I	127	Melanoma	MELA-0019		T	P278F
S129.	129	Cervical	TCGA-EK-A2RK-01	Diploid	T	wt
D133N/D130N	133	Pediatric medulloblastoma	ICGC_MB240		T	wt
M134V	134	Thyroid carcinoma	TCGA-DJ-A1QE	Diploid	T	wt
Q145.	145	Cutaneous SCC	CSCC-16-T	N/A	T	Y236N
Q146R	146	Colon carcinoma	I2L-P7-Tumor-Organoid			
E153-splice	153	Lung adenocarcinoma	TCGA-05-4424	Gain	T	L194R
Q154.	154	Desmoplastic melanoma	40M	N/A	T	A86fs
R155C	155	Ovarian carcinoma	OC-314 [CL]		CL	R273H
R155H	155	B-cell lymphoma	WSU-NHL [CL]		CL	R248Q
R155H	155	Colon adenocarcinoma	587228	N/A	T	R280K
R165C	165	Esophageal carcinoma	TCGA-JY-A6FG-01	Gain	T	R273H
R165C	165	Renal carcinoma	BK0073		T	wt
R165H	165	Myeloma	SK-MM-2 [CL]		CL	K132N
E174K	174	Colorectal carcinoma	coadread_dfci_2016_365383	N/A	T	wt
R182Q	182	Endometrial carcinoma	TCGA-AP-A0LM	Diploid	CL	wt
E184.	184	Endometrial carcinoma	TCGA-BS-A0UF	Diploid	T	R213.
N188I	188	Mantle cell lymphoma	JEKO-1 [CL]		CL	P58fs
R194C	194	Cutaneous melanoma	Mewo [CL]		CL	E258K, Q317.
R194C	194	Skin SCC	CSCC-37-T		T	F270I, E286K
R194C	194	Rectum adenocarcinoma	TCGA-EI-6917		T	R213.
D201N	201	Colon carcinoma	TCGA-AY-A54L-01		T	C242fs
I205V	205	Lung adenocarcinoma	HF-23896, 1946219		T	T231del
N209Y	209	Melanoma	Mel-BRAFi-03	N/A	T	wt
G208D	208	Lower grade glioma	TCGA-DU-5874		T	wt
X211_splice	211	Prostate adenocarcinoma	1115161	Diploid	T	wt
A219V	219	Skin melanoma	IPC-298 [CL]		CL	R213.
F221L	221	Stomach adenocarcinoma	TCGA-CG-4466	Diploid	T	P152fs
Q222R	222	Melanoma	MELA-0169		T	wt

L233del	233	Lung SCC	TCGA-63-A5MT	Diploid	T	X187_splice
L233fs	233	Colon adenocarcinoma	TCGA-A6-5665		T	D207G
R234fs	234	Papillary thyroid carcinoma	DO223895		T	D21fs
Q237.	237	Cervical	TCGA-EK-A3GK-01	Diploid	T	wt
Q240K	240	Papillary RCC	TCGA-5P-A9JZ	Diploid	T	wt
K246N	246	Colon carcinoma	SNU-81 [CL]		CL	R213., K132T
G248R	248	CSCC	CSCC-38-T		T	R248W
T261M	261	Renal clear cell carcinoma	TCGA-A3-3365		T	wt
R270T	270	HPV- squamous cell carcinoma from hypopharynx of patient HN_62421 _vs_ matched normal DNA_SM	HN_62421	N/A	T	wt
R278C	278	Cutaneous melanoma	TCGA-D9-A6EC-06	Gain	T	R213.
Q285H	285	Colon adenocarcinoma	TCGA-AA-A010	Diploid	T	R175C
L289F	289	Neuroendocrine prostate cancer	WCMC10362	N/A	T	R110L
R299H	299	Colon adenocarcinoma	TCGA-D5-6928	N/A	T	wt
M301T	301	Endometrial carcinoma	EN [CL]		CL	wt
R309.	309	Esophageal carcinoma	ESCC_BICR_041T		T	wt
R309.	309	Colorectal carcinoma	T469		T	wt
R314C	314	Lymphoblastic leukemia	MOLT-4 [CL]	Diploid	CL	R306.
R314C	314	Colon adenocarcinoma	587338	N/A	T	R213., T125M
R316H	316	Oral cancer - Gingivobuccal	OSCC-GB_011601		T	wt
K321E	321	Breast invasive carcinoma	MRK-nu-1 [CL]		CL	wt
I322fs	322	Ovarian carcinoma	OC-314 [CL]		CL	R273H
I322fs	322	Stomach adenocarcinoma	TCGA-BR-8081	Diploid	T	P301fs
I322fs	322	Liver carcinoma	TCGA-ED-A97K	ShallowDel	T	wt
Q323.	323	CSCC	CSCC-52-T		T	P278S
P333T	333	Mouth carcinoma	OSCC-GB_00660111			
L337M	337	Liver carcinoma	H080217	Diploid	T	L145P
S340L	340	Melanoma	MELA-0193		T	R175H
S340L	340	Colorectal carcinoma	coadread_dfci_2016_3321	N/A	T	wt
S340L	340	Cervical SCC	SW756 [CL]		CL	wt
A344T	344	Rhabdomyosarcoma	SJRH30 [CL]		CL	Y205C, R273C, R280S
V346M	346	CLL	1179	N/A	T	wt
Q351.	351	Cutaneous melanoma	TCGA-EE-A20H	Gain	T	wt
G356R	356	Stomach adenocarcinoma	TCGA-CG-4462	Diploid	T	wt
G356R	356	Rectum adenocarcinoma	TCGA-F5-6863		T	R248W
P364T	364	Liver carcinoma	H090156	Diploid	T	wt
P367S	367	Colon adenocarcinoma	587332	N/A	T	L257Q
S369C	369	Cervical carcinoma	SW756 [CL]		CL	wt
S369P	369	Endometrial carcinoma	TCGA-AP-A0LM	Diploid	T	wt
I372V	372	Lymphoblastic leukemia	KARPAS-45 [CL]		CL	R175H, R273C

S374L	374	Esophageal carcinoma	399		T	Q331H
A376G	376	Osteosarcoma	G-292 Clone A141B1 [CL]			
G379R	379	Pancreas endocrine tumour	ITNET_0797_T		T	wt
G379R	379	Melanoma	C067		T	
S383splice	383	Endometrial carcinoma	TCGA-B5-A11U-01	Diploid	T	wt
G387A	387	Lung squamous cell carcinoma	TCGA-66-2785	Diploid	T	wt
Q394H	394	Colon adenocarcinoma	TCGA-AA-3864	Diploid	T	wt
V400L	400	Anaplastic large cell lymphoma	SU-DHL-1 [CL]		CL	R273H
A406T	406	Colon carcinoma	TCGA-CK-4951-01		T	wt
W410R	410	Stomach adenocarcinoma	PGM63	N/A	T	wt
W410.	410	Colorectal carcinoma	T1772		T	R273C
V415M	415	Colon adenocarcinoma	587220	N/A	T	wt
E416D	416	Prostate carcinoma	TCGA-XK-AAIW-01	Diploid	T	P82L, A74T
S432A	432	Liver carcinoma	TCGA-ED-A7PZ-01	ShallowDel	T	R248fs
S478F	478	CSCC	CSCC-15-T		T	G279E, R280K
E480K	480	Desmoplastic melanoma	Au1	N/A	T	wt
E484K	484	Melanoma	MELA-0035		T	wt
E484K	484	Cutaneous melanoma	TCGA-EE-A2M5	Gain	T	wt
S486T	486	Colon carcinoma	LIM2405 [CL]		CL	wt
S496I	496	Lung adenocarcinoma	LUAD-E01217	Diploid	T	wt
S496I	496	Thyroid cancer	PTC-14C		T	wt
R497Q	497	Prostate carcinoma	PCA78-1		T	wt
P500T	500	Lung squamous cell carcinoma	TCGA-66-2777	ShallowDel	T	R248L
T501I	501	Colorectal carcinoma	coadread_dfci_2016_354	N/A	T	wt
P504L	504	Lung adenocarcinoma	LC_C9		T	R248L
P504S	504	Stomach adenocarcinoma	TCGA-BR-4261		T	wt
T509P	509	Pancreatic ductal carcinoma	ICGC_0503	N/A	T	wt
Q511fs	511	Colorectal carcinoma	coadread_dfci_2016_3446	N/A	T	wt
Q511fs	511	Colon adenocarcinoma	LS174T [CL]		CL	wt
Q511fs	511	Colon adenocarcinoma	GP5d [CL]		CL	wt
Q511fs	511	Colon adenocarcinoma	LS-180 [CL]		CL	wt
Q511fs	511	Colon adenocarcinoma	GP2D [CL]		CL	wt
Q511fs	511	Colon adenocarcinoma	SNU-1040 [CL]		CL	R248W
Q511fs	511	Colon adenocarcinoma	T3446		T	wt
Q511fs	511	Colon adenocarcinoma	T464		T	wt
Q511fs	511	Stomach adenocarcinoma	TCGA-HF-7132	Diploid	T	splice
Q511fs	511	Stomach adenocarcinoma	TCGA-CG-5723	Diploid	T	M169T
Q511fs	511	Stomach adenocarcinoma	TCGA-BR-6852	Diploid	T	R273C
Q511fs	511	Stomach adenocarcinoma	TCGA-BR-8487	Diploid	T	R248W
Q511fs	511	Colon adenocarcinoma	TCGA-AA-3877	Diploid	T	wt

V524I	524	Colorectal carcinoma	coadread_dfci_2016_50	N/A	T	I162N
P525L	525	Colorectal carcinoma	coadread_dfci_2016_92	N/A	T	wt
P525Q	525	Lung SCC	TCGA-18-3419	Gain	T	D259F
Y530S	530	Ovarian cystadenocarcinoma	AOCS-093		T	Y220C
P531L	531	Rectum adenocarcinoma	TCGA-DC-6154		T	R196.
G534R	534	Cutaneous melanoma	TCGA-EE-A2A2	ShallowDel	T	Q317.
D542N	542	Uterine carcinosarcoma	MM18T	N/A	T	R273C, R213.
S543G	543	Colorectal carcinoma	coadread_dfci_2016_1222	N/A	T	wt
P555T	555	AML	CN-AML-NR-08-Dx		T	wt
A558V	558	Colon adenocarcinoma	587224	N/A	T	wt
A558V	558	Renal clear cell carcinoma	TCGA-BP-5176	Diploid	T	wt
G561V	561	Breast invasive carcinoma	TCGA-AR-A24Q	ShallowDel	T	R175H
G561A	561	Glioma	KNS-42 [CL]		CL	R342.
G561A	561	Liver cancer	JHH-2 [CL]		CL	wt
S562.	562	Lung SCC	TCGA-33-6737	Diploid	T	L111Q, T284P
S562fs	562	Lung adenocarcinoma	TCGA-83-5908	Gain	T	C275F
P564S	564	Lymphoblastic leukemia	GR-ST [CL]		CL	M246T
Q565H	565	Lung small cell carcinoma	NCI-H1105 [CL]		CL	R249S
S566F	566	Lung SCC	TCGA-98-8021	Diploid	T	R65Efs *58
S577fs	576	Kidney Renal Clear Cell Carcinoma	TCGA-A3-3322	Diploid	T	wt
K591N	591	Endometrial carcinoma	TCGA-B5-A0JY	Diploid	T	wt
P595Q	595	Prostate adenocarcinoma	MO_1020	ShallowDel	T	V173L
G629W	629	Lung adenocarcinoma	LUAD-NYU315-	Diploid	T	R248P
S632F	632	Lung SCC	TCGA-98-8021	Diploid	T	R65Efs *58
S632P	632	Liver carcinoma	TCGA-CC-A3MA-01	ShallowDel	T	F113C
T633fs	633	Colorectal carcinoma	coadread_dfci_2016_140903			
T633M	633	Colorectal carcinoma	coadread_dfci_2016_391	N/A	T	wt
G634R	634	Esophageal carcinoma	ESO-0133	N/A	T	R306.
S644N	644	Germinal center B-cell derived	4150895		T	wt
Q651H	651	Colorectal carcinoma	coadread_dfci_2016_354	N/A	T	wt
G653D	653	Adult T cell lymphoma-leukaemia (HTLV+)	ATL012		T	V157G
A655F	655	Lung adenocarcinoma	LUAD-2GUGK	Diploid	T	wt
A655V	655	Colorectal carcinoma	coadread_dfci_2016_546	N/A	T	wt
P657S	657	Endometrial carcinoma	TCGA-AP-A0LT	Diploid	T	wt
A658fs	658	Colorectal carcinoma	coadread_dfci_2016_60	N/A	T	R110C, R175C
T662I	662	Cutaneous melanoma	TCGA-EE-A2GO	Gain	T	wt
R668W	668	Lung adenocarcinoma	LC_C23	N/A	T	wt
P672S	672	Liver carcinoma	TCGA-LG-A6GG-01	Diploid	T	wt
L675F	675	Cutaneous melanoma	TCGA-EE-A29R	Diploid	T	wt

A689T	689	Stomach adenocarcinoma	TCGA-BR-6709	ShallowDel	T	wt
A701V	701	Invasive breast cancer	TCGA-BH-A0H3	Diploid	T	wt
A701V	701	Colorectal carcinoma	coadread_dfci_2016_407	N/A	T	wt
A701T	701	Lung adenocarcinoma	TCGA-78-7161	Amp/gain	T	wt
R708C	708	Breast invasive carcinoma	CGP_donor_1964513	N/A		
S710F	710	Melanoma	TCGA-ER-A3PL	ShallowDel	T	wt
E713K	713	Colon adenocarcinoma	TCGA-AD-6964		T	wt
G719R	719	Stomach adenocarcinoma	TCGA-D7-A74A-01	ShallowDel	T	wt
N721T	721	B-cell lymphoma	HT [CL]		CL	V216M
Y727F	727	Papillary RCC	TCGA-4A-A93W	Diploid	T	wt
R729C	729	Prostate adenocarcinoma	MO_1176	Gain	T	W91fs
T740A	740	Colon carcinoma	RKO [CL]		CL	wt
F742L	742	Testicular germ cell tumor	TCGA-XE-A8H5-01	ShallowDel	T	wt
P745S	745	Esophageal carcinoma	ESCC_BICR_041T		T	wt
P747L	747	CSCC	CSCC-38-T		T	R248W
T754I	754	Melanoma	TCGA-EB-A6QZ	N/A	T	wt
G761R	761	Melanoma	MELA-0247		T	R342., G245S, C275S
N762K	762	Mouth SCC	OSCC-GB_00770111			
N766S	766	Head and neck SCC	PCI-15A [CL]		CL	
N766fs	766	Renal clear cell carcinoma	TCGA-A3-3385	Diploid	T	wt
P774fs	774	Colorectal carcinoma	coadread_dfci_2016_3181	N/A	T	V10I
P774fs	774	Colorectal carcinoma	TCGA-G4-6302-01		T	R267Q
P774fs	774	Stomach adenocarcinoma	TCGA-CG-5721	Diploid	T	R273C
P774fs	774	Stomach adenocarcinoma	TCGA-BR-4370	Diploid	T	wt
P780S	780	Esophageal carcinoma	ESCC_37		T	C141fs
A786S	786	Bladder urothelial carcinoma	TCGA-G2-A3IB	Diploid	T	C176F, Q331.
D790N	790	melanoma	Pat_45_B		T	Y103fs* 1
S798Y	798	Malignant melanoma	YUMOKI		T	R280K
E800fs	800	Colon carcinoma	HCT-116 [CL]	Diploid	CL	wt
E800K	800	Synovial sarcoma	SW982 [CL]		CL	wt
E803K	803	Esophageal carcinoma	ZZUFHECRKL-G071T			
Q808K	808	Stomach adenocarcinoma	TCGA-F1-A448-01	Diploid	T	wt
Q812.	812	Melanoma	MELA-0176		T	P27s, P151F
P816L	816	Esophageal carcinoma	TCGA-L5-A8NK	Diploid	T	V197L, R273C
D819E	819	Esophageal carcinoma	TCGA-VR-AA4G-01	ShallowDel	T	S94., G154fs
V824M	824	Endometrial carcinoma	TCGA-D1-A17D	Diploid	T	wt
T826M	826	Liver carcinoma	182T		T	M246V
S835T	835	B-cell lymphoma	SLVL [CL]		CL	
E846D	846	Colorectal carcinoma	TCGA-AZ-4315		T	wt
P849S	849	Desmoplastic melanoma	46M	N/A	T	wt
P855F	855	Head and neck SCC	CSCC-38-T		T	R248W

P855S	855	Head and neck SCC	CSCC-49-T		T	P142S, V216G, E286K
P860S	860	Head and neck SCC	CSCC-20-T		T	R248Q, V272G
P875L	875	Lung SCC	TCGA-85-8350	ShallowDel	T	D207fs
S875L	875	Colon adenocarcinoma	TCGA-CM-6162		T	R290S
R877W	877	DLBC	DLBCL-MAYO_DLBCL_234	N/A	T	wt
T878M	878	Lung adenocarcinoma	SW1573 [CL]		CL	wt
H880Y	880	Stomach adenocarcinoma	TCGA-BR-8059	Diploid	T	wt
G881E	881	Melanoma	MELA-0247		T	C275S
G881R	881	Ewing's sarcoma	ES5 [CL]		CL	R248Q
R883K	883	Head and neck SCC	TCGA-CV-A6K2	Diploid	T	M246I
R885Q	885	Colorectal carcinoma	coadread_dfci_2016_1208	N/A	T	R248Q
R885W	885	Colorectal carcinoma	coadread_dfci_2016_437	N/A	T	V157F
R885W	885	Sarcoma	TCGA-3B-A9HT-01	Diploid	T	W91.
P888H	888	Bladder urothelial carcinoma	B105	N/A	T	wt
L889fs	889	Colorectal carcinoma	coadread_dfci_2016_2564	N/A	T	wt
S896del	896	Prostate adenocarcinoma	TCGA-G9-6496	Diploid	T	wt
E898K	898	Lung adenocarcinoma	TCGA-17-Z049-01		T	P250L
D902N	902	Uterine carcinosarcoma	MM18T	N/A	T	R213., R273C
P914H	914	Endometrial carcinoma	TCGA-D1-A16F	N/A	T	wt
E920K	920	melanoma	Pat_16_B		T	wt
P924fs	924	Colorectal carcinoma	coadread_dfci_2016_2624	N/A	T	wt
H926Y	926	Liver	R308_C01		T	R273C
A928S	928	Lung adenocarcinoma	TCGA-95-8039	Diploid	T	G105D
V929I	929	Liver carcinoma	TCGA-MI-A75G-01	Diploid	T	H193R
I936T	936	Colorectal carcinoma	coadread_dfci_2016_207430	N/A	T	P36Afs*7
V947M	947	Kidney Chromophobe	TCGA-KO-8409	Gain	T	wt
X955_splice	955	Liver	TCGA-G3-A25S-01	ShallowDel	T	wt
A961S	961	SCLC	S02242	N/A	T	H168L
Q972H	972	Rectum adenocarcinoma	TCGA-AG-A002	Diploid	T	wt
S983A	983	Stomach adenocarcinoma	TCGA-CG-5726	Diploid	T	R306.
K994T	994	Colorectal carcinoma	coadread_dfci_2016_352566	N/A	T	wt
I1003N	1003	T-cell leukemia	Jurkat [CL]		CL	R196.
Q1004H	1004	Lung adenocarcinoma	H522 [CL]	Gain	CL	wt
C1005Y	1005	Stomach adenocarcinoma	TCGA-FP-A9TM-01	ShallowDel	T	H179R, R248W
K1015Q	1015	Lung SCC	TCGA-92-7341	Gain	T	wt
M1019I	1019	Cervical SCC	TCGA-C5-A1BL		T	wt
A1024V	1024	Colorectal carcinoma	coadread_dfci_2016_2448	N/A	T	H178fs, E294fs
A1024T	1024	Colon	DLD1 [CL]		CL	wt
F1040V	1040	Stomach adenocarcinoma	pfg043T	N/A	T	wt

D1044E	1044	Cholangiocarcinoma	TCGA-ZH-A8Y4-01	ShallowDel	T	wt
A1045T	1045	Colon carcinoma	sysucc-783T		T	wt
R1051C	1051	Colorectal carcinoma	coadread_dfci_2016_116	N/A	T	D49N
G1065E	1065	Head and neck SCC	BICR_22 [CL]		CL	F109fs, A307V
G1065E	1065	Melanoma	14T		T	
P1084L	1084	Pancreas	8013014	N/A	T	C124.
R1085Q	1085	Bladder urothelial carcinoma	TCGA-DK-A3X1	Amp	T	E271K
R1087Q	1087	Cutaneous melanoma	TCGA-D3-A1Q6	Diploid	T	wt
A1090T	1090	Pancreatic ductal carcinoma	Capan-2 [CL]		CL	wt
A1090T	1090	Stomach adenocarcinoma	TCGA-BR-4370	Diploid	T	wt

Appendix Table S5. Manually curated somatic mutations in *TP53BP2*.

Version snapshot 5 September 2016, updated to cBio v1.2.4, COSMIC v78, ICGC r23 plus legacy mutations.

Mutation	Actual residue	Reported residue	Cancer	Sample number	ASPP2 copy number	Tumour/cell line	TP53 status
translation start site			Colorectal adenocarcinoma	TCGA-AA-3672		T	wt
chr1:g.223988401C>T			melanoma	MELA-0180		T	wt
L12P	6	12	stomach adenocarcinoma	TCGA-BR-8382	Diploid	T	wt
T13I	7	13	cutaneous melanoma	TCGA-EE-A29E		T	L330R
V14L	8	14	Lung small cell carcinoma	631076		T	wt
V14L	8	14	lung SCC	TCGA-39-5021	Gain	T	C141W
Y15C	9	15	Head and neck SCC	TCGA-QK-A8ZA-01	Gain	T	H193L
N19K	13	19	Liver carcinoma	TCGA-DD-A73D-01	Gain	T	wt
I33fs	27	33	malignant melanoma	YUMOBER		T	R209fs
T23I	23	29	melanoma	2492729		T	P191S
T23P	23	29	melanoma	MEL-UKRV-Mel-6	N/A	T	wt
C34Y	28	34	endometrial carcinoma	TCGA-D1-A174-01	Diploid	T	wt
C41R	35	41	colorectal carcinoma	Gp2D [CL]		CL	wt
E43.	37	43	endometrial carcinoma	TCGA-B5-A0JY-01	Diploid	T	wt
P44L	38	44	cutaneous melanoma	TCGA-EE-A3J7		T	S241F
G45S	39	45	T-cell leukemia	Jurkat [CL]		CL	R196.

R60C	54	60	neuroblastoma	MHH-NB-11 [CL]		CL	wt
R60C	54	60	Thymoma	TCGA-ZB-A96E-01	Diploid	T	wt
R60C	54	60	colon adenocarcinoma	TCGA-AA-3994	Gain	T	wt
A63V	57	63	colorectal carcinoma	coadread_dfci_2016_335135		T	R283H
N65K	59	65	colon adenocarcinoma	GP2d [CL]		CL	wt
N65K	59	65	colon adenocarcinoma	GP5d [CL]		CL	wt
R67Q	61	67	thyroid cancer	TT2609-C02 [CL]		CL	
R67Q	61	67	endometrial carcinoma	TCGA-D1-A17Q-01	Diploid	T	wt
R67.	61	67	Liver carcinoma	TCGA-BC-A216-01	AMP	T	R248W
R68.	68	74	colon adenocarcinoma	TCGA-D5-6930		T	wt
R74.	68	74	esophageal carcinoma	ESCC_86		T	wt
D70H	64	70	Lung SCC	TCGA-56-8624-01	Diploid	T	N268Tfs*77
L72F	66	72	cutaneous melanoma	TCGA-EE-A3AF		T	wt
R74Q	68	74	breast cancer	TCGA-AN-A046	AMP	T	E11K
R74Q	68	74	colorectal carcinoma	coadread_dfci_2016_2227	N/A	T	wt
R74Q	68	74	colorectal carcinoma	coadread_dfci_2016_3451	N/A	T	wt
E81K	75	81	colorectal carcinoma	coadread_dfci_2016_2674	N/A	T	R175H
F84L	78	84	B-cell lymphoma	HT [CL]		CL	V216M, R273H
R87L	81	87	esophageal carcinoma	TCGA-LN-A9FP	Gain	T	E294., splice
R90H	84	90	Marginal zone lymphoma	12D		T	wt
P85fs	85	91	stomach adenocarcinoma	TCGA-B7-5816		T	wt
P86fs	86	92	stomach adenocarcinoma	TCGA-HF-A5NB	Gain	T	wt
P86fs, Q103fs	86	92	stomach adenocarcinoma	TCGA-CG-5728	Diploid	T	wt
P86fs	86	92	pancreatic adenocarcinoma	ICGC_0548	N/A	T	S215N
P86fs	86	92	Head and neck SCC	TCGA-F7-A624	Diploid	T	N239D, T256I
P86fs	86	92	breast cancer	TCGA-D8-A1XQ		T	L252fs
P86fs	86	92	colon adenocarcinoma	C135		T	wt
P86fs	86	92	colon adenocarcinoma	TCGA-AZ-6598		T	fs

P86fs	86	92	cervical cancer	C-33-A [CL]		CL	R273C
P86fs	86	92	stomach adenocarcinoma	TGBC11TKB [CL]		CL	R273C
P87fs	87	93	endometrial carcinoma	EN [CL]		CL	R273H
V97M	91	97	Lung SCC	TCGA-56-8624-01	Diploid	T	N268Tfs*77
P100A	94	100	breast lobular carcinoma	CGP_donor_1451422	Missense		
R109I	103	109	endometrial carcinoma	TCGA-B5-A0JY-01	Diploid	T	wt
Y118.	112	118	pancreatic carcinoma	8058178, 8068579	N/A	T	H179Y
G124R	118	124	lung non-small cell cancer	NCI-H2172 [CL]		CL	wt
L134P	128	134	liver carcinoma	HCC079T		T	R249S
A135S	129	135	colon adenocarcinoma	TCGA-CM-6674		T	wt
E133Q	133	139	lung adenocarcinoma	TCGA-MP-A4SW	Gain	T	wt
Q138.	132	138	bladder urothelial carcinoma	TCGA-K4-A3WS-01	Gain	T	wt
S142F	136	142	sarcoma	TCGA-DX-AB2E-01	Diploid	T	R342.
S142Y	136	142	colorectal carcinoma	coadread_dfci_2016_92	N/A	T	wt
S142T	136	142	colorectal carcinoma	2476_PT		T	G244D
S142Y	136	142	endometrial carcinoma	TCGA-AX-A05Z-01	Gain	T	wt
R143H	137	143	glioblastoma	TCGA-28-2506			wt
G145.	139	145	lung non-small cell cancer	NCI-H2228 [CL]		CL	Q331.
Q151R	145	151	neuroblastoma	LAN-6 [CL]		CL	wt
K157Q	157	34	Colon carcinoma	I2L-P26-Tumor-Organoid	Missense		
E159.	153	159	breast ductal carcinoma	PD24194a		T	C176W
F164C	158	164	B-cell lymphoma	BC-1 [CL]		CL	wt
L165F	159	165	lung non-small cell cancer	LU-65 [CL]		CL	E11Q
K166fs	160	166	renal non-clear cell carcinoma	9266826	N/A	T	wt
Q167L	161	167	lung mesothelioma	H2810 [CL]		CL	R273L
R171Q	165	171	endometrial carcinoma	HEC-1 [CL]		CL	R248Q
A177V	171	177	stomach adenocarcinoma	STC252		T	wt
A189V	183	189	DLBC	TCGA-G8-6914-01	Diploid	T	wt

E190.	184	190	cutaneous melanoma	TCGA-EE-A2A6			wt
Q186.	186	192	lung adenocarcinoma	TCGA-49-AARE-01	Gain	T	R273G
K192fs	192	198	stomach adenocarcinoma	TCGA-BR-4298			mut
K195E	189	195	colon adenocarcinoma	KM12 [CL]		CL	R72fs
V199fs	193	199	melanoma	Pat_06_A		T	wt
V199fs	193	199	melanoma	Pat_53_B		T	wt
V199fs	193	199	breast cancer	TCGA-EW-A1IZ-01		T	wt
V199fs	193	199	colon adenocarcinoma	TCGA-AA-3811-01		T	H178fs
V199fs	193	199	Ewing's sarcoma	07-P1079	N/A	T	wt
V199fs	193	199	colon adenocarcinoma	TCGA-A6-6781-01		T	wt
V199fs	193	199	colon adenocarcinoma	TCGA-AA-A01P-01		T	wt
V199.	193	199	colon adenocarcinoma	coadread_dfci_2016_4500	N/A	T	wt
V199.	193	199	breast cancer ER+ Her2-	CGP_donor_1964513	Nonsense		
V199.	193	199	stomach adenocarcinoma	TCGA-FP-A4BE	N/A	T	R158H, R273C
V199.	193	199	stomach adenocarcinoma	TCGA-BR-8078-01	Diploid	T	wt
V199.	193	199	stomach adenocarcinoma	TCGA-BR-8368	Diploid	T	wt
V199.	193	199	stomach adenocarcinoma	TCGA-BR-8382	Diploid	T	wt
V199.	193	199	stomach adenocarcinoma	pfg181T	N/A	T	wt
A201S	195	201	endometrial carcinoma	TCGA-AP-A051-01	Gain	T	wt
V206M	200	206	Prostate carcinoma	TCGA-XK-AAIW	Diploid	T	P82L
G214R	208	214	cutaneous melanoma	TCGA-EE-A3AG	Diploid	T	R213.
Q216H	216	216	small cell lung carcinoma	DMS-79 [CL]		CL	wt
M217I	217	217	tongue SCC	SCC-25 [CL]		CL	R209fs
Q229K	223	229	DLBC	TCGA-G8-6914-01	Diploid	T	wt
Q229K	223	229	colon adenocarcinoma	TCGA-AM-5821		T	
A237D	231	237	endometrial carcinoma	TCGA-D1-A17F-01		T	wt
A237T	231	237	glioblastoma	TCGA-06-5416			R213.

E243.	237	243	endometrial carcinoma	TCGA-BS-A0UF-01	Diploid	T	R213.
E249D	243	249	invasive breast cancer	TCGA-A8-A07I	AMP	T	wt
K252E	246	252	lung adenocarcinoma	M4945	N/A	T	wt
S258N	252	258	endometrial carcinoma	MFE-319 [CL]		CL	Y220C, R273C
H260R	254	260	lymphoid neoplasm	JVM-3 [CL]		CL	wt
A267T	261	267	Hodgkin lymphoma	SUP-HD1 [CL]		CL	wt
E268K	262	268	breast cancer	HCC202 [CL]		CL	T284fs
D270A	264	270	esophageal carcinoma	ESCC_11		T	P152Q, V274L
D270N	264	270	breast cancer	TCGA-D8-A1JA	Gain	T	wt
R271C	265	271	melanoma	Pat_24_A		T	wt
R271C	265	271	colon adenocarcinoma	C13		T	wt
R271H	265	271	colon adenocarcinoma	TCGA-A6-6141		T	wt
Q385E	379	385	breast cancer	TCGA-A8-A08R-01	Gain		N239.
P292S	286	292	CLL	Missense			
L297W	291	297	Prostate cancer (met)	05-165M1_LN, 05-165K5_LUNG	Diploid	T	wt
S302L	296	302	melanoma	COLO-792 [CL]		CL	wt
E303G	297	303	ovarian cystadenocarcinoma	TCGA-23-2649		T	Y234.
R310H	304	310	Acute myeloid leukaemia	CN-AML-CR-33-Dx		T	wt
V311I	305	311	Head and neck SCC	WSU-HN8 [CL]		CL	R196.
K316del	316	316	lower grade glioma	TCGA-HT-7481		wt	
R317W	311	317	pancreatic carcinoma	8		T	wt
W319G	313	319	renal papillary carcinoma	TCGA-B9-4117-01	Diploid	T	wt
W319fs	313	319	melanoma	2492709, 2492710		T	wt
K321N	315	321	oesophageal carcinoma	KYSE-520 [CL]		CL	wt
K328E	322	328	melanoma	YUWAND	N/A	T	H233fs, Q317.
A324V	318	324	Melanoma	TCGA-FS-A4F0-06	Diploid	T	wt
D336N	330	336	Head and neck SCC	TCGA-CR-6481	Diploid	T	wt
N338S	332	338	esophageal carcinoma	ESCC_142		T	V274F
P340S	334	340	stomach adenocarcinoma	TCGA-D7-A4YX-01	Diploid	T	wt
A340T	340	346	colon adenocarcinoma	TCGA-AY-6197		T	R175H
A344V	338	344	Ewing's sarcoma	ES8 [CL]		CL	C135F

A344V	338	344	colon adenocarcinoma	SNU-1040 [CL]		CL	R248W, V73fs, D48N
S354F	354	231	malignant melanoma	ME009T	N/A	T	wt
Y356C	350	356	ovarian cancer	OV-56 [CL]		CL	K101fs
Q358L	352	358	liver carcinoma	041T		T	wt
P363L	357	363	melanoma	MELA-0054		T	wt
P363L	357	363	pancreatic carcinoma	8065126, ICGC_0217	N/A	T	F134C
S367.	360	367	Bladder cancer	DFCI-04_R	N/A	T	D207.
P360S	360	237	melanoma	C084		T	wt
K374fs	368	374	papillary renal carcinoma	TCGA-A4-A772-01	Diploid	T	wt
P375S	369	375	colon adenoma	61			R110C
D379G	373	379	endometrial carcinoma	TCGA-D1-A17Q-01	Diploid	T	wt
I384T	378	384	Liver carcinoma	CHC304		T	in-frame ins
I384V	378	384	stomach adenocarcinoma	TCGA-HU-A4GU-01	Diploid	T	P191del
Q385E	379	385	invasive breast cancer	TCGA-A8-A08R	Gain	T	N239.
P390S	384	390	T-cell leukemia	Jurkat [CL]		CL	R196.
N398K	392	398	malignant melanoma	YUMOBER		T	R209fs
A403V	397	403	melanoma	Pat_53_B		T	wt
S410Y	404	410	stomach adenocarcinoma	TCGA-BR-8680	Gain	T	wt
G409D	403	409	Mouth SCC	201, OSCJM-PT43-201-T	N/A	T	Y126.
X409_splice	403	409	melanoma	TCGA-D3-A51F-06	Diploid	T	wt
P414L	408	414	Melanoma	TCGA-D9-A4Z3-01	N/A	T	wt
P422Q	422	428	liver carcinoma	H081665	Diploid	T	wt
P428L	422	428	Head and neck SCC	CSCC-41-T		T	Q317.
P428S	422	428	Neuroendocrine prostate tumour	WCMC139_1_C	Gain	T	wt
A442T	436	442	esophageal carcinoma	OCCAMS-SH-020		T	wt
D447N	441	447	Cervical SCC	TCGA-EK-A3GM-01	Gain	T	wt
P452S	446	452	cutaneous melanoma	MELA-0165		T	wt
E455D	449	455	Lung SCC	TCGA-34-A5IX-01	Gain	T	A159V
K459N	453	459	Head and neck SCC	TCGA-CV-7089	Diploid	T	Y163C
R461C	455	461	Colon carcinoma	coadread_dfci_2016_335135	N/A	T	R283H
R455C	455	332	esophageal carcinoma	ESO-717			wt
R461C	455	461	liver carcinoma	HuH-7 [CL]		CL	Y220C

R461C	455	461	colon adenocarcinoma	TCGA-AA-3949	Diploid	T	R273H
R461H	455	461	melanoma	Pat_66_A		T	wt
R461H	455	461	stomach adenocarcinoma	TCGA-BR-4361	Diploid	T	R273H, H168R
P462L	456	462	Colon carcinoma	TCGA-AG-3728-01		T	H193N
P456Q	456	462	liver carcinoma	H081665	Diploid	T	wt
F457L	457	334	T-cell leukemia	PF-382 [CL]		CL	R248Q, R273C
S464L	458	464	breast cancer	TCGA-AC-A23H	Gain	T	wt
D467H	461	467	Colon carcinoma	I2L-P11-Tumor-Biopsy	Missense		
A468T	462	468	endometrial carcinoma	TCGA-D1-A176-01		T	wt
N473S	467	473	kidney carcinoma	RCC-MF [CL]		CL	wt
S477F	471	477	sarcoma	TCGA-DX-A8BP-01	Diploid	T	C135R
R482S	476	482	gall bladder carcinoma	TGBC24TKB [CL]		CL	C242fs
G479D	473	479	Head and neck SCC	TCGA-BA-A6DJ	Diploid	T	wt
S486G	482	486	Colon carcinoma	coadread_dfci_2016_1221	N/A	T	wt
L491F	485	491	bladder cancer	MSKCC-0741_R	N/A	T	Y236C, R282W
R492Q	486	492	skin cutaneous melanoma	TCGA-EE-A20B	Gain	T	wt
D493A	487	493	lymphoblastic leukemia	KARPAS-45 [CL]		CL	R175H, T125M, R272C
D493E	487	493	skin cutaneous melanoma	TCGA-EE-A3J3	Gain	T	wt
V504A	498	504	colon adenocarcinoma	TCGA-AA-3815-01		T	R158H, G245S
P507H	501	507	pancreatic acinar cell carcinoma	ACINAR09	n/a	T	wt
P509S	503	509	melanoma	TCGA-FR-A726-01	N/A	T	wt
T510R	504	510	lymphoblastic T-cell leukemia	P12-ICHIKAWA [CL]	Missense		
K513fs	513	513	prostate adenocarcinoma	TCGA-M7-A724-01	ShallowDel	T	wt
D523H	523	523	NS	STS-0421 [CL]		CL	wt
P526S	520	526	esophageal carcinoma	TCGA-L5-A8NV-01	Gain	T	R175H, R196.
P527fs	521	527	thyroid cancer	PTC_14	N/A	T	D21fs
P527L	521	527	Prostate carcinoma	EOPC-01_tumor		T	wt
P527R	521	527	cutaneous melanoma	ML_24		T	wt
S540L	534	540	prostate adenocarcinoma	TCGA-HC-7233	Diploid	T	wt

A553T	547	553	liver carcinoma HBV+	HCC005		T	wt
A553T	547	553	stomach adenocarcinoma	TCGA-CG-4305	Diploid	T	wt
R558K	552	558	cutaneous melanoma	TCGA-FS-A1ZD	Gain	T	wt
S561L	561	567	small cell lung carcinoma	2334188	N/A	T	X331_splice
S567L	561	567	Desmoplastic melanoma	11M	N/A	T	wt
Q570K	564	570	renal clear cell carcinoma	TCGA-AK-3460	Diploid	T	wt
A586V	580	586	Prostate carcinoma	MO_1263	Diploid	T	V157F
A587del	571	587	Head and neck SCC	CSCC-49-T		T	P142S, V216G, E286K
V588I	582	588	colorectal carcinoma	coadread_dfci_2016_3319	N/A	T	wt
R589W	583	589	esophageal carcinoma	ESCC-123T	N/A	T	I255S
S596F	590	596	pancreatic carcinoma	ASHPC_0029_Pa_P		T	W91.
D598Y	592	598	breast cancer	CGP_donor_1478543	Missense		
V610M	604	610	B-cell Acute Lymphoblastic Leukemia	TARGET-10-PAPNNX-04A-01D	Missense		
V610M	604	610	Pancreatic cancer	TCGA-HZ-A9TJ-01	Gain	T	R175H
S617F	611	617	melanoma	ML_92		T	wt
T620M	614	620	colon adenocarcinoma	SK-CO-1 [CL]		CL	wt
A624V	618	624	pancreatic adenocarcinoma	PCSI_0477		T	wt
G626E	620	626	Uterine carcinomsarcoma	MM04T	N/A	T	P223S
F629L	623	629	Liver carcinoma (virus+)	HX32		T	wt
N628fs	622	628	stomach adenocarcinoma	TCGA-D7-A6EY-01	Diploid	T	wt
N628fs	622	628	colon adenocarcinoma	I2L-P19Ta-Tumor-Biopsy	Frame shift		
F629L	623	629	liver carcinoma	HX32T		T	wt
A632V	626	632	endometrial carcinoma	TCGA-AP-A051-01	Gain	T	wt
Q634H	628	634	DLBC	TCGA-G8-6914-01	Diploid	T	wt
A636fs	630	636	prostate adenocarcinoma	TCGA-CH-5769-01	Diploid	T	wt
A636T	630	636	endometrial carcinoma	TCGA-D1-A167-01	Diploid	T	wt
P645T	639	645	endometrial carcinoma	ESS-1 [CL]		CL	R213.

G652S	646	652	bladder urothelial carcinoma	TCGA-GC-A3BM	Gain	T	wt
G652A	646	652	lung adenocarcinoma	TCGA-95-7039	Gain	T	R273L
Q660H	654	660	pancreatic carcinoma	PDA_006	Diploid	T	wt
Q660R	654	660	liver carcinoma	TCGA-CC-A8HV-01	Gain	T	R249S
E667.	661	667	breast cancer HER2amp	PI066		T	V216M
E667.	661	667	breast cancer	CGP_donor_1589431	N/A	T	V216M
S671F	665	671	melanoma	TCGA-EE-A29N	Diploid	T	wt
S671F	665	671	melanoma	TCGA-EE-A2MT	Gain	T	wt
G678D	672	678	Colon carcinoma	coadread_dfci_2016_354	N/A	T	wt
P686S	680	686	Melanoma	TCGA-FS-A4F0-06	Diploid	T	wt
P700L	694	700	cutaneous melanoma	TCGA-DA-A1HV	Diploid	T	R280K
R701Q	695	695	bladder urothelial carcinoma	TCGA-K4-A3WV	Diploid	T	wt
P705R	699	705	small cell lung carcinoma	585205	N/A	T	Y163C
S713F	707	713	malignant melanoma	C008		T	wt
S713Y	707	713	breast cancer	TCGA-A2-A04Y-01		T	wt
R717Q	711	717	Colon carcinoma	coadread_dfci_2016_2365	N/A	T	wt
S717.	711	717	colon adenocarcinoma	sysucc-274T		T	wt
L724P	718	724	Colon carcinoma	1517_CLM		T	A276V
E725D	719	725	Colon carcinoma	coadread_dfci_2016_2641	N/A	T	R273C
A726T	720	726	endometrial carcinoma	TCGA-BG-A0M4	Gain	T	wt
R728Q	722	728	Colon carcinoma	coadread_dfci_2016_2365	N/A	T	wt
S732C	726	732	Head and neck SCC	TCGA-BB-4223	Diploid	T	wt
K733N	733	739	colon adenocarcinoma	587376	N/A	T	wt
S737C	737	743	breast cancer	PD4601a	N/A	T	wt
S737Y	737	743	colorectal carcinoma	coadread_dfci_2016_593	N/A	T	wt
T745A	739	745	lung non-small cell cancer	NCI-H2342 [CL]		CL	Y220C
E742D	742	619	melanoma	ME048	N/A	T	wt
I744V	738	744	thyroid cancer	PTC_448		T	wt
P750L	744	750	Desmoplastic melanoma	40M	N/A	T	A86fs
L758P	752	758	Colon carcinoma	coadread_dfci_2016_116	N/A	T	D49N

Y754C	754	760	stomach adenocarcinoma	TCGA-BR-7715	Diploid	T	F270S
T764I	758	764	T-cell leukemia	Jurkat [CL]		CL	R196.
A766V	760	766	stomach adenocarcinoma	PGM65	N/A	T	wt
M768I	762	768	colon adenocarcinoma	TCGA-AA-3984	Diploid	T	wt
E769K	763	769	glioblastoma	P10_Rec	N/A	T	R273C
S772C	766	772	bladder urothelial carcinoma	TCGA-DK-A3X1	ShallowDel	T	E271K
P774S	768	774	melanoma	Pat_63_B		T	wt
S775.	769	769	bladder urothelial carcinoma	TCGA-DK-A3X1	ShallowDel	T	E271K
S775.	769	775	breast cancer	TCGA-E9-A5FL	Diploid	T	H193R
Y776C	770	776	Colon carcinoma	coadread_dfci_2016_683	N/A	T	wt
P777A	771	777	liver carcinoma	TCGA-CC-A7IJ-01	DeepDel	T	V157F
E792D	786	792	lung adenocarcinoma	LUAD-B01970	Diploid	T	G154V
P796S	790	796	melanoma	MELA-0209		T	wt
E801S	795	801	Head and neck SCC	TCGA-UF-A7JO-01	Diploid	T	R213L
E801K	795	801	Head and neck SCC	CSCC-10-T		T	Q104., R249S
E803K	797	803	stomach adenocarcinoma	NUGC-3 [CL]		CL	Y220C
E803K	797	803	endometrial carcinoma	TCGA-AP-A059-01	Diploid	T	R273C, S240G, P64T
V807A	801	807	endometrial carcinoma	TCGA-B5-A11E	Gain	T	R342.
P810S	810	810	melanoma	MEL-13473	N/A	T	wt
V819M	813	819	T-cell leukemia	DND-41 [CL]		CL	R248Q
S827C	821	827	liver carcinoma	TCGA-MI-A75E-01	Gain	T	wt
P824L	824	701	melanoma	2427TIL			
P830L	824	830	Neuroendocrine prostate tumour	WCMC170_1_C	Diploid	T	wt
A831T	825	831	lung small cell cancer	SHP-77 [CL]		CL	C176W
D837A	831	837	colon adenocarcinoma	587376	N/A	T	wt
P840L	834	840	malignant melanoma	YUSMI		T	R248W
G842R	836	842	stomach adenocarcinoma	RERF-GC-1B [CL]		CL	R213.
P844Q	838	844	Liver carcinoma (HBV+)	HCC013		T	K132T

S847N	841	847	malignant melanoma	YUKAT		T	wt
E855K	849	855	bladder urothelial carcinoma	TCGA-DK-A3IT	Gain	T	wt
P859A	853	859	Renal carcinoma	C0020T		T	wt
H863R	857	863	stomach adenocarcinoma	MKN45 [CL]		CL	wt
E870D	864	870	papillary renal carcinoma	TCGA-B1-A656-01	Diploid	T	wt
Y872C	866	872	esophageal carcinoma	ESO-859	N/A	T	wt
Y872S	866	872	stomach adenocarcinoma	BJ-GC-9538803		T	F109S
P873L	867	873	renal clear cell carcinoma	TCGA-BP-4164	Diploid	T	wt
Y875.	869	875	T cell lymphoma-leukaemia (adult) HTLV+	ATL043		T	wt
P879T	873	879	esophageal carcinoma	TCGA-LN-A49K-01	Amp	T	V122Dfs*26
P878fs	872	878	colon adenocarcinoma	LS174T [CL]		CL	wt
P878fs	872	878	colon adenocarcinoma	LS-180 [CL]		CL	wt
P875L	875	752	skin SCC	cSCCP1			R248W
P881S	875	881	cutaneous melanoma	TCGA-EE-A2GP	Diploid	T	R248W
S882F	876	882	Head and neck SCC	CSCC-6-T		T	R196., P278T
P885L	879	885	renal non-clear cell carcinoma	9259814	N/A	T	wt
G883R	883	883	multiple myeloma	MM-0633	N/A	T	wt
P888fs	882	888	cervical SCC	TCGA-DR-A0ZM			wt
R896C	890	896	Colon carcinoma	coadread_dfci_2016_2423	Missense		
R896H	890	896	pancreatic adenocarcinoma	TCGA-IB-7651	Gain	T	wt
R896C	890	896	lung adenocarcinoma	TCGA-50-5944	Gain	T	E285.
P898fs	898	898	lung carcinoma	TCGA-60-2711-01	ShallowDel	T	L265fs
G902W	896	902	Lung SCC	TCGA-37-A5EM	Gain	T	R174W
Q903.	897	903	colon adenocarcinoma	LS-411N [CL]		CL	Y126.
R915H	909	915	colon adenocarcinoma	T2932		T	wt
R915H	915	921	CLL	1100	N/A	T	wt

A923T	917	923	colon adenocarcinoma	T3024	N/A	T	wt
A923T	917	923	stomach adenocarcinoma	TCGA-CG-5733	Diploid	T	wt
H924R	918	924	colon adenocarcinoma	587224	N/A	T	P152L
A934fs	928	934	Bile duct carcinoma	BD72T		T	R248Q
A934fs	928	934	colon adenocarcinoma	LS-411N [CL]		CL	Y126.
S939L	933	939	Colon carcinoma	coadread_dfci_2016_2936	N/A	T	wt
D946G	940	946	liver carcinoma (virus+)	RK245		T	wt
D946G	940	946	pancreatic adenocarcinoma	PCSI_0348		T	wt
R950I	944	950	endometrial carcinoma	TCGA-B5-A0JY-01	Diploid	T	wt
E954K	948	952	small cell lung carcinoma	S02375	N/A	T	wt
D956H	950	956	ovarian cystadenocarcinoma	TCGA-24-0966	AMP	T	G245V
S959I	953	959	pancreatic carcinoma	8016470		T	G245S
T967M	961	967	prostate adenocarcinoma	TCGA-KK-A7B3	Diploid	T	wt
A975T	969	975	colon adenocarcinoma	TCGA-CK-4951-01		T	wt
A975V	969	975	esophageal carcinoma	3N25-VS-3T25		T	wt
V975F	975	975	Head and neck SCC	JHU-028 [CL]		CL	wt
Q986.	980	986	colon adenocarcinoma	TCGA-AZ-6598		T	C229fs
N990S	984	990	small cell lung carcinoma	S02295	N/A	T	wt
N990T	984	990	AML	CN-AML-CR-45-Dx		T	wt
G998E	992	998	Head and neck SCC	CSCC-42-T		T	wt
999splice	993	999	Lung SCC	TCGA-77-A5G3	Gain	T	G245C, R248L
H1003Y	997	1003	bladder urothelial carcinoma	TCGA-H4-A2HQ	Diploid	T	wt
S1007A	1001	1007	esophageal carcinoma	430		T	R282W
C1014R	1008	1014	breast cancer	BR-V-027	N/A	T	Y126_splice
F1016L	1010	1016	B-cell lymphoma	SU-DHL-10 [CL]		CL	splice
A1023T	1017	1023	B-cell leukemia	REH [CL]		CL	wt

S1030I	1024	1030	rectum adenocarcinoma	TCGA-AG-3598	Diploid	T	R175H
D1031E	1025	1031	esophageal carcinoma	TCGA-LN-A8I0-01	Gain	T	V272M, R273H
E1043K	1037	1043	lung mesothelioma	H513 [CL]		CL	N239fs
F1046C	1046	923	colon adenocarcinoma	587376		T	wt
M1063I	1057	1063	pancreatic acinar cell carcinoma	ACINAR26	N/A	T	S90T
A1070G	1064	1070	lung adenocarcinoma	TCGA-55-1594	Diploid	T	S166.
A1070V	1064	1070	stomach adenocarcinoma	TCGA-CD-5801	Diploid	T	wt
A1070V	1064	1070	pancreatic adenocarcinoma	8016470	N/A	T	G245S
A1070V	1064	1070	colon adenocarcinoma	coadread_dfci_2016_4536	N/A	T	R175H
A1070V	1064	1070	colon adenocarcinoma	TCGA-AA-3811	Diploid	T	H178fs
W1072.	1066	1072	malignant melanoma	YUCHIME		T	E286K
P1083L	1077	1083	melanoma	MZ7-mel [CL]		CL	wt
M1084I	1078	1084	Cervical SCC	TCGA-IR-A3LA-01	Diploid	T	L194F, E271K
E1086del	1080	1086	stomach adenocarcinoma	TCGA-HU-A4GX-01	Diploid	T	P72S
E1086del	1080	1086	stomach adenocarcinoma	TCGA-BR-8368	Diploid	T	wt
G1087.	1081	1087	stomach adenocarcinoma	TCGA-BR-6452	Diploid	T	wt
C1089F	1083	1089	Colon carcinoma	T368	N/A	T	wt
M1084I	1084	1090	liver carcinoma	H090798	Diploid	T	wt
M1090V	1084	1090	uterine leiomyosarcoma	SK-UT-1 [CL]		CL	R175H, R248Q
I1093M	1087	1093	breast cancer	TCGA-AR-A24N	Gain	T	wt
E1098K	1092	1098	Prostate cancer	TCGA-EJ-7782-01		T	wt
E1102K	1096	1102	endometrial carcinoma	TCGA-BS-A0UF	Diploid	T	R213.
W1104R	1098	1104	endometrial carcinoma	EN [CL]		CL	R273H
W1105C	1099	1105	Lung SCC	TCGA-56-8082	Diploid	T	E180_S185del
A1106V	1100	1106	colorectal carcinoma	sysucc-CRC-1370		T	wt
A1106V	1100	1106	Prostate carcinoma	00-010N_PROSTATE	Diploid	T	R273C
R1107H	1101	1107	Prostate carcinoma	SC_9081	Diploid	T	wt

N1109S	1103	1109	Lung SCC	TCGA-22-1005-01	Diploid	T	wt
N1109I	1103	1109	pancreatic cancer	6T, 6P2-1,2		T	S241F
V1115F	1109	1115	B cell lymphoma	QC2-35-T2		T	H168R
R1117	1111	1117	Colon carcinoma	coadread_dfci_2016_2441	N/A	T	wt
G1121splice	1115	1121	endometrial carcinoma	TCGA-AP-A056		T	wt
R1129I	1123	1129	rectum adenocarcinoma	TCGA-AG-A002	Diploid	T	wt
Q1130E	1124	1130	liver carcinoma	TCGA-DD-A4NN-01	Gain	T	wt

Appendix Table S6. Manually curated somatic mutations in *PPP1R13L.*

Version snapshot 5 September 2016, updated to cBio v1.2.4, COSMIC v78, ICGC r23 plus legacy mutations.

Mutation	Actual residue	Reported residue	Cancer	Sample number	Tumour/cell line	TP53 status
M1.	1	1	Lung adenocarcinoma	LUAD-B02594	T	V157F
E4K	4	4	Breast carcinoma	CGP_donor_1988714	T	wt
A9V	9	9	DLBCL	DLBCL-MAYO_DLBCL_234	T	wt
R10W	10	10	colon adenoma	61	T	R110C
D11N	11	11	Thyroid carcinoma	TCGA-DJ-A4UT	T	wt
M15V	15	15	Renal clear cell carcinoma	TCGA-CJ-4916	T	wt
Q18.	18	18	Melanoma	YUBER	T	G361E
A36V	36	36	Uterine carcinosarcoma	MM12T	T	R248Q, R342.
A36T	36	36	Melanoma	Pat_45_B	T	Y103fs*1
A36V	36	36	Thyroid carcinoma	TCGA-EL-A3T7	T	wt
K44E	44	44	Colorectal cancer	coadread_dfci_2016_354	T	wt
Q45fs, P90fs, P691fs	45	45	Thyroid carcinoma	PTC_14	T	D21fs
S48L	48	48	Cutaneous melanoma	MELA-0165	T	wt
A55T	55	55	Ewing's sarcoma	12-P412	T	wt
A55V	55	55	Stomach adenocarcinoma	TCGA-CG-4442	T	wt
G58fs	58	58	Cutaneous SCC	CSCC-52-T	T	P278S
G62V	62	62	Lung adenocarcinoma	TCGA-50-5049	T	E289.
P64L	64	64	Cutaneous melanoma	MELA-0035	T	wt
P64L	64	64	Cutaneous melanoma	TCGA-EE-A2M5	T	wt

R66M	66	66	Lymphoblastic leukemia	KARPAS-45 [CL]	CL	R175H, R273C
S74L	74	74	Stomach adenocarcinoma	TCGA-BR-4370-01	T	wt
I75S	75	75	Lung carcinoma	TCGA-58-8390-01	T	K132N
86ins-fs	86	86	Stomach adenocarcinoma	TCGA-BR-A4QL	T	wt
S81I	81	81	Colon carcinoma	CRC-33T	T	G245S
R86Q	86	86	Uterine endometrioid carcinoma	TCGA-BG-A0VZ	T	wt
R86W	86	86	Head and neck SCC	TCGA-CV-7568	T	P278S
A88V	88	88	Stomach adenocarcinoma	pfg016T	T	wt
G92A	92	92	PCPG	TCGA-QR-A70Q-01	T	wt
G92C	92	92	Melanoma	ME043	T	wt
G92C	92	92	Uterine endometrioid carcinoma	TCGA-BS-A0TJ	T	wt
H107Y	107	107	Colon carcinoma	CRC-31T, Patient-4693207	T	R306.
H107Q	107	107	Hairy cell leukemia	Mo-T [CL]	CL	wt
P108S	108	108	Melanoma	YUPAT	T	wt
S110N	110	110	Invasive breast cancer tumor	TCGA-A7-A0DA	T	I195T
P111L	111	111	Colorectal cancer	T3090	T	wt
P114fs	114	114	Stomach adenocarcinoma	TCGA-CD-A4MJ	T	wt
S120L	120	120	Esophageal carcinoma	TCGA-LN-A4MQ	T	C238_N239delinsY
S120P	120	120	Colon adenocarcinoma	587350	T	wt
R122C	122	122	Breast carcinoma	CGP_donor_1503133	T	wt
R122H	122	122	Embryonal Tumor (Neuroblastoma)	NBL32	T	wt
T123I	123	123	Cutaneous SCC	CSCC-27-T	T	H179Y, P80S
G133S	133	133	Colorectal cancer	coadread_dfci_2016_60	T	R110C, R175C
G133S	133	133	Colorectal cancer	coadread_dfci_2016_2197	T	G244C
R144H	144	144	Colon adenocarcinoma	587336	T	R175H
R144S	144	144	Liver carcinoma HBV+	ICC009	T	wt
A145T	145	145	Colon adenocarcinoma	T2944	T	K382fs
G148F	148	148	NSCLC	TCGA-55-7913-01	T	splice
A149T	149	149	Colon adenocarcinoma	587278	T	P152fs
T173M	173	173	Uterine carcinosarcoma	MM06T	T	wt
R180C	180	180	Colorectal cancer	coadread_dfci_2016_2624	T	wt
R180C	180	180	Stomach adenocarcinoma	TCGA-F1-6874	T	F113L

P188L	188	188	Cutaneous melanoma	TCGA-EE-A2M5	T	wt
E191K	191	191	Liver cancer	CHC892T	T	wt
G192E	192	192	Liver cancer	CHC892T	T	wt
Q194fs	194	194	Bile duct carcinoma	BD124T	T	G245S
R205P	205	205	Glioblastoma	TCGA-76-4925	T	R249T
P207fs	207	207	Liver carcinoma	BCM723	T	wt
P207fs	207	207	Colorectal cancer	T3021	T	wt
P207fs	207	207	Colorectal cancer	T3724	T	R267W
P207T	207	207	Stomach adenocarcinoma	TCGA-CG-5728	T	wt
A213T	213	213	Stomach adenocarcinoma	TCGA-BR-7723	T	splice
A215T	215	215	Colorectal cancer	coadread_dfci_2016_2271	T	wt
G219R	219	219	Colorectal cancer	coadread_dfci_2016_224049	T	wt
G219W	219	219	Lung SCC	TCGA-43-6143	T	H214R
G224V	224	224	Lung adenocarcinoma	TCGA-55-7995-01	T	H179Y
A231T	231	231	Colon carcinoma	SNU-1040 [CL]	CL	R248W
A236T	236	236	Colon adenocarcinoma	587284	T	S90fs
R243C	243	243	Colon carcinoma	sysucc-1370T	T	wt
R243C	243	243	Esophageal SCC	386	T	M246V
R243C	243	243	Colon adenocarcinoma	TCGA-AA-3864	T	wt
R243H	243	243	Colorectal cancer	coadread_dfci_2016_302124	T	wt
R243H	243	243	Stomach adenocarcinoma	TCGA-CD-A4MI	T	wt
R244Q	244	244	Lung SCC	TCGA-60-2722	T	H179R
P246L	246	246	Cutaneous melanoma	TCGA-EE-A2MR	T	wt
P246L	246	246	Cutaneous melanoma	MELA-0247	T	C275S, R342.
P247S/L	247	247	Sarcoma	TCGA-QC-A7B5	T	E286K
E252V	252	252	Lung adenocarcinoma	TCGA-99-8028	T	C242Y
E252V	252	252	Lung SCC	TCGA-NC-A5HJ-01	T	wt
D254N	254	254	Lung SCC	TCGA-66-2793	T	H179L
D254Y	254	254	Lung SCC	TCGA-77-8144	T	H214R, E294.
L255Q	255	255	Lung SCC	TCGA-34-8455	T	R249S
D256Y	256	256	Ovarian cystadenocarcinoma	TCGA-13-2071	T	TR155del
A258V	258	258	Esophageal carcinoma	ESO-732	T	R273C
E260K	260	260	Liver carcinoma	CHC1545	T	wt
K262Q	262	262	Colon carcinoma	CL-40 [CL]	CL	R248Q
Y269C	269	269	Pleural mesothelioma	NYU517	T	wt

R271C	271	271	Bladder urothelial carcinoma	TCGA-DK-A1A3	T	Q192., C242F
R271L	271	271	Lung NSCLC	TCGA-55-8506	T	G154V
V274I	274	274	Colorectal cancer	T3603	T	C238Y
F275V	275	275	Head and neck carcinoma	WSU-HN6	T	H179L
P281L	281	281	Cutaneous SCC	CSCC-55-T	T	S127F
L283V	283	283	Head and neck carcinoma	BICR22 [CL]	CL	F109fs, A307V
W288.	288	288	Lung SCC	TCGA-22-4599	T	G245C
W288R	288	288	Melanoma	TCGA-EE-A29L-06	T	P177_C182del
E290K	290	290	Breast carcinoma	TCGA-AR-A0TX	T	Y205C
L293F	293	293	Skin melanoma	MEL-Ma-Mel-48	T	
D294H	294	294	Breast carcinoma	TCGA-AR-A0TX	T	Y205C
G297W	297	297	Prostate carcinoma	PR-2915	T	wt
G298fs	298	298	Esophagogastric cancer	TCGA-BR-8487-01	T	R248W
G298fs	298	298	Gall bladder carcinoma	BD236	T	wt
G298fs	298	298	Colorectal cancer	SW48 [CL]	CL	wt
P317del	317	317	Prostate carcinoma	SC_9008	T	R273C
R322Q	322	322	Colorectal cancer	T3658	T	Y126C
A336T	336	336	Colon carcinoma	HCT116 [CL]	CL	wt
R344C	344	344	Uterine carcinosarcoma	MM12T	T	R248Q, R342.
W346.	346	346	small intestine adenoma	61	T	R110C
S357P	357	357	SCLC	S02342	T	R158L
R362C	362	362	Melanoma	3466TIL	T	
P365L	365	365	Colon carcinoma	sysucc-1370T	T	wt
P371L	371	371	Melanoma	TCGA-D9-A6EC-06	T	R213.
W383L	383	383	Lung adenocarcinoma	ZA6505	T	wt
S388I	388	388	Pancreatic carcinoma	8053200	T	wt
L397fs	397	397	Stomach adenocarcinoma	TCGA-BR-4362	T	wt
P403R	403	403	Liver cancer	TCGA-CC-A5UE	T	R249S
Q440fs	440	440	Colon adenocarcinoma	TCGA-G4-6586	T	Nonsense
P448S	448	448	Cutaneous melanoma	TCGA-GF-A3OT	T	wt
E451K	451	451	Melanoma	ML_98	T	wt
G452E	452	452	Melanoma	MELA-0253	T	wt
K455T	455	455	melanoma	2492700, 2492701, 2492702, 2492703 (primary+mets)	T	C275W
K455T	455	455	Desmoplastic melanoma	Au2	T	C275W

K455fs	455	455	Colorectal cancer	C135 [CL]	CL	wt
P456L	456	456	Colorectal cancer	coadread_dfci_2016_603	T	wt
P456L	456	456	Pancreatic carcinoma	QGP-1 [CL]	CL	wt
P457L	457	457	Cutaneous SCC	CSCC-32-T	T	Y107.
P457S	457	457	Cutaneous SCC	CSCC-35-T	T	H179Y
E463Q	463	463	Head and neck SCC	TCGA-DQ-5625	T	wt
P464L	464	464	Bile duct carcinoma	BD53T	T	wt
P464L	464	464	Endometrial carcinoma	TCGA-D1-A16R	T	wt
E467V	467	467	Liver carcinoma HBV+	HCC058	T	wt
P472T	472	472	Renal clear cell carcinoma	C0074T	T	wt
V479M	479	479	CLL	ICGC_CLL-194-TD	T	
E481V	481	481	Lung non-small-cell carcinoma	NCI-H2342 [CL]	CL	Y220C
L488fs	488	488	Stomach adenocarcinoma	TCGA-CG-4305-01	T	wt
P490S	490	490	Cutaneous melanoma	TCGA-DA-A1HY	T	wt
P490S	490	490	Cutaneous melanoma	TCGA-EE-A3J5	T	R213Q
T491M	491	491	Colon adenocarcinoma	587224	T	P152L
Q494R	494	494	Lung NSCLC	TCGA-77-8130-01	T	K305.
P499L	499	499	Colorectal cancer	coadread_dfci_2016_60	T	R110C, R175C
E500K	500	500	Sarcoma	TCGA-QC-A7B5	T	E286K
E506fs	506	506	Cutaneous SCC	CSCC-31-T	T	A138V, W146., R110C
E509A	509	509	Stomach adenocarcinoma	SK-GT-2 [CL]	CL	
V510G	510	510	Thyroid carcinoma	PTC_10	T	wt
R512L	512	512	Lung adenocarcinoma	TCGA-62-A46P	T	wt
R512W	512	512	Colorectal cancer	coadread_dfci_2016_3249	T	wt
A515V	515	515	Myeloma	OPM-2 [CL]	CL	wt
P518S	518	518	Skin melanoma	MEL-Ma-Mel-65	T	Y126 splice
A530V	530	530	Lower grade glioma	TCGA-QH-A65Z	T	wt
P536S	536	536	Uterine endometrioid carcinoma	TCGA-AP-A059	T	R273C, S240G
P537S	537	537	Cutaneous melanoma	TCGA-EE-A2MI	T	wt
R549H	549	549	Uterine endometrioid carcinoma	TCGA-B5-A0K9	T	wt
R553C	553	553	Cutaneous melanoma	TCGA-EE-A29B	T	wt
R553H	553	553	Colorectal cancer	coadread_dfci_2016_2406	T	R282W
R553H	553	553	Astrocytoma	P151	T	wt
R553H	553	553	Neuroblastoma	CHP-134 [CL]	CL	wt

G556E	556	556	Cutaneous SCC	CSCC-29-T	T	R248W
G561R	561	561	B-cell lymphoma	DB [CL]	CL	R248.
G561fs	561	561	Breast carcinoma	CGP_donor_1347732	T	wt
G561fs	561	561	Stomach adenocarcinoma	TCGA-HU-A4G8	T	wt
P562fs	562	562	PAAC	ACINAR03	T	wt
P562fs	562	562	Endometrial carcinoma	EN [CL]	CL	wt
E565Q	565	565	Lung squamous cell carcinoma	TCGA-22-5478	T	R280I
S567F	567	567	Colon carcinoma	COLO-741 [CL]	CL	K321fs
E571D	571	571	Prostate carcinoma	TCGA-J9-A52C	T	wt
E574K	574	574	Desmoplastic melanoma	40M	T	A86fs
P589S	589	589	Cutaneous melanoma	TCGA-FW-A3R5	T	P152L, P191L
P589S	589	589	Cutaneous melanoma	TCGA-EE-A3AE	T	R213.
P589S	589	589	Cutaneous melanoma	MELA-0075	T	wt
P591L	591	591	Cutaneous SCC	CSCC-54-T	T	W23., R248W
P591S	591	591	Colon adenocarcinoma	TCGA-D5-6928	T	wt
P593S	593	593	Melanoma	MELA-0241	T	wt
S596R	596	596	Uterine endometrioid carcinoma	TCGA-A5-A0R8	T	wt
P598L	598	598	Cutaneous melanoma	C086	T	E286.
P602A	602	602	Lung squamous cell carcinoma tumors	TCGA-66-2778	T	mutated TSS
606splice	606	606	Lung adenocarcinoma	LUAD-RT-S01707	T	E285K
S612Q	612	612	Stomach adenocarcinoma	TCGA-HF-A5NB-01	T	wt
S616F	616	616	Thyroid carcinoma	TCGA-DO-A2HM	T	wt
P617S	617	617	melanoma	Pat_41_B	T	R267G
K619E	619	619	Papillary RCC	TCGA-UZ-A9PL	T	wt
R622H	622	622	Cutaneous SCC	CSCC-4-T	T	wt
A623T	623	623	Colorectal cancer	coadread_dfci_2016_3187	T	wt
A623T	623	623	Colorectal cancer	coadread_dfci_2016_197	T	wt
A623T	623	623	Colorectal cancer	coadread_dfci_2016_267922	T	P278H
A623V	623	623	Stomach adenocarcinoma	pfg057T	T	wt
V629M	629	629	Pancreatic carcinoma	TCGA-IB-7651	T	wt
L632del	631	631	Stomach adenocarcinoma	TCGA-CD-A4MJ	T	wt
A634V	634	634	Head and neck SCC	TCGA-F7-A624-01	T	N239D, T256I
A635V	635	635	Colon adenocarcinoma	TCGA-AZ-6598	T	V10I, T81S, S96P, R158C, C229fs

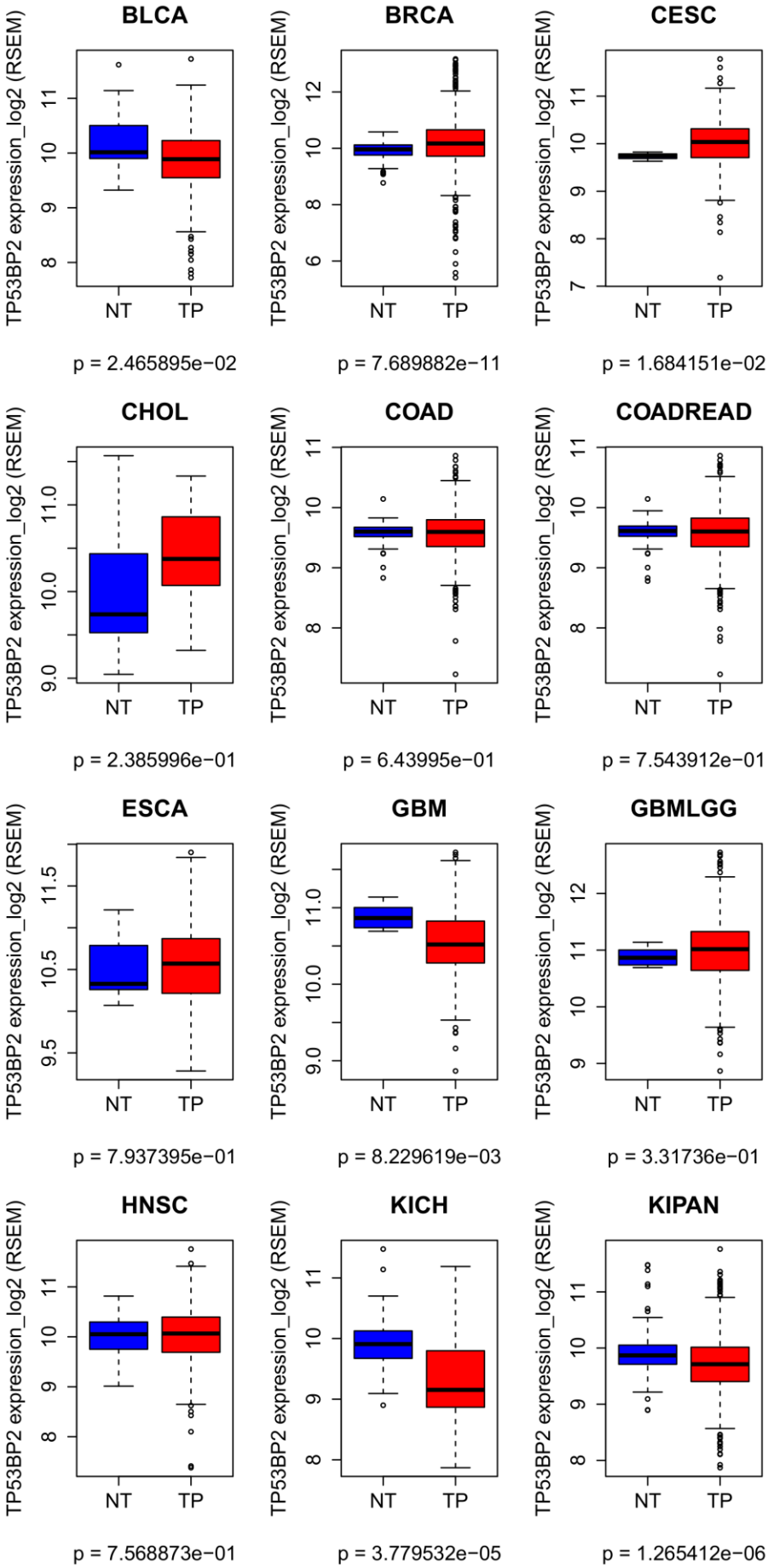
Q644E	644	644	Adenoid cystic carcinoma	ACYC-MDA_AC05	T	wt
K648N	648	648	Stomach adenocarcinoma	TCGA-BR-8487-01	T	R248W
X650splice	650	650	Head and neck SCC	TCGA-CQ-7063-01	T	H179N
D652N	652	652	Uterine endometrioid carcinoma	TCGA-D1-A16X	T	wt
A671V	671	671	Glioma	P21_Rec	T	R181P, R273C, P85S
Y673F	673	673	Lung squamous cell carcinoma	TCGA-39-5027	T	F270I
S674F	674	674	Breast carcinoma	TCGA-A2-A0YK	T	wt
A682T	682	682	Breast carcinoma	TCGA-AO-A128	T	R342.
N685S	685	685	Liver cancer	CHC1191T	T	wt
D690N	690	690	Uterine endometrioid carcinoma	TCGA-D1-A17D	T	wt
H692Y	692	692	Lung NSCLC	TCGA-43-8115-01	T	Y163C
G693S	693	693	Lung adenocarcinoma	TCGA-55-8506-01	T	G154V
A700V	700	700	Prostate carcinoma	MO_1179	T	R282W
A700V	700	700	Colon adenocarcinoma	TCGA-AA-3492	T	P301fs
S702L	702	702	Colon carcinoma	CRC-11T	T	R248Q
Q714P	714	714	Esophageal adenocarcinoma	OCCAMS-SA-013	T	Y205C
A717T	717	717	Colorectal cancer	coadread_dfci_2016_1231	T	wt
F720S	720	720	Chronic myeloid leukemia	RPMI-8866 [CL]	CL	wt
T723M	723	723	Prostate carcinoma	TCGA-XK-AAIW	T	A74T, P82L
T723M	723	723	Colon carcinoma	I2L-P19Tb-Tumor-Biopsy	T	R273C, Q331R
A730T	730	730	DLBCL	DLBCL-Ls1899	T	wt
A730T	730	730	Stomach adenocarcinoma	TGBC11TKB [CL]	CL	R273C, K382fs
A730T	730	730	Uterine endometrioid carcinoma	TCGA-B5-A0JY	T	wt
D735N	735	735	Colorectal cancer	coadread_dfci_2016_189255	T	wt
D750splice	750	750	Melanoma	TCGA-OD-A75X	T	H179Y
D750splice	750	750	Melanoma	TCGA-D9-A6E9	T	wt
M758I	758	758	Lung adenocarcinoma	TCGA-55-A493	T	R273S
G761R	761	761	Lymphoblastic leukemia	P30-OHK [CL]	CL	wt
G761R	761	761	Cutaneous melanoma	TCGA-EE-A3AC	T	wt
Y764H	764	764	Stomach adenocarcinoma	TCGA-BR-8487	T	R248W
A771S	771	771	Liver cancer	CHC1568T	T	H193L, E286K
F773C	773	773	Lung squamous cell carcinoma	LK-2 [CL]	CL	V272M

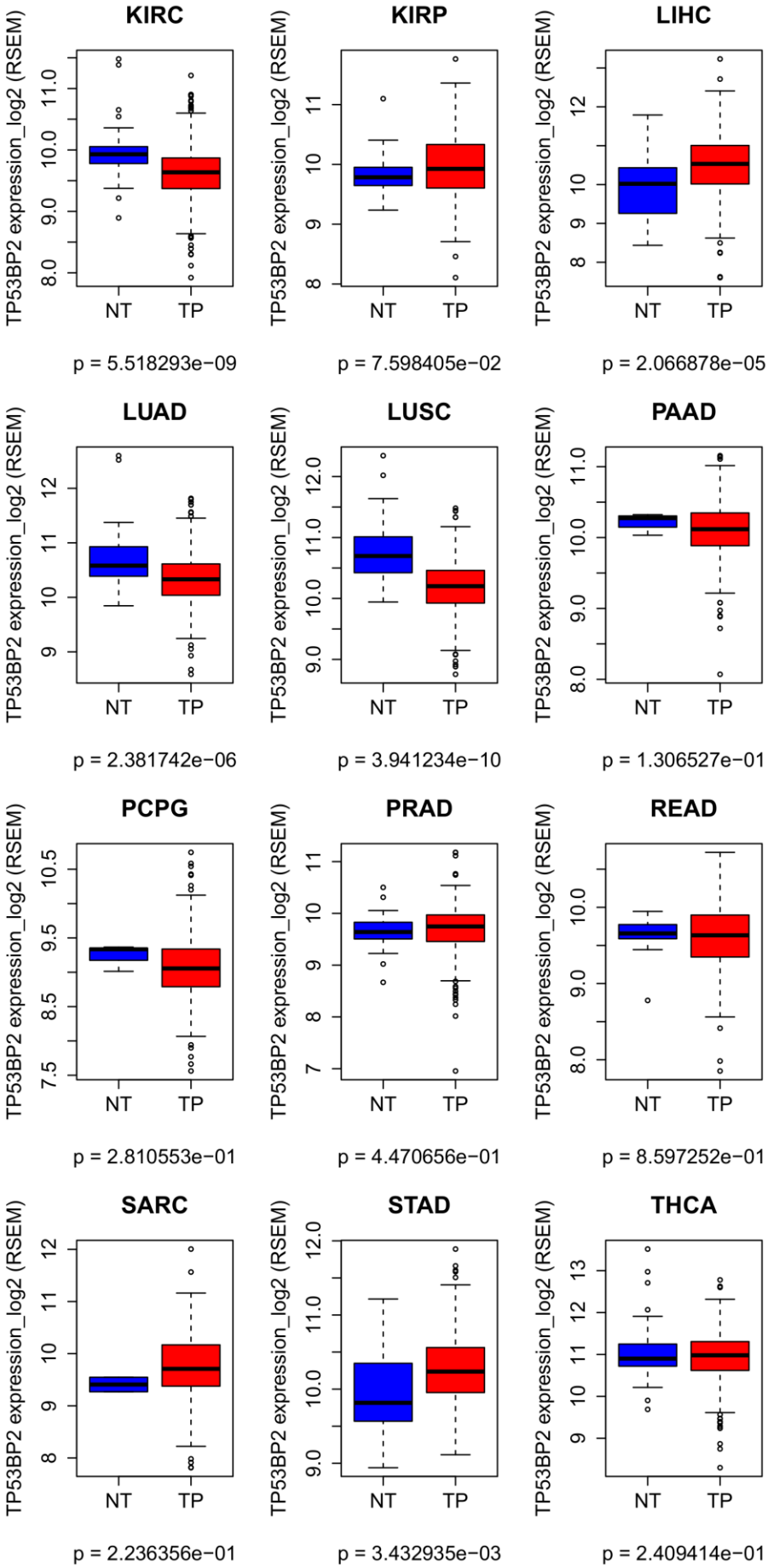
G774W	774	774	Lung small cell carcinoma	NCI-H187 [CL]	CL	S241C
S778C	778	778	Lung adenocarcinoma	TCGA-55-7907	T	R158P
F779L	779	779	Breast carcinoma	TCGA-AC-A23H	T	wt
R780H	780	780	Esophageal carcinoma	SM-4AX86	T	wt
T786I	786	786	Head and neck SCC	TCGA-D6-6516	T	wt
V787M	787	787	Multiple myeloma	MM-0038	T	wt
R789W	789	789	Uterine carcinosarcoma	MM12T	T	R248Q, R342.
G792R	792	792	Colorectal cancer	T3225	T	wt
E794K	794	794	Melanoma	MEL-Ma-Mel-19	T	wt
E795K	795	795	Melanoma	YULAPE	T	wt
T796I	796	796	DLBCL	DLBCL-LS4619	T	wt
W799C	799	799	Cutaneous SCC	T2197	T	G244C
W799.	799	799	Lung adenocarcinoma	TCGA-38-4628	T	wt
W800.	800	800	Melanoma	MEL-HO [CL]	CL	wt
A802V	802	802	Lung adenocarcinoma	TCGA-93-A4JP	T	E258K
R812L	812	812	Kidney renal carcinoma	SN12C [CL]	CL	E336.
R812W	812	812	Colorectal cancer	T3152	T	wt
R812W	812	812	Stomach adenocarcinoma	TCGA-BR-8059	T	wt
R812W	812	812	Lung adenocarcinoma	TCGA-55-8615	T	wt
Y814H	814	814	Lip carcinoma	1604875	T	wt
Y814.	814	814	Lung SCC	TCGA-58-8387	T	A189Pfs*58

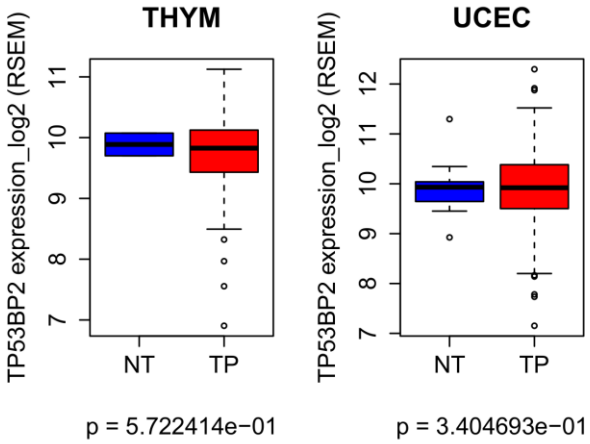
Appendix Table S7. Top brain regions by *PPP1R13B* expression (human brain)

Age	Region with the highest *PPP1R13B* expression	Top level structure	ASPP1 level
8 pcw	upper (rostral) rhombic lip	neural plate	2.5409
9 pcw	dorsal thalamus	thalamus	2.6319
12 pcw (3 samples)	dorsal thalamus	thalamus	2.62-2.86
13 pcw	mediodorsal nucleus of thalamus	thalamus	3.328
13 pcw	cerebellum	neural plate	2.91
16 pcw	mediodorsal nucleus of thalamus	thalamus	3.22

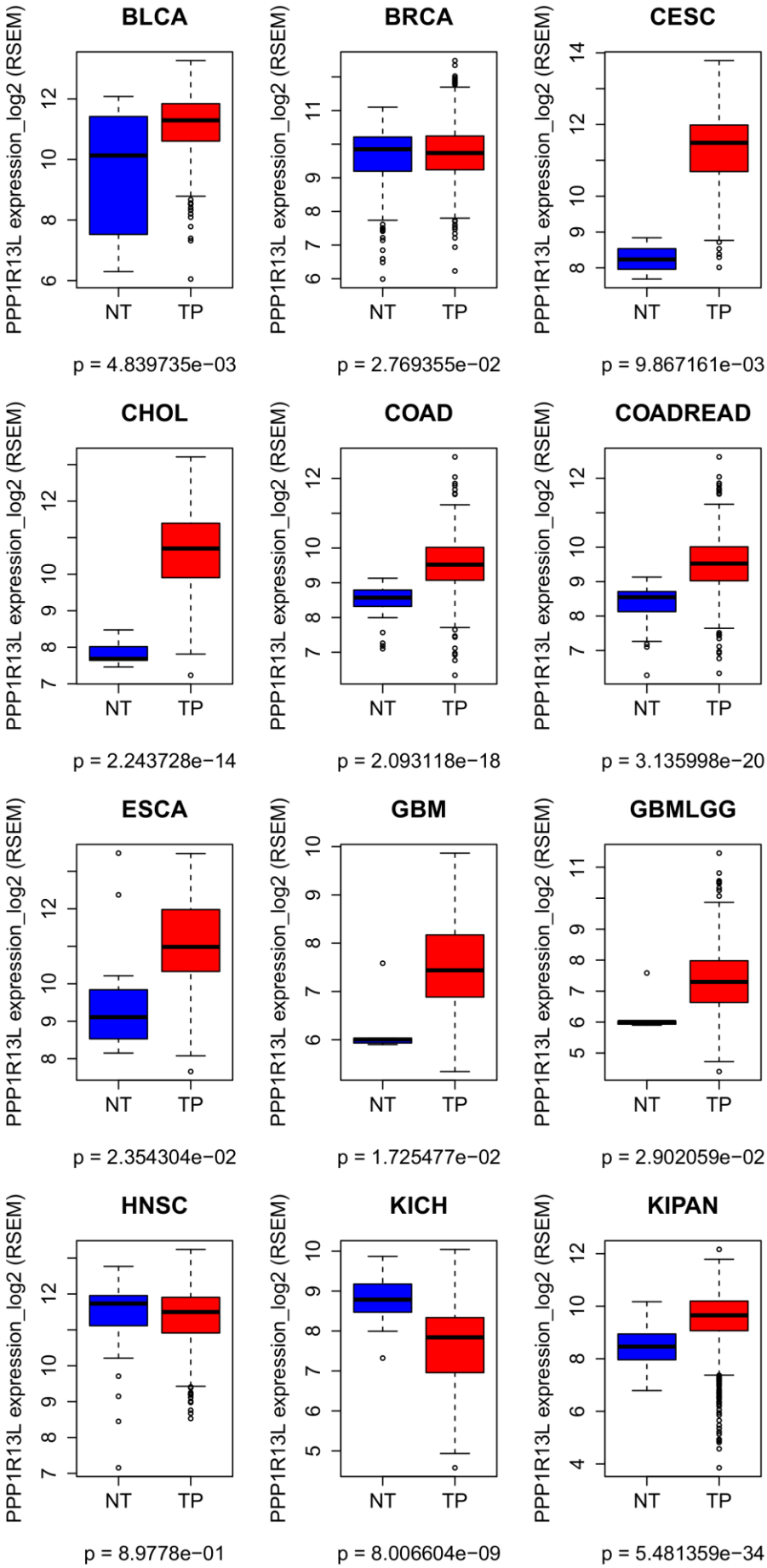
17 pcw	striatum	basal nuclei (basal ganglia)	3.25
19 pcw	mediodorsal nucleus of thalamus	thalamus	3.10
21 pcw	striatum	basal nuclei (basal ganglia)	3.15
24 pcw	primary somatosensory cortex (area S1, areas 3,1,2)	neural plate	2.67
37 pcw	primary auditory cortex (core)	neural plate	2.94
4m (2 samples)	primary visual cortex (striate cortex, area V1/17)	neural plate	3.07; 2.99
10m	cerebellar cortex	Neural plate	3.77
1 yr	primary visual cortex (striate cortex, area V1/17)	Neural plate	3.18
2 yr	primary somatosensory cortex (area S1, areas 3,1,2)	Neural plate	2.20
3 yr (2 samples)	cerebellar cortex	Neural plate	3.39; 3.11
4 yr	cerebellar cortex	Neural plate	2.70
8 yr (2 samples)	cerebellar cortex	Neural plate	3.30
11 yr	cerebellar cortex	Neural plate	2.99
13 yr	cerebellar cortex	Neural plate	3.19
15 yr	cerebellar cortex	Neural plate	3.13
18 yr	cerebellar cortex	Neural plate	2.79
19 yr	cerebellar cortex	Neural plate	3.28
21 yr	cerebellar cortex	Neural plate	3.20
23 yr	cerebellar cortex	Neural plate	3.27
30 yr	cerebellar cortex	Neural plate	3.18
36 yr	cerebellar cortex	Neural plate	3.31
37 yr	cerebellar cortex	Neural plate	3.20
40 yr	primary visual cortex (striate cortex, area V1/17)	Neural plate	2.84

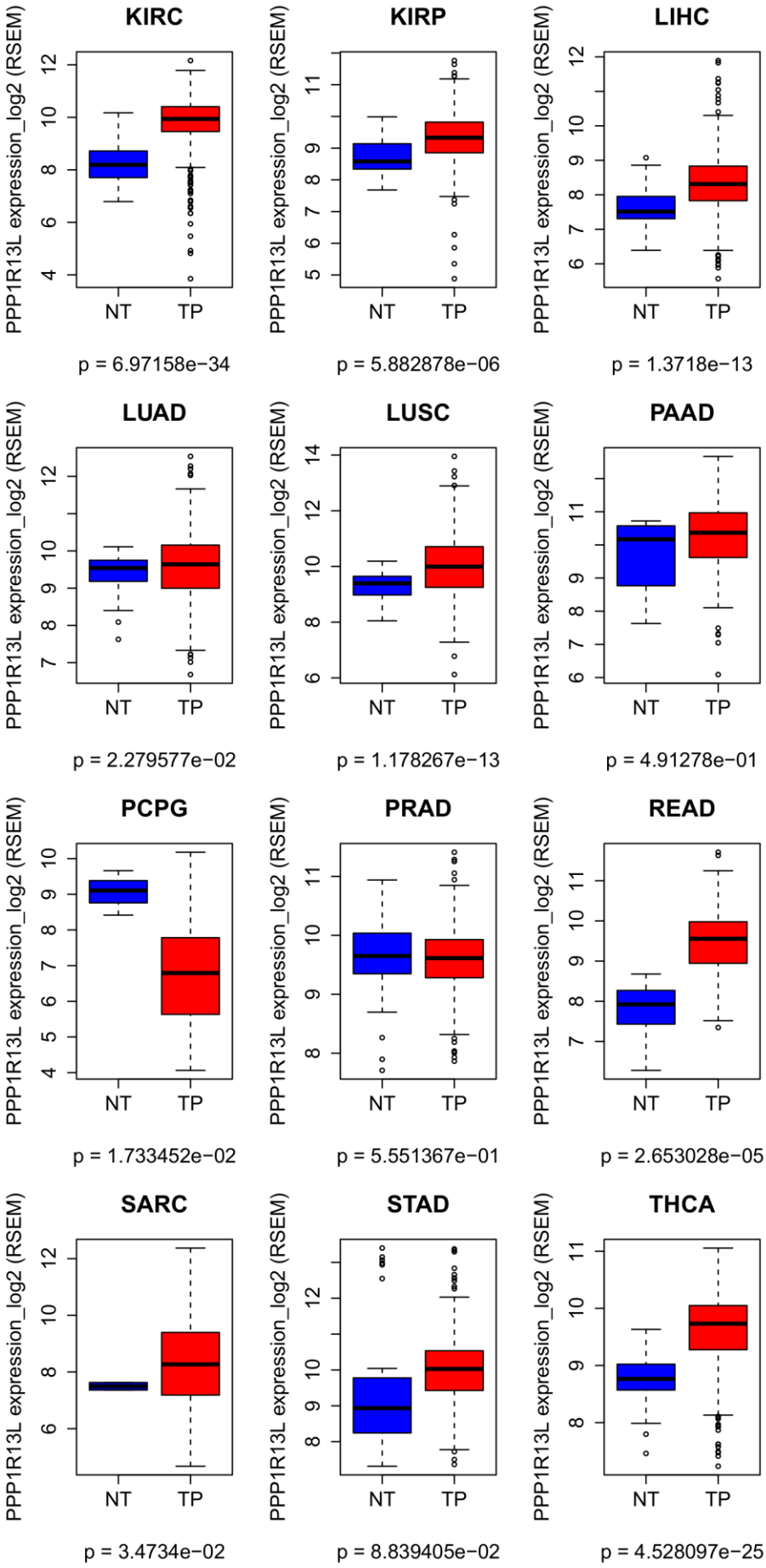

BLCA
TP53BP2 expression_log2 (RSEM)
NT
TP
p = 2.465895e−02
BRCA
p = 7.689882e−11
CESC
p = 1.684151e−02
CHOL
p = 2.385996e−01
COAD
p = 6.43995e−01
COADREAD
p = 7.543912e−01
ESCA
p = 7.937395e−01
GBM
p = 8.229619e−03
GBMLGG
p = 3.31736e−01
HNSC
p = 7.568873e−01
KICH
p = 3.779532e−05
KIPAN
p = 1.265412e−06

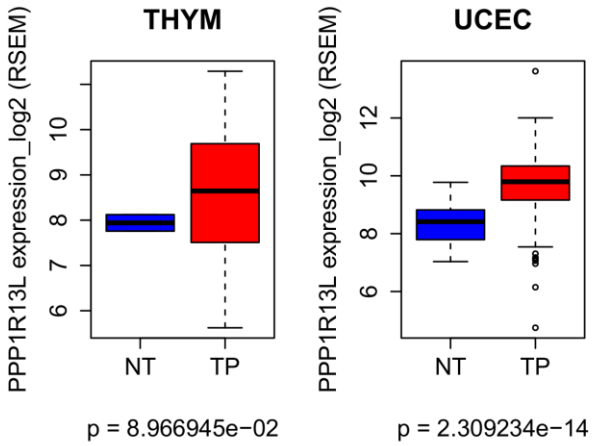

KIRC
TP53BP2 expression_log2 (RSEM)
8.0
9.0
10.0
11.0
NT
TP
p = 5.518293e−09
KIRP
TP53BP2 expression_log2 (RSEM)
8
9
10
11
NT
TP
p = 7.598405e−02
LIHC
TP53BP2 expression_log2 (RSEM)
8
9
10
12
NT
TP
p = 2.066878e−05
LUAD
TP53BP2 expression_log2 (RSEM)
9
10
11
12
NT
TP
p = 2.381742e−06
LUSC
TP53BP2 expression_log2 (RSEM)
9.0
10.0
11.0
12.0
NT
TP
p = 3.941234e−10
PAAD
TP53BP2 expression_log2 (RSEM)
8.0
9.0
10.0
11.0
NT
TP
p = 1.306527e−01
PCPG
TP53BP2 expression_log2 (RSEM)
7.5
8.5
9.5
10.5
NT
TP
p = 2.810553e−01
PRAD
TP53BP2 expression_log2 (RSEM)
7
8
9
10
11
NT
TP
p = 4.470656e−01
READ
TP53BP2 expression_log2 (RSEM)
8.0
9.0
10.0
NT
TP
p = 8.597252e−01
SARC
TP53BP2 expression_log2 (RSEM)
8
9
10
11
12
NT
TP
p = 2.236356e−01
STAD
TP53BP2 expression_log2 (RSEM)
9.0
10.0
11.0
12.0
NT
TP
p = 3.432935e−03
THCA
TP53BP2 expression_log2 (RSEM)
9
10
11
12
13
NT
TP
p = 2.409414e−01



Appendix Figure S8. *TP53BP2* expression TP vs NT from Firehose data run 2016_01_28.


BLCA
PPP1R13L expression_log2 (RSEM)
NT
TP
p = 4.839735e−03
BRCA
p = 2.769355e−02
CESC
p = 9.867161e−03
CHOL
p = 2.243728e−14
COAD
p = 2.093118e−18
COADREAD
p = 3.135998e−20
ESCA
p = 2.354304e−02
GBM
p = 1.725477e−02
GBMLGG
p = 2.902059e−02
HNSC
p = 8.9778e−01
KICH
p = 8.006604e−09
KIPAN
p = 5.481359e−34


KIRC
KIRP
LIHC
LUAD
LUSC
PAAD
PCPG
PRAD
READ
SARC
STAD
THCA
PPP1R13L expression_log2 (RSEM)
NT
TP
p = 6.97158e−34
p = 5.882878e−06
p = 1.3718e−13
p = 2.279577e−02
p = 1.178267e−13
p = 4.91278e−01
p = 1.733452e−02
p = 5.551367e−01
p = 2.653028e−05
p = 3.4734e−02
p = 8.839405e−02
p = 4.528097e−25



Appendix Figure S9. *PPP1R13L* expression TP vs NT from Firehose data run 2016_01_28.

REFERENCES

1. Rossier BC, Baker ME, Studer RA. Epithelial Sodium Transport and Its Control by Aldosterone: The Story of Our Internal Environment Revisited. Physiological Reviews. 2014;95(1):297.

2. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009;41(6):666-76.

3. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. 2011;43(4):333-8.

4. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet. 2008;40(2):161-9.

5. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide Association Analysis of Coronary Artery Disease. New England Journal of Medicine. 2007;357(5):443-53.

6. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences. 2009;106(23):9362-7.

7. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53.

8. Singleton AB, Hardy J, Traynor BJ, Houlden H. Towards a complete resolution of the genetic architecture of disease. Trends in Genetics. 2010;26(10):438-42.

9. Lifton RP. Molecular Genetics of Human Blood Pressure Variation. Science. 1996;272(5262):676.

10. Lifton RP, Gharavi AG, Geller DS. Molecular Mechanisms of Human Hypertension. Cell. 2001;104(4):545-56.

11. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. Nature reviews Genetics. 2015;16(5):275-84.

12. Monroe CL, Dahiya S, Gutmann DH. Dissecting Clinical Heterogeneity in Neurofibromatosis Type 1. Annual Review of Pathology: Mechanisms of Disease. 2016.

13. Lindor NM, McMaster ML, Lindor CJ, Greene MH. Concise Handbook of Familial Cancer Susceptibility Syndromes - Second Edition. JNCI Monographs. 2008;2008(38):3-93.

14. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.

15. The Deciphering Developmental Disorders S. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015;519(7542):223-8.

16. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature. 2014;506(7487):179-84.

17. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. Science. 2012;337(6090):64.

18. Nelson MR, Wegmann D, Ehm MG, Kessner D, St. Jean P, Verzilli C, et al. An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People. Science. 2012;337(6090):100.

19. Keinan A, Clark AG. Recent Explosive Human Population Growth Has Resulted in an Excess of Rare Genetic Variants. Science. 2012;336(6082):740.

20. The UKKC. The UK10K project identifies rare variants in health and disease. Nature. 2015;526(7571):82-90.

21. Weischenfeldt J, Symmons O, Spitz F, Korbel JO. Phenotypic impact of genomic structural variation: insights from and for human disease. Nat Rev Genet. 2013;14(2):125-38.

22. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010;464(7289):704-12.

23. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004;36(9):949-51.

24. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-Scale Copy Number Polymorphism in the Human Genome. Science. 2004;305(5683):525.

25. The Genomes Project C. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.

26. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526(7571):75-81.

27. Evrony Gilad D, Lee E, Mehta Bhaven K, Benjamini Y, Johnson Robert M, Cai X, et al. Cell Lineage Analysis in Human Brain Using Endogenous Retroelements. Neuron. 2015;85(1):49-59.

28. Erwin JA, Paquola ACM, Singer T, Gallina I, Novotny M, Quayle C, et al. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. Nat Neurosci. 2016;advance online publication.

29. Jacobs P, Court Brown WM, Baikie AG, Strong JA. THE SOMATIC CHROMOSOMES IN MONGOLISM. The Lancet. 1959;273(7075):710.

30. LEJEUNE J TR, GAUTIER M. Chromosomic diagnosis of mongolism. Arch Fr Pediatr. 1959;16:962-3.

31. Jacobs PA, Matsuura JS, Mayer M, Newlands IM. A cytogenetic survey of an institution for the mentally retarded: I. Chromosome abnormalities. Clinical Genetics. 1978;13(1):37-60.

32. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong Association of De Novo Copy Number Mutations with Autism. Science. 2007;316(5823):445.

33. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural Variation of Chromosomes in Autism Spectrum Disorder. The American Journal of Human Genetics. 2008;82(2):477-88.

34. Stefansson H, Rujescu D, Cichon S, Pietilainen OPH, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008;455(7210):232-6.

35. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet. 2008;40(7):880-5.

36. Malhotra D, McCarthy S, Michaelson Jacob J, Vacic V, Burdick Katherine E, Yoon S, et al. High Frequencies of De Novo CNVs in Bipolar Disorder and Schizophrenia. Neuron. 2011;72(6):951-63.

37. Georgieva L, Rees E, Moran JL, Chambert KD, Milanova V, Craddock N, et al. De novo CNVs in bipolar affective disorder and schizophrenia. Human Molecular Genetics. 2014.

38. Malhotra D, Sebat J. CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics. Cell. 2012;148(6):1223-41.

39. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, et al. Relative Burden of Large CNVs on a Range of Neurodevelopmental Phenotypes. PLoS Genet. 2011;7(11):e1002334.

40. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511(7509):344-7.

41. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. Nat Genet. 2011;43(9):838-46.

42. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature. 2014;505(7483):361-6.

43. Kirov G, Rees E, Walters JTR, Escott-Price V, Georgieva L, Richards AL, et al. The Penetrance of Copy Number Variations for Schizophrenia and Developmental Delay. Biological Psychiatry. 2014;75(5):378-85.

44. Shaffer LG, Theisen A, Bejjani BA, Ballif BC, Aylsworth AS, Lim C, et al. The discovery of microdeletion syndromes in the post-genomic era: review of the methodology and characterization of a new 1q41q42 microdeletion syndrome. Genet Med. 2007;9(9):607-16.

45. Peca J, Feliciano C, Ting JT, Wang W, Wells MF, Venkatraman TN, et al. Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. Nature. 2011;472(7344):437-42.

46. Yi F, Danko T, Botelho SC, Patzke C, Pak C, Wernig M, et al. Autism-associated SHANK3 haploinsufficiency causes Ih channelopathy in human neurons. Science. 2016;352(6286).

47. Mei Y, Monteiro P, Zhou Y, Kim J-A, Gao X, Fu Z, et al. Adult restoration of Shank3 expression rescues selective autistic-like phenotypes. Nature. 2016;530(7591):481-4.

48. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;515(7526):209-15.

49. O/'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012;485(7397):246-50.

50. McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y, et al. De novo Mutations in Schizophrenia Implicate Chromatin Remodeling and Support a

Genetic Overlap with Autism and Intellectual Disability. Molecular psychiatry. 2014;19(6):652-8.

51. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry. 2012;17(2):142-53.

52. Ruderfer DM, Hamamsy T, Lek M, Karczewski KJ, Kavanagh D, Samocha KE, et al. Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. Nat Genet. 2016;48(10):1107-11.

53. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. The American Journal of Human Genetics. 2010;86(5):749-64.

54. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. Nature. 2016;536(7614):41-7.

55. Taylor JC, Martin HC, Lise S, Broxholme J, Cazier J-B, Rimmer A, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet. 2015;47(7):717-26.

56. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet. 2011;12(9):628-40.

57. Greene NDE, Stanier P, Copp AJ. Genetics of human neural tube defects. Human Molecular Genetics. 2009;18(R2):R113-R29.

58. Harris MJ, Juriloff DM. An update to the list of mouse mutants with neural tube closure defects and advances toward a complete genetic perspective of neural tube closure. Birth Defects Res A Clin Mol Teratol. 2010;88(8):653-69.

59. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. Nat Med. 2016;22(4):345-61.

60. McCarroll Steven A, Hyman Steven E. Progress in the Genetics of Polygenic Brain Disorders: Significant New Challenges for Neurobiology. Neuron. 2013;80(3):578-87.

61. Hindorff LA, Gillanders EM, Manolio TA. Genetic architecture of cancer and other complex diseases: lessons learned and future directions. Carcinogenesis. 2011;32(7):945-54.

62. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014;508(7497):469-76.

63. Asimit J, Zeggini E. Rare Variant Association Analysis Methods for Complex Traits. Annual Review of Genetics. 2010;44(1):293-308.

64. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008;40(6):695-701.

65. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-23.

66. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med. 2011;13(7):680-5.

67. Cady J, Allred P, Bali T, Pestronk A, Goate A, Miller TM, et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. Annals of Neurology. 2015;77(1):100-13.

68. Ramagopalan SV, Dyment DA, Cader MZ, Morrison KM, Disanto G, Morahan JM, et al. Rare variants in the CYP27B1 gene are associated with multiple sclerosis. Annals of Neurology. 2011;70(6):881-6.

69. Ament SA, Szelinger S, Glusman G, Ashworth J, Hou L, Akula N, et al. Rare variants in neuronal excitability genes influence risk for bipolar disorder. Proceedings of the National Academy of Sciences. 2015;112(11):3576-81.

70. Chen P-C, Yin J, Yu H-W, Yuan T, Fernandez M, Yung CK, et al. Next-generation sequencing identifies rare variants associated with Noonan syndrome. Proceedings of the National Academy of Sciences. 2014;111(31):11473-8.

71. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare Variants of IFIH1, a Gene Implicated in Antiviral Responses, Protect Against Type 1 Diabetes. Science. 2009;324(5925):387.

72. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014;506(7487):185-90.

73. Sifrim A, Hitz M-P, Wilsdon A, Breckpot J, Turki SHA, Thienpont B, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Nat Genet. 2016;48(9):1060-5.

74. Fearnhead NS, Wilding JL, Winney B, Tonks S, Bartlett S, Bicknell DC, et al. Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(45):15992-7.

75. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet. 2014;46(7):736-41.

76. Bonnefond A, Clement N, Fawcett K, Yengo L, Vaillant E, Guillaume J-L, et al. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet. 2012;44(3):297-301.

77. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet. 2011;43(11):1066-73.

78. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, et al. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet. 2010;42(8):684-7.

79. Kibar Z, Torban E, McDearmid JR, Reynolds A, Berghout J, Mathieu M, et al. Mutations in VANGL1 Associated with Neural-Tube Defects. New England Journal of Medicine. 2007;356(14):1432-7.

80. De Marco P, Merello E, Rossi A, Piatelli G, Cama A, Kibar Z, et al. FZD6 is a novel gene for human neural tube defects. Human Mutation. 2012;33(2):384-90.

81. Chen X, Shen Y, Gao Y, Zhao H, Sheng X, Zou J, et al. Detection of Copy Number Variants Reveals Association of Cilia Genes with Neural Tube Defects. PLoS ONE. 2013;8(1):e54492.

82. Marini NJ, Hoffmann TJ, Lammer EJ, Hardin J, Lazaruk K, Stein JB, et al. A Genetic Signature of Spina Bifida Risk from Pathway-Informed Comprehensive Gene-Variant Analysis. PLoS ONE. 2011;6(11):e28408.

83. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome Research. 2005;15(7):901-13.

84. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47-54.

85. Kosmicki JA, Churchhouse CL, Rivas MA, Neale BM. Discovery of rare variants for complex phenotypes. Hum Genet. 2016;135(6):625-34.

86. Stankiewicz P, Lupski JR. Structural Variation in the Human Genome and its Role in Disease. Annual Review of Medicine. 2010;61(1):437-55.

87. Weise A, Mrasek K, Klein E, Mulatinho M, Llerena JC, Hardekopf D, et al. Microdeletion and Microduplication Syndromes. Journal of Histochemistry and Cytochemistry. 2012;60(5):346-58.

88. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. Trends in Genetics. 2002;18(2):74-82.

89. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, et al. Recent Segmental Duplications in the Human Genome. Science. 2002;297(5583):1003.

90. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet. 2006;38(9):1038-42.

91. Carvalho CMB, Ramocki MB, Pehlivan D, Franco LM, Gonzaga-Jauregui C, Fang P, et al. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. Nat Genet. 2011;43(11):1074-81.

92. Stankiewicz P, Pursley AN, Cheung SW. Challenges in clinical interpretation of microduplications detected by array CGH analysis. American Journal of Medical Genetics Part A. 2010;152A(5):1089-100.

93. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nat Rev Genet. 2009;10(8):551-64.

94. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet. 2016;17(4):224-38.

95. Carvalho CMB, Pehlivan D, Ramocki MB, Fang P, Alleva B, Franco LM, et al. Replicative mechanisms for CNV formation are error prone. Nat Genet. 2013;45(11):1319-26.

96. Tyshchenko N, Lurie I, Schinzel A. Chromosomal map of human brain malformations. Hum Genet. 2008;124(1):73-80.

97. Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick D. A Chromosomal Deletion Map of Human Malformations. Am J Hum Genet. 1998;63(4):1153-9.

98. Rosenfeld JA, Lacassie Y, El-Khechen D, Escobar LF, Reggin J, Heuer C, et al. New cases and refinement of the critical region in the 1q41q42 microdeletion syndrome. Eur J Med Genet. 2011;54(1):42-9.

99. Jun KR, Hur YJ, Lee JN, Kim HR, Shin JH, Oh SH, et al. Clinical characterization of DISP1 haploinsufficiency: A case report. Eur J Med Genet. 2013;56(6):309-13.

100. Au PYB, Argiropoulos B, Parboosingh JS, Micheil Innes A. Refinement of the critical region of 1q41q42 microdeletion syndrome identifies FBXO28 as a candidate causative gene for intellectual disability and seizures. Am J Med Genet A. 2014;164(2):441-8.

101. Cepeda D, Ng HF, Sharifi HR, Mahmoudi S, Cerrato VS, Fredlund E, et al. CDK-mediated activation of the SCFFBXO28 ubiquitin ligase promotes MYC-driven transcription and tumourigenesis and predicts poor survival in breast cancer. EMBO Molecular Medicine. 2013;5(7):1067-86.

102. Gorina S, Pavletich NP. Structure of the p53 Tumor Suppressor Bound to the Ankyrin and SH3 Domains of 53BP2. Science. 1996;274(5289):1001-5.

103. Samuels-Lev Y, O'Connor DJ, Bergamaschi D, Trigiante G, Hsieh J-K, Zhong S, et al. ASPP Proteins Specifically Stimulate the Apoptotic Function of p53. Mol Cell. 2001;8(4):781-94.

104. Kampa KM, Acoba JD, Chen D, Gay J, Lee H, Beemer K, et al. Apoptosis-stimulating protein of p53 (ASPP2) heterozygous mice are tumor-prone and have attenuated cellular damage–response thresholds. P Natl Acad Sci USA. 2009;106(11):4390-5.

105. Vives V, Su J, Zhong S, Ratnayaka I, Slee E, Goldin R, et al. ASPP2 is a haploinsufficient tumor suppressor that cooperates with p53 to suppress tumor growth. Gene Dev. 2006;20(10):1262-7.

106. Sottocornola R, Royer C, Vives V, Tordella L, Zhong S, Wang Y, et al. ASPP2 Binds Par-3 and Controls the Polarity and Proliferation of Neural Progenitors during CNS Development. Dev Cell. 2010;19(1):126-37.

107. Bergamaschi D, Samuels Y, Jin B, Duraisingham S, Crook T, Lu X. ASPP1 and ASPP2: Common Activators of p53 Family Members. Mol Cell Biol. 2004;24(3):1341-50.

108. Wallingford JB, Niswander LA, Shaw GM, Finnell RH. The Continuing Challenge of Understanding, Preventing, and Treating Neural Tube Defects. Science. 2013;339(6123).

109. Gu X, Lin L, Zheng X, Zhang T, Song X, Wang J, et al. High prevalence of NTDs in Shanxi Province: a combined epidemiological approach. Birth defects research Part A, Clinical and molecular teratology. 2007;79(10):702-7.

110. Murdoch JN, Copp AJ. The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. Birth defects research Part A, Clinical and molecular teratology. 2010;88(8):633-52.

111. Leck I. Causation of Neural Tube Defects: Clues from Epidemiology. Brit Med Bull. 1974;30(2):158-63.

112. Harris MJ, Juriloff DM. Mouse mutants with neural tube closure defects and their role in understanding human neural tube defects. Birth Defects Research Part A: Clinical and Molecular Teratology. 2007;79(3):187-210.

113. Mitchell LE, Adzick NS, Melchionne J, Pasquariello PS, Sutton LN, Whitehead AS. Spina bifida. The Lancet.364(9448):1885-95.

114. Rampersaud E, Bassuk A, Enterline D, George T, Siegel D, Melvin E, et al. Whole genomewide linkage screen for neural tube defects reveals regions of interest on chromosomes 7 and 10. Journal of Medical Genetics. 2005;42(12):940-6.

115. Stamm DS, Rampersaud E, Slifer SH, Mehltretter L, Siegel DG, Xie J, et al. High-density single nucleotide polymorphism screen in a large multiplex neural

tube defect family refines linkage to loci at 7p21.1–pter and 2q33.1–q35. Birth Defects Research Part A: Clinical and Molecular Teratology. 2006;76(6):499-505.

116. Lei Y, Zhu H, Duhon C, Yang W, Ross ME, Shaw GM, et al. Mutations in Planar Cell Polarity Gene <italic>SCRIB</italic> Are Associated with Spina Bifida. PLoS ONE. 2013;8(7):e69262.

117. Robinson A, Escuin S, Doudney K, Vekemans M, Stevenson RE, Greene NDE, et al. Mutations in the planar cell polarity genes CELSR1 and SCRIB are associated with the severe neural tube defect craniorachischisis. Human Mutation. 2012;33(2):440-7.

118. Lei Y, Zhu H, Yang W, Ross ME, Shaw GM, Finnell RH. Identification of Novel CELSR1 Mutations in Spina Bifida. PLoS ONE. 2014;9(3):e92207.

119. Lemay P, Guyot M-C, Tremblay É, Dionne-Laporte A, Spiegelman D, Henrion É, et al. Loss-of-function de novo mutations play an important role in severe human neural tube defects. Journal of Medical Genetics. 2015.

120. Lei Y-P, Zhang T, Li H, Wu B-L, Jin L, Wang H-Y. VANGL2 Mutations in Human Cranial Neural-Tube Defects. New England Journal of Medicine. 2010;362(23):2232-5.

121. Shi Y, Ding Y, Lei Y-P, Yang X-Y, Xie G-M, Wen J, et al. Identification of novel rare mutations of DACT1 in human neural tube defects. Human Mutation. 2012;33(10):1450-5.

122. Blom HJ, Shaw GM, den Heijer M, Finnell RH. Neural tube defects and folate: case far from closed. Nat Rev Neurosci. 2006;7(9):724-31.

123. van der Linden IJM, Afman LA, Heil SG, Blom HJ. Genetic variation in genes of folate metabolism and neural-tube defect risk. Proceedings of the Nutrition Society. 2006;65(02):204-15.

124. Ishiuchi T, Takeichi M. Willin and Par3 cooperatively regulate epithelial apical constriction through aPKC-mediated ROCK phosphorylation. Nat Cell Biol. 2011;13(7):860-6.

125. David DJV, Wang Q, Feng JJ, Harris TJC. Bazooka inhibits aPKC to limit antagonism of actomyosin networks during amnioserosa apical constriction. Development. 2013;140(23):4719-29.

126. Plageman TF, Chauhan BK, Yang C, Jaudon F, Shang X, Zheng Y, et al. A Trio-RhoA-Shroom3 pathway is required for apical constriction and epithelial invagination. Development. 2011;138(23):5177.

127. Das D, Zalewski JK, Mohan S, Plageman TF, VanDemark AP, Hildebrand JD. The interaction between Shroom3 and Rho-kinase is required for neural tube morphogenesis in mice. Biology Open. 2014.

128. Gao Y, Chen X, Shangguan S, Bao Y, Lu X, Zou J, et al. Association Study of PARD3 Gene Polymorphisms With Neural Tube Defects in a Chinese Han Population. Reprod Sci. 2012;19(7):764-71.

129. Ray HJ, Niswander LA. Grainyhead-like 2 downstream targets act to suppress epithelial-to-mesenchymal transition during neural tube closure. Development. 2016;143(7):1192.

130. Wang Y, Bu F, Royer C, Serres S, Larkin JR, Soto Manuel S, et al. ASPP2 controls epithelial plasticity and inhibits metastasis through β-catenin-dependent regulation of ZEB1. Nat Cell Biol. 2014;16(11):1092-104.

131. Sah VP, Attardi LD, Mulligan GJ, Williams BO, Bronson RT, Jacks T. A subset of p53-deficient embryos exhibit exencephaly. Nat Genet. 1995;10(2):175-80.

132. Armstrong JF, Kaufman MH, Harrison DJ, Clarke AR. High-frequency developmental abnormalities in p53-deficient mice. Curr Biol. 1995;5(8):931-6.

133. Patterson AD, Hildesheim J, Fornace AJ, Jr., Hollander MC. Neural tube development requires the cooperation of p53- and Gadd45a-associated pathways. Birth defects research Part A, Clinical and molecular teratology. 2006;76(2):129-32.

134. Yang A, Walker N, Bronson R, Kaghad M, Oosterwegel M, Bonnin J, et al. p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. Nature. 2000;404(6773):99-103.

135. Talos F, Abraham A, Vaseva AV, Holembowski L, Tsirka SE, Scheel A, et al. p73 is an essential regulator of neural stem cell maintenance in embryonal and adult CNS neurogenesis. Cell Death Differ. 2010;17(12):1816-29.

136. Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: A decade of new twin studies. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2011;156(3):255-74.

137. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat Rev Genet. 2012;13(8):537-51.

138. Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder. American Journal of Medical Genetics Part C: Seminars in Medical Genetics. 2003;123C(1):48-58.

139. Owen MJ, Craddock N, O'Donovan MC. SUggestion of roles for both common and rare risk variants in genome-wide studies of schizophrenia. Archives of General Psychiatry. 2010;67(7):667-73.

140. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia. Science. 2008;320(5875):539.

141. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature. 2008;455(7210):237-41.

142. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, et al. Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nat Genet. 2011;43(9):864-8.

143. Iossifov I, O/'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515(7526):216-21.

144. Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. Nat Rev Genet. 2014;15(2):133-41.

145. Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, et al. Contribution of SHANK3 Mutations to Autism Spectrum Disorder. The American Journal of Human Genetics. 2007;81(6):1289-97.

146. Gauthier J, Spiegelman D, Piton A, Lafrenière RG, Laurent S, St-Onge J, et al. Novel de novo SHANK3 mutation in autistic patients. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2009;150B(3):421-4.

147. Gauthier J, Champagne N, Lafrenière RG, Xiong L, Spiegelman D, Brustein E, et al. De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. Proceedings of the National Academy of Sciences. 2010;107(17):7863-8.

148. Hara M, Ohba C, Yamashita Y, Saitsu H, Matsumoto N, Matsuishi T. De novo SHANK3 mutation causes Rett syndrome-like phenotype in a female patient. American Journal of Medical Genetics Part A. 2015;167(7):1593-6.

149. Durand CM, Perroy J, Loll F, Perrais D, Fagni L, Bourgeron T, et al. SHANK3 mutations identified in autism lead to modification of dendritic spine morphology via an actin-dependent mechanism. Mol Psychiatry. 2011.

150. Bozdagi O, Sakurai T, Papapetrou D, Wang X, Dickstein DL, Takahashi N, et al. Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. Molecular Autism. 2010;1(1):1-15.

151. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. Science. 2013;339(6127):1546-58.

152. Knudson AG. Mutation and Cancer: Statistical Study of Retinoblastoma. Proceedings of the National Academy of Sciences. 1971;68(4):820-3.

153. Comings DE. A General Theory of Carcinogenesis. Proceedings of the National Academy of Sciences. 1973;70(12):3324-8.

154. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell.61(5):759-67.

155. Berger AH, Knudson AG, Pandolfi PP. A continuum model for tumour suppression. Nature. 2011;476(7359):163-9.

156. Payne SR, Kemp CJ. Tumor suppressor genetics. Carcinogenesis. 2005;26(12):2031-45.

157. Fero ML, Randel E, Gurley KE, Roberts JM, Kemp CJ. The murine gene p27Kip1 is haplo-insufficient for tumour suppression. Nature. 1998;396(6707):177-80.

158. Philipp-Staheli J, Kim K-H, Payne SR, Gurley KE, Liggitt D, Longton G, et al. Pathway-specific tumor suppression: Reduction of p27 accelerates gastrointestinal tumorigenesis in Apc mutant mice, but not in Smad3 mutant mice. Cancer Cell. 2002;1(4):355-68.

159. Tordella L, Koch S, Salter V, Pagotto A, Doondeea JB, Feller SM, et al. ASPP2 suppresses squamous cell carcinoma via RelA/p65–mediated repression of p63. Proceedings of the National Academy of Sciences. 2013.

160. Solimini NL, Xu Q, Mermel CH, Liang AC, Schlabach MR, Luo J, et al. Recurrent Hemizygous Deletions in Cancers May Optimize Proliferative Potential. Science. 2012;337(6090):104-9.

161. Lee EYHP, Abbondante S. Tissue-specific tumor suppression by BRCA1. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(12):4353-4.

162. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014;505(7484):495-501.

163. Pelletier J, Bruening W, Li FP, Haber DA, Glaser T, Housman DE. WT1 mutations contribute to abnormal genital system development and hereditary Wilms' tumour. Nature. 1991;353(6343):431-4.

164. King-Underwood L, Pritchard-Jones K. Gene Mutations Occur Mainly in Acute Myeloid Leukemia and May Confer Drug Resistance. Blood. 1998;91(8):2961.

165. Dietrich WF, Lander ES, Smith JS, Moser AR, Gould KA, Luongo C, et al. Genetic identification of Mom-1, a major modifier locus affecting Min-induced intestinal neoplasia in the mouse. Cell. 1993;75(4):631-9.

166. Barnes DR, Antoniou AC. Unravelling modifiers of breast and ovarian cancer risk for BRCA1 and BRCA2 mutation carriers: update on genetic modifiers. Journal of Internal Medicine. 2012;271(4):331-43.

167. Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge JV, Thomas JL, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat Genet. 2015;47(2):106-14.

168. Klijn C, Bot J, Adams DJ, Reinders M, Wessels L, Jonkers J. Identification of Networks of Co-Occurring, Tumor-Related DNA Copy Number Changes Using a Genome-Wide Scoring Approach. PLoS Comput Biol. 2010;6(1):e1000631.

169. Emery CM, Vijayendran KG, Zipser MC, Sawyer AM, Niu L, Kim JJ, et al. MEK1 mutations confer resistance to MEK and B-RAF inhibition. Proceedings of the National Academy of Sciences. 2009;106(48):20411-6.

170. Villanueva J, Infante Jeffrey R, Krepler C, Reyes-Uribe P, Samanta M, Chen H-Y, et al. Concurrent MEK2 Mutation and BRAF Amplification Confer Resistance to BRAF and MEK Inhibitors in Melanoma. Cell Reports. 2013;4(6):1090-9.

171. Merlo LMF, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. Nat Rev Cancer. 2006;6(12):924-35.

172. Caignard A, Martin MS, Michel MF, Martin F. Interaction between two cellular subpopulations of a rat colonic carcinoma when inoculated to the syngeneic host. International Journal of Cancer. 1985;36(2):273-9.

173. Miller BE, Miller FR, Leith J, Heppner GH. Growth Interactionin Vivo between Tumor Subpopulations Derived from a Single Mouse Mammary Tumor. Cancer Research. 1980;40(11):3977-81.

174. McFarland CD, Korolev KS, Kryukov GV, Sunyaev SR, Mirny LA. Impact of deleterious passenger mutations on cancer progression. Proceedings of the National Academy of Sciences. 2013;110(8):2910-5.

175. Castro-Giner F, Ratcliffe P, Tomlinson I. The mini-driver model of polygenic cancer evolution. Nat Rev Cancer. 2015;15(11):680-5.

176. Davoli T, Xu Andrew W, Mengwasser Kristen E, Sack Laura M, Yoon John C, Park Peter J, et al. Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome. Cell.155(4):948-62.

177. Copeland NG, Jenkins NA. Harnessing transposons for cancer gene discovery. Nat Rev Cancer. 2010;10(10):696-706.

178. DeNicola GM, Karreth FA, Adams DJ, Wong CC. The utility of transposon mutagenesis for cancer studies in the era of genome editing. Genome Biology. 2015;16(1):1-15.

179. Rad R, Rad L, Wang W, Strong A, Ponstingl H, Bronner IF, et al. A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. Nat Genet. 2015;47(1):47-56.

180. Imamura Y, Lochhead P, Yamauchi M, Kuchiba A, Qian ZR, Liao X, et al. Analyses of clinicopathological, molecular, and prognostic associations of KRAS codon 61 and codon 146 mutations in colorectal cancer: cohort study and literature review. Molecular Cancer. 2014;13(1):1-15.

181. Segditsas S, Rowan AJ, Howarth K, Jones A, Leedham S, Wright NA, et al. APC and the three-hit hypothesis. Oncogene. 2008;28(1):146-55.

182. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330-7.

183. Tuupanen S, Niittymäki I, Nousiainen K, Vanharanta S, Mecklin J-P, Nuorva K, et al. Allelic Imbalance at rs6983267 Suggests Selection of the Risk Allele in Somatic Colorectal Tumor Evolution. Cancer Research. 2008;68(1):14-7.

184. Shen C, Beroukhim R, Schumacher SE, Zhou J, Chang M, Signoretti S, et al. Genetic and Functional Studies ImplicateHIF1α as a 14q Kidney Cancer Suppressor Gene. Cancer Discovery. 2011;1(3):222-35.

185. Favero F, McGranahan N, Salm M, Birkbak NJ, Sanborn JZ, Benz SC, et al. Glioblastoma adaptation traced through decline of an IDH1 clonal driver and macro-evolution of a double-minute chromosome. Annals of Oncology. 2015;26(5):880-7.

186. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499(7457):214-8.

187. Kim E, Ilic N, Shrestha Y, Zou L, Kamburov A, Zhu C, et al. Systematic functional interrogation of rare cancer variants identifies oncogenic alleles. Cancer Discovery. 2016.

188. Berger Alice H, Brooks Angela N, Wu X, Shrestha Y, Chouinard C, Piccioni F, et al. High-throughput Phenotyping of Lung Cancer Somatic Mutations. Cancer Cell. 2016;30(2):214-28.

189. Kim JW, Botvinnik OB, Abudayyeh O, Birger C, Rosenbluh J, Shrestha Y, et al. Characterizing genomic alterations in cancer by complementary functional associations. Nat Biotech. 2016;34(5):539-46.

190. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat Genet. 2016;48(8):838-47.

191. Prasad V. Perspective: The precision-oncology illusion. Nature. 2016;537(7619):S63-S.

192. Wong WC, Kim D, Carter H, Diekhans M, Ryan MC, Karchin R. CHASM and SNVBox: toolkit for detecting biologically important single nucleotide mutations in cancer. Bioinformatics. 2011;27(15):2147-8.

193. Shihab HA, Gough J, Cooper DN, Day INM, Gaunt TR. Predicting the functional consequences of cancer-associated amino acid substitutions. Bioinformatics. 2013;29(12):1504-10.

194. Kumar RD, Swamidass SJ, Bose R. Unsupervised detection of cancer driver mutations with parsimony-guided learning. Nat Genet. 2016;advance online publication.

195. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nat Rev Cancer. 2004;4(3):177-83.

196. Garraway Levi A, Lander Eric S. Lessons from the Cancer Genome. Cell. 2013;153(1):17-37.

197. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). Current Protocols in Human Genetics: John Wiley & Sons, Inc.; 2001.

198. Chung I-F, Chen C-Y, Su S-C, Li C-Y, Wu K-J, Wang H-W, et al. DriverDBv2: a database for human cancer driver gene research. Nucleic Acids Research. 2016;44(D1):D975-D9.

199. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandoth C, Reimand J, et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. Scientific Reports. 2013;3:2650.

200. An O, Dall'Olio GM, Mourikis TP, Ciccarelli FD. NCG 5.0: updates of a manually curated repository of cancer genes and associated properties from cancer mutational screenings. Nucleic Acids Research. 2016;44(D1):D992-D9.

201. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a Shared Vision for Cancer Genomic Data. New England Journal of Medicine. 2016;375(12):1109-12.

202. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A Direct Estimate of the Human αβ T Cell Receptor Diversity. Science. 1999;286(5441):958-61.

203. Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, et al. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proceedings of the National Academy of Sciences. 2009;106(48):20216-21.

204. Perelson AS, Oster GF. Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-non-self discrimination. Journal of Theoretical Biology. 1979;81(4):645-70.

205. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. New England Journal of Medicine. 2010;363(8):711-23.

206. Wolchok JD, Kluger H, Callahan MK, Postow MA, Rizvi NA, Lesokhin AM, et al. Nivolumab plus Ipilimumab in Advanced Melanoma. New England Journal of Medicine. 2013;369(2):122-33.

207. Robert C, Schachter J, Long GV, Arance A, Grob JJ, Mortier L, et al. Pembrolizumab versus Ipilimumab in Advanced Melanoma. New England Journal of Medicine. 2015;372(26):2521-32.

208. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD, et al. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. New England Journal of Medicine. 2015;373(1):23-34.

209. Barber DL, Wherry EJ, Masopust D, Zhu B, Allison JP, Sharpe AH, et al. Restoring function in exhausted CD8 T cells during chronic viral infection. Nature. 2006;439(7077):682-7.

210. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature. 2014;515(7528):568-71.

211. Nguyen LT, Ohashi PS. Clinical blockade of PD1 and LAG3 [mdash] potential mechanisms of action. Nat Rev Immunol. 2015;15(1):45-56.

212. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. Science. 2015;348(6230):124.

213. McGranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science. 2016;351(6280):1463.

214. Im SJ, Hashimoto M, Gerner MY, Lee J, Kissick HT, Burger MC, et al. Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. Nature. 2016;537(7620):417-21.

215. Sala C, Vicidomini C, Bigi I, Mossa A, Verpelli C. Shank synaptic scaffold proteins: keys to understanding the pathogenesis of autism and other synaptic disorders. Journal of Neurochemistry. 2015;135(5):849-58.

216. Iwabuchi K, Bartel PL, Li B, Marraccino R, Fields S. Two cellular proteins that bind to wild-type but not mutant p53. Proceedings of the National Academy of Sciences. 1994;91(13):6098-102.

217. Marston NJ CT, Vousden KH. Interaction of p53 with MDM2 is independent of E6 and does not mediate wild type transformation suppressor function. Oncogene. 1994;9(9):2707-16.

218. Scheffner M, Takahashi T, Huibregtse JM, Minna JD, Howley PM. Interaction of the human papillomavirus type 16 E6 oncoprotein with wild-type and mutant human p53 proteins. Journal of Virology. 1992;66(8):5100-5.

219. Iwabuchi K, Li B, Massa HF, Trask BJ, Date T, Fields S. Stimulation of p53-mediated Transcriptional Activation by the p53-binding Proteins, 53BP1 and 53BP2. Journal of Biological Chemistry. 1998;273(40):26061-8.

220. Naumovski L, Cleary ML. The p53-binding protein 53BP2 also interacts with Bc12 and impedes cell cycle progression at G2/M. Molecular and Cellular Biology. 1996;16(7):3884-92.

221. Nagase T, Ishikawa K-i, Suyama M, Kikuno R, Hirosawa M, Miyajima N, et al. Prediction of the Coding Sequences of Unidentified Human Genes. XII. The Complete Sequences of 100 New cDNA Clones from Brain Which Code for Large Proteins in vitro. DNA Research. 1998;5(6):355-64.

222. Yang J-P, Hori M, Sanda T, Okamoto T. Identification of a Novel Inhibitor of Nuclear Factor-κB, RelA-associated Inhibitor. Journal of Biological Chemistry. 1999;274(22):15662-70.

223. Morgan AA, Rubenstein E. Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome. PLoS ONE. 2013;8(1):e53785.

224. Robinson RA, Lu X, Jones EY, Siebold C. Biochemical and Structural Studies of ASPP Proteins Reveal Differential Binding to p53, p63, and p73. Structure. 2008;16(2):259-68.

225. Bergamaschi D, Samuels Y, O'Neil NJ, Trigiante G, Crook T, Hsieh J-K, et al. iASPP oncoprotein is a key inhibitor of p53 conserved from worm to human. Nat Genet. 2003;33(2):162-7.

226. Trigiante G, Lu X. ASPPs and cancer. Nat Rev Cancer. 2006;6(3):217-26.

227. Lettre G, Kritikou EA, Jaeggi M, Calixto A, Fraser AG, Kamath RS, et al. Genome-wide RNAi identifies p53-dependent and -independent regulators of germ cell apoptosis in C. elegans. Cell Death Differ. 2004;11(11):1198-203.

228. Song X, Du J, Zhu W, Jin P, Ma F. Identification and characterization of an apoptosis-stimulating protein of p53 (ASPP) gene from Branchiostoma belcheri: Insights into evolution of ASPP gene family. Fish & Shellfish Immunology. 2016;49:268-74.

229. Slee EA, O'Connor DJ, Lu X. To die or not to die: how does p53 decide? Oncogene. 2004;23(16):2809-18.

230. Helps NR BH, Elledge SJ, Cohen PT. Protein phosphatase 1 interacts with p53BP2, a protein which binds to the tumour suppressor p53. FEBS Letters. 1995;377(3):295-300.

231. Liu C-Y, Lv X, Li T, Xu Y, Zhou X, Zhao S, et al. PP1 Cooperates with ASPP2 to Dephosphorylate and Activate TAZ. Journal of Biological Chemistry. 2011;286(7):5558-66.

232. Royer C, Koch S, Qin X, Zak J, Buti L, Dudziec E, et al. ASPP2 Links the Apical Lateral Polarity Complex to the Regulation of YAP Activity in Epithelial Cells. PLoS ONE. 2014;9(10):e111384.

233. Espanel X, Sudol M. Yes-associated Protein and p53-binding Protein-2 Interact through Their WW and SH3 Domains. Journal of Biological Chemistry. 2001;276(17):14514-23.

234. Couzens AL, Knight JDR, Kean MJ, Teo G, Weiss A, Dunham WH, et al. Protein Interaction Network of the Mammalian Hippo Pathway Reveals Mechanisms of Kinase-Phosphatase Interactions. Science Signaling. 2013;6(302):rs15.

235. Hauri S, Wepf A, van Drogen A, Varjosalo M, Tapon N, Aebersold R, et al. Interaction proteome of human Hippo signaling: modular control of the co-activator YAP1. Molecular Systems Biology. 2013;9(1).

236. Gao K, An J, Zhang Y, Jin X, Ma J, Peng J, et al. The E3 ubiquitin ligase Itch and Yap1 have antagonistic roles in the regulation of ASPP2 protein stability. FEBS Letters. 2015;589(1):94-101.

237. Zhao B, Wei X, Li W, Udan RS, Yang Q, Kim J, et al. Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. Genes & Development. 2007;21(21):2747-61.

238. Vigneron AM, Ludwig RL, Vousden KH. Cytoplasmic ASPP1 inhibits apoptosis through the control of YAP. Genes & Development. 2010;24(21):2430-9.

239. Fardilha M, Esteves SLC, Korrodi-Gregório L, Vintém AP, Domingues SC, Rebelo S, et al. Identification of the human testis protein phosphatase 1 interactome. Biochemical Pharmacology. 2011;82(10):1403-15.

240. Esteves SLC, Korrodi-Gregório L, Cotrim CZ, van Kleeff PJM, Domingues SC, da Cruz e Silva OAB, et al. Protein Phosphatase 1γ Isoforms Linked Interactions in the Brain. Journal of Molecular Neuroscience. 2013;50(1):179-97.

241. Zhang P, Zhang Y, Gao K, Wang Y, Jin X, Wei Y, et al. ASPP1/2-PP1 complexes are required for chromosome segregation and kinetochore-microtubule attachments. Oncotarget. 2015;6(39):41550-65.

242. Cong W, Hirose T, Harita Y, Yamashita A, Mizuno K, Hirano H, et al. ASPP2 Regulates Epithelial Cell Polarity through the PAR Complex. Current Biology. 2010;20(15):1408-14.

243. Buti L, Spooner E, Van der Veen AG, Rappuoli R, Covacci A, Ploegh HL. Helicobacter pylori cytotoxin-associated gene A (CagA) subverts the apoptosis-stimulating protein of p53 (ASPP2) tumor suppressor pathway of the host. Proceedings of the National Academy of Sciences. 2011;108(22):9238-43.

244. Nešić D, Buti L, Lu X, Stebbins CE. Structure of the Helicobacter pylori CagA oncoprotein bound to the human tumor suppressor ASPP2. Proceedings of the National Academy of Sciences. 2014;111(4):1562-7.

245. Cao Y, Hamada T, Matsui T, Date T, Iwabuchi K. Hepatitis C virus core protein interacts with p53-binding protein, 53BP2/Bbp/ASPP2, and inhibits p53-mediated apoptosis. Biochemical and Biophysical Research Communications. 2004;315(4):788-95.

246. Wang Y, Wang XD, Lapi E, Sullivan A, Jia W, He Y-W, et al. Autophagic activity dictates the cellular response to oncogenic RAS. Proceedings of the National Academy of Sciences. 2012.

247. Wang Y, Godin-Heymann N, Dan Wang X, Bergamaschi D, Llanos S, Lu X. ASPP1 and ASPP2 bind active RAS, potentiate RAS signalling and enhance p53 activity in cancer cells. Cell Death Differ. 2013;20(4):525-34.

248. Wang Z, Liu Y, Takahashi M, Van Hook K, Kampa-Schittenhelm KM, Sheppard BC, et al. N terminus of ASPP2 binds to Ras and enhances

Ras/Raf/MEK/ERK activation to promote oncogene-induced senescence. Proceedings of the National Academy of Sciences. 2013;110(1):312-7.

249. Posada IMD, Serulla M, Zhou Y, Oetken-Lindholm C, Abankwa D, Lectez B. ASPP2 Is a Novel Pan-Ras Nanocluster Scaffold. PLoS ONE. 2016;11(7):e0159677.

250. Liu K, Shi Y, Guo X, Wang S, Ouyang Y, Hao M, et al. CHOP mediates ASPP2-induced autophagic apoptosis in hepatoma cells by releasing Beclin-1 from Bcl-2 and inducing nuclear translocation of Bcl-2. Cell Death Dis. 2014;5:e1323.

251. Shi Y, Han Y, Xie F, Wang A, Feng X, Li N, et al. ASPP2 enhances Oxaliplatin (L-OHP)-induced colorectal cancer cell apoptosis in a p53-independent manner by inhibiting cell autophagy. Journal of Cellular and Molecular Medicine. 2015;19(3):535-43.

252. Lu M, Zak J, Chen S, Sanchez-Pulido L, Severson David T, Endicott J, et al. A Code for RanGDP Binding in Ankyrin Repeats Defines a Nuclear Import Pathway. Cell. 2014;157(5):1130-45.

253. Rotem S, Katz C, Benyamini H, Lebendiker M, Veprintsev D, Rüdiger S, et al. The Structure and Interactions of the Proline-rich Domain of ASPP2. Journal of Biological Chemistry. 2008;283(27):18990-9.

254. Rotem-Bamberger S, Katz C, Friedler A. Regulation of ASPP2 Interaction with p53 Core Domain by an Intramolecular Autoinhibitory Mechanism. PLoS ONE. 2013;8(3):e58470.

255. Godin-Heymann N, Wang Y, Slee E, Lu X. Phosphorylation of ASPP2 by RAS/MAPK Pathway Is Critical for Its Full Pro-Apoptotic Function. PLoS ONE. 2013;8(12):e82022.

256. Reingewertz TH, Iosub-Amir A, Bonsor DA, Mayer G, Amartely H, Friedler A, et al. An Intrinsically Disordered Region in the Proapoptotic ASPP2 Protein Binds to the Helicobacter pylori Oncoprotein CagA. Biochemistry. 2015;54(21):3337-47.

257. Lu M, Breyssens H, Salter V, Zhong S, Hu Y, Baer C, et al. Restoring p53 Function in Human Melanoma Cells by Inhibiting MDM2 and Cyclin B1/CDK1-Phosphorylated Nuclear iASPP. Cancer cell. 2013;23(5):618-33.

258. Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, et al. High-throughput discovery of novel developmental phenotypes. Nature. 2016;537(7621):508-14.

259. Marin O, Rubenstein JLR. A long, remarkable journey: Tangential migration in the telencephalon. Nat Rev Neurosci. 2001;2(11):780-90.

260. Calcagnotto ME, Ruiz LP, Blanco MM, Santos-Junior JG, Valente MF, Patti C, et al. Effect of neuronal precursor cells derived from medial ganglionic eminence in an acute epileptic seizure model. Epilepsia. 2010;51:71-5.

261. Hirashima M, Sano K, Morisada T, Murakami K, Rossant J, Suda T. Lymphatic vessel assembly is impaired in Aspp1-deficient mouse embryos. Developmental Biology. 2008;316(1):149-59.

262. Herron BJ, Rao C, Liu S, Laprade L, Richardson JA, Olivieri E, et al. A mutation in NFkB interacting protein 1 results in cardiomyopathy and abnormal skin development in wa3 mice. Human Molecular Genetics. 2005;14(5):667-77.

263. Toonen J, Liang L, Sidjanin DJ. Waved with open eyelids 2 (woe2) is a novel spontaneous mouse mutation in the protein phosphatase 1, regulatory (inhibitor) subunit 13 like (Ppp1r13l)gene. BMC Genetics. 2012;13(1):1-10.

264. Simpson MA, Cook RW, Solanki P, Patton MA, Dennis JA, Crosby AH. A mutation in NFκB interacting protein 1 causes cardiomyopathy and woolly haircoat syndrome of Poll Hereford cattle. Animal Genetics. 2009;40(1):42-6.

265. Notari M, Hu Y, Sutendra G, Dedeić Z, Lu M, Dupays L, et al. iASPP, a previously unidentified regulator of desmosomes, prevents arrhythmogenic right ventricular cardiomyopathy (ARVC)-induced sudden death. Proceedings of the National Academy of Sciences. 2015;112(9):E973-E81.

266. Bailey P, Chang DK, Nones K, Johns AL, Patch A-M, Gingras M-C, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016;531(7592):47-52.

267. Park SW, An CH, Kim SS, Yoo NJ, Lee SH. Mutational Analysis of ASPP1 and ASPP2 Genes, a p53-related Gene, in Gastric and Cololorectal Cancers with Microsatellite Instability. Gut Liver. 2010;4(2):292-3.

268. Yi Ao LHRaLN. p53-interacting protein 53BP2 inhibits clonogenic survival and sensitizes cells to doxorubicin but not paclitaxel-induced apoptosis. Oncogene. 2001;20(21):2720-5.

269. Mak VCY, Lee L, Siu MKY, Wong OGW, Lu X, Ngan HYS, et al. Downregulation of ASPP1 in gestational trophoblastic disease: correlation with hypermethylation, apoptotic activity and clinical outcome. Mod Pathol. 2011;24(4):522-32.

270. Schittenhelm MM, Illing B, Ahmut F, Rasp KH, Blumenstock G, Döhner K, et al. Attenuated Expression of <underline>A</underline>poptosis <underline>S</underline>timulating <underline>P</underline>rotein of <underline>p</underline>53-2 (ASPP2) in Human Acute Leukemia Is Associated with Therapy Failure. PLoS ONE. 2013;8(11):e80193.

271. Liu K, Jiang T, Ouyang Y, Shi Y, Zang Y, Li N, et al. Nuclear EGFR impairs ASPP2-p53 complex-induced apoptosis by inducing SOS1 expression in hepatocellular carcinoma2015.

272. Song Q, Song J, Wang Q, Ma Y, Sun N, Ma J, et al. miR-548d-3p/TP53BP2 axis regulates the proliferation and apoptosis of breast cancer cells. Cancer Medicine. 2016;5(2):315-24.

273. Xu L, Tong X, Zhang S, Yin F, Li X, Wei H, et al. ASPP2 suppresses stem cell-like characteristics and chemoresistance by inhibiting the Src/FAK/Snail axis in hepatocellular carcinoma. Tumor Biology. 2016:1-9.

274. Lossos IS, Natkunam Y, Levy R, Lopez CD. Apoptosis Stimulating Protein of p53 (ASPP2) Expression Differs in Diffuse Large B-cell and Follicular Center Lymphoma: Correlation with Clinical Outcome. Leukemia & Lymphoma. 2002;43(12):2309-17.

275. Liu ZJ, Zhang Y, Zhang XB, Yang X. Abnormal mRNA expression of ASPP members in leukemia cell lines. Leukemia. 2004;18(4):880-.

276. Liu Z-J, Lu X, Zhang Y, Zhong S, Gu S-Z, Zhang X-B, et al. Downregulated mRNA expression of ASPP and the hypermethylation of the 5′-untranslated region in cancer cell lines retaining wild-type p53. FEBS Letters. 2005;579(7):1587-90.

277. Zhao J, Wu G, Bu F, Lu B, Liang A, Cao L, et al. Epigenetic silence of ankyrin-repeat–containing, SH3-domain–containing, and proline-rich-region–containing protein 1 (ASPP1) and ASPP2 genes promotes tumor growth in hepatitis B virus–positive hepatocellular carcinoma. Hepatology. 2010;51(1):142-53.

278. Liu W, Jiang X, Ren J, Zhang Z. Expression Pattern of the ASPP Family Members in Endometrial Endometrioid Adenocarcinoma. Onkologie. 2010;33(10):500-3.

279. Kampa-Schittenhelm KM, Lopez CD, Schittenhelm M. ASPP2 Expression In Acute Leukemia Modulates Response to Therapy. Blood. 2010;116(21):3143.

280. Kampa-Schittenhelm KM, Akmut F, Illing B, Lopez C, Schittenhelm MM. ASPP2 Attenuation Is an Early Step in Leukemogenesis Facilitating Acquisition of Mutations. Blood. 2014;124(21):2221.

281. Mori T, Okamoto H, Takahashi N, Ueda R, Okamoto T. Aberrant overexpression of 53BP2 mRNA in lung cancer cell lines. FEBS Letters. 2000;465(2–3):124-8.

282. Levan K, Partheen K, ÖSterberg L, Olsson B, Delle U, Eklind S, et al. Identification of a Gene Expression Signature for Survival Prediction in Type I Endometrial Carcinoma. Gene Expression. 2010;14(6):361-70.

283. Agirre X, Roman-Gomez J, Jimenez-Velasco A, Garate L, Montiel-Duarte C, Navarro G, et al. ASPP1, a common activator of TP53, is inactivated by aberrant methylation of its promoter in acute lymphoblastic leukemia. Oncogene. 2005;25(13):1862-70.

284. Yamashita M, Nitta E, Suda T. Aspp1 Preserves Hematopoietic Stem Cell Pool Integrity and Prevents Malignant Transformation. Cell Stem Cell. 2015;17(1):23-34.

285. Vigneron AM, Vousden KH. An indirect role for ASPP1 in limiting p53-dependent p21 expression and cellular senescence. The EMBO Journal. 2012;31(2):471-80.

286. Aylon Y, Ofir-Rosenfeld Y, Yabuta N, Lapi E, Nojima H, Lu X, et al. The Lats2 tumor suppressor augments p53-mediated apoptosis by promoting the nuclear proapoptotic function of ASPP1. Genes & Development. 2010;24(21):2420-9.

287. Nexø BA, Vogel U, Olsen A, Nyegaard M, Bukowy Z, Rockenbauer E, et al. Linkage disequilibrium mapping of a breast cancer susceptibility locus near RAI/PPP1R13L/iASPP. BMC Medical Genetics. 2008;9(1):1-12.

288. Deng Q, Sheng L, Su D, Zhang L, Liu P, Lu K, et al. Genetic polymorphisms in ATM, ERCC1, APE1 and iASPP genes and lung cancer risk in a population of southeast China. Medical Oncology. 2011;28(3):667-72.

289. Chae YS, Kim JG, Kang BW, Lee SJ, Jeon H-S, Park JS, et al. PPP1R13L variant associated with prognosis for patients with rectal cancer. Journal of Cancer Research and Clinical Oncology. 2013;139(3):465-73.

290. Yin J, Wang H, Vogel U, Wang C, Hou W, Ma Y. Association and interaction of NFKB1 rs28362491 insertion/deletion ATTG polymorphism and PPP1R13L and CD3EAP related to lung cancer risk in a Chinese population. Tumor Biology. 2016;37(4):5467-73.

291. Skjelbred CF, Sæbø M, Nexø BA, Wallin H, Hansteen I-L, Vogel U, et al. Effects of polymorphisms in ERCC1, ASE-1 and RAIon the risk of colorectal carcinomas and adenomas: a case control study. BMC Cancer. 2006;6(1):1-6.

292. Hansen RD, Sørensen M, Tjønneland A, Overvad K, Wallin H, Raaschou-Nielsen O, et al. A haplotype of polymorphisms in ASE-1, RAI and ERCC1and the effects of tobacco smoking and alcohol consumption on risk of colorectal cancer: a danish prospective case-cohort study. BMC Cancer. 2008;8(1):1-8.

293. Sæbø M, Skjelbred CF, Nexø BA, Wallin H, Hansteen I-L, Vogel U, et al. Increased mRNA expression levels of ERCC1, OGG1 and RAIin colorectal adenomas and carcinomas. BMC Cancer. 2006;6(1):1-7.

294. Zhang X, Wang M, Zhou C, Chen S, Wang J. The expression of iASPP in acute leukemias. Leukemia Research. 2005;29(2):179-83.

295. CHEN Yanping SL, HUANG Bencheng,WANG Changsong. Expression and significance of iASSP in invasive lobular breast carcinoma. Chinese Journal of General Surgery. 2010;11.

296. Liu Z, Zhang X, Huang D, Liu Y, Zhang X, Liu L, et al. Elevated expression of iASPP in head and neck squamous cell carcinoma and its clinical significance. Medical Oncology. 2012;29(5):3381-8.

297. Cao L, Huang Q, He J, Lu J, Xiong Y. Elevated expression of iASPP correlates with poor prognosis and chemoresistance/radioresistance in FIGO Ib1-IIa squamous cell cervical cancer. Cell and Tissue Research. 2013;352(2):361-9.

298. Kong F, Shi X, Li H, Li P, Yu J, Li X, et al. Increased expression of iASPP correlates with poor prognosis in FIGO IA2–IIA cervical adenocarcinoma following a curative resection. American Journal of Cancer Research. 2015;5(3):1217-24.

299. Laska MJ, Lowe SW, Zender L, Hearn S, Vogel U, Jensen UB, et al. Enforced expression of PPP1R13L increases tumorigenesis and invasion through p53-dependent and p53-independent mechanisms. Molecular Carcinogenesis. 2009;48(9):832-42.

300. Liu H, Wang M, Diao S, Rao Q, Zhang X, Xing H, et al. siRNA-mediated down-regulation of iASPP promotes apoptosis induced by etoposide and daunorubicin in leukemia cells expressing wild-type p53. Leukemia Research. 2009;33(9):1243-8.

301. Chen J, Xie F, Zhang L, Jiang W. iASPP is over-expressed in human non-small cell lung cancer and regulates the proliferation of lung cancer cells through a p53 associated pathway. BMC Cancer. 2010;10(1):694.

302. Li G, Wang R, Gao J, Deng K, Wei J, Wei Y. RNA interference-mediated silencing of iASPP induces cell proliferation inhibition and G0/G1 cell cycle arrest in U251 human glioblastoma cells. Molecular and Cellular Biochemistry. 2011;350(1):193-200.

303. Zhang B, Xiao HJ, Chen J, Tao X, Cai LH. Inhibitory member of the apoptosis-stimulating protein of p53 (ASPP) family promotes growth and tumorigenesis in human p53-deficient prostate cancer cells. Prostate Cancer Prostatic Dis. 2011;14(3):219-24.

304. Chen Y, Yan W, He S, Chen J, Chen D, Zhang Z, et al. In vitro effect of iASPP on cell growth of oral tongue squamous cell carcinoma. Chinese Journal of Cancer Research. 2014;26(4):382-90.

305. Notari M, Hu Y, Koch S, Lu M, Ratnayaka I, Zhong S, et al. Inhibitor of apoptosis-stimulating protein of p53 (iASPP) prevents senescence and is required for epithelial stratification. Proceedings of the National Academy of Sciences. 2011;108(40):16645-50.

306. Liu X, Kang J, Liu F, Wen S, Zeng X, Liu K, et al. Overexpression of iASPP-SV in glioma is associated with poor prognosis by promoting cell viability and antagonizing apoptosis. Tumor Biology. 2016;37(5):6323-30.

307. Liu Y, Fei T, Zheng X, Brown M, Zhang P, Liu XS, et al. An Integrative Pharmacogenomic Approach Identifies Two-drug Combination Therapies for Personalized Cancer Medicine. Scientific Reports. 2016;6:22120.

308. Mazzeu JF, Krepischi-Santos AC, Rosenberg C, Szuhai K, Knijnenburg J, Weiss JMM, et al. Chromosome abnormalities in two patients with features of autosomal dominant Robinow syndrome. Am J Med Genet A. 2007;143A(15):1790-5.

309. Mazzeu JF, Vianna-Morgante AM, Krepischi ACV, Oudakker A, Rosenberg C, Szuhai K, et al. Deletions encompassing 1q41q42.1 and clinical features of autosomal dominant Robinow syndrome. Clin Genet. 2010;77(4):404-7.

310. Rice GM, Qi Z, Selzer R, Richmond T, Thompson K, Pauli RM, et al. Microdissection-based high-resolution genomic array analysis of two patients with cytogenetically identical interstitial deletions of chromosome 1q but distinct clinical phenotypes. Am J Med Genet A. 2006;140A(15):1637-43.

311. Filges I, Röthlisberger B, Boesch N, Weber P, Wenzel F, Huber AR, et al. Interstitial deletion 1q42 in a patient with agenesis of corpus callosum: Phenotype–genotype comparison to the 1q41q42 microdeletion suggests a contiguous 1q4 syndrome. Am J Med Genet A. 2010;152A(4):987-93.

312. Spreiz A, Haberlandt E, Baumann M, Baumgartner Sigl S, Fauth C, Gautsch K, et al. Chromosomal microaberrations in patients with epilepsy, intellectual disability, and congenital anomalies. Clin Genet 2014;86(4):361-6.

313. Christensen RD, Yaish HM. A neonate with the Pelger-Huet anomaly, cleft lip and palate, and agenesis of the corpus callosum, with a chromosomal microdeletion involving 1q41 to 1q42.12. J Perinatol. 2012;32(3):238-40.

314. Wat MJ, Veenma D, Hogue J, Holder AM, Yu Z, Wat JJ, et al. Genomic alterations that contribute to the development of isolated and non-isolated congenital diaphragmatic hernia. J Med Genet. 2011;48(5):299-307.

315. Ono Y, Sorimachi H. Calpains — An elaborate proteolytic system. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2012;1824(1):224-36.

316. Takano J, Mihira N, Fujioka R, Hosoki E, Chishti AH, Saido TC. Vital Role of the Calpain-Calpastatin System for Placental-Integrity-Dependent Embryonic Survival. Mol Cell Biol. 2011;31(19):4097-106.

317. Dutt P, Croall D, Arthur JS, Veyra TD, Williams K, Elce J, et al. m-Calpain is required for preimplantation embryonic development in mice. BMC Dev Biol. 2006;6(1):3.

318. Hata S, Abe M, Suzuki H, Kitamura F, Toyama-Sorimachi N, Abe K, et al. Calpain 8/nCL-2 and Calpain 9/nCL-4 Constitute an Active Protease Complex, G-Calpain, Involved in Gastric Mucosal Defense. PLoS Genet. 2010;6(7):e1001040.

319. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5.

320. Hawrylycz M, Ng L, Feng D, Sunkin S, Szafer A, Dang C. The Allen Brain Atlas. In: Kasabov N, editor. Springer Handbook of Bio-/Neuroinformatics: Springer Berlin Heidelberg; 2014. p. 1111-26.

321. Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. Nature. 2014;508(7495):199-206.

322. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014;46(10):1063-71.

323. Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, et al. High-resolution mapping and analysis of copy number variations in the human genome:

A data resource for clinical and research applications. Genome Res. 2009;19(9):1682-90.

324. Weninger WJ, Meng S, Streicher J, Mueller GB. A new episcopic method for rapid 3-D reconstruction: Applications in anatomy and embryology. Anat Embryol. 1998;197(5):341-8.

325. Rowland CA, Correa A, Cragan JD, Alverson CJ. Are Encephaloceles Neural Tube Defects? Pediatrics. 2006;118(3).

326. National Guideline C. Neural tube defects. Rockville MD: Agency for Healthcare Research and Quality (AHRQ).

327. Brown SDM, Moore MW. The International Mouse Phenotyping Consortium: past and future perspectives on mouse phenotyping. Mammalian Genome. 2012;23(9):632-40.

328. Libicher M, Tröger J. US measurement of the subarachnoid space in infants: normal values. Radiology. 1992;184(3):749-51.

329. Lam WWM, Ai VHG, Wong V, Leong LLY. Ultrasonographic measurement of subarachnoid space in normal infants and children. Pediatric Neurology. 2001;25(5):380-4.

330. Strachan DP, Rudnicka AR, Power C, Shepherd P, Fuller E, Davis A, et al. Lifecourse influences on health among British adults: Effects of region of residence in childhood and adulthood. International Journal of Epidemiology. 2007;36(3):522-31.

331. Albin RL, Young AB, Penney JB. The functional anatomy of basal ganglia disorders. Trends in Neurosciences. 1989;12(10):366-75.

332. Filion M, Tremblay Lo. Abnormal spontaneous activity of globus pallidus neurons in monkeys with MPTP-induced parkinsonism. Brain Research. 1991;547(1):140-4.

333. Filion M, Tremblay Lo, Be´dard PJ. Effects of dopamine agonists on the spontaneous activity of globus pallidus neurons in monkeys with MPTP-induced parkinsonism. Brain Research. 1991;547(1):145-9.

334. Takase K, Tamagaki C, Okugawa G, Nobuhara K, Minami T, Sugimoto T, et al. Reduced White Matter Volume of the Caudate Nucleus in Patients with Schizophrenia. Neuropsychobiology. 2004;50(4):296-300.

335. Goes FS, McGrath J, Avramopoulos D, Wolyniec P, Pirooznia M, Ruczinski I, et al. Genome-wide association study of schizophrenia in Ashkenazi Jews. Am J Med Genet B Neuropsychiatr Genet. 2015;168(8):649-59.

336. Feenstra B, Pasternak B, Geller F, Carstensen L, Wang T, Huang F, et al. Common variants associated with general and MMR vaccine-related febrile seizures. Nat Genet. 2014;46(12):1274-82.

337. Wain LV, Pedroso I, Landers JE, Breen G, Shaw CE, Leigh PN, et al. The Role of Copy Number Variation in Susceptibility to Amyotrophic Lateral Sclerosis: Genome-Wide Association Study and Comparison with Published Loci. PLoS ONE. 2009;4(12):e8175.

338. Ambalavanan A, Girard SL, Ahn K, Zhou S, Dionne-Laporte A, Spiegelman D, et al. De novo variants in sporadic cases of childhood onset schizophrenia. Eur J Hum Genet. 2016;24(6):944-8.

339. Birchler JA. Expression of cis-regulatory mutations of the white locus in metafemales of Drosophila melanogaster. Genetical Research. 1992;59(1):11-8.

340. Schuster-Böckler B, Conrad D, Bateman A. Dosage Sensitivity Shapes the Evolution of Copy-Number Varied Regions. PLoS ONE. 2010;5(3):e9474.

341. Cassina M, Rigon C, Casarin A, Vicenzi V, Salviati L, Clementi M. FBXO28 is a critical gene of the 1q41q42 microdeletion syndrome. Am J Med Genet A. 2015;167(6):1418-20.

342. The International Mouse Knockout C. A Mouse for All Reasons. Cell. 2007;128(1):9-13.

343. Van Nostrand JL, Brady CA, Jung H, Fuentes DR, Kozak MM, Johnson TM, et al. Inappropriate p53 activation during development induces features of CHARGE syndrome. Nature. 2014;514.

344. Sanlaville D, Verloes A. CHARGE syndrome: an update. Eur J Hum Genet. 2007;15(4):389-99.

345. Ray HJ, Niswander LA. Grainyhead-like 2 downstream targets act to suppress EMT during neural tube closure. Development. 2016.

346. Grunert M, Dorn C, Schueler M, Dunkel I, Schlesinger J, Mebus S, et al. Rare and private variations in neural crest, apoptosis and sarcomere genes define the polygenic background of isolated Tetralogy of Fallot. Human Molecular Genetics. 2014;23(12):3115-28.

347. Kempton MJ, Underwood TSA, Brunton S, Stylios F, Schmechtig A, Ettinger U, et al. A comprehensive testing protocol for MRI neuroanatomical segmentation techniques: Evaluation of a novel lateral ventricle segmentation method. NeuroImage. 2011;58(4):1051-9.

348. Sottocornola R, Royer C, Vives V, Tordella L, Zhong S, Wang Y, et al. ASPP2 binds Par-3 and controls the polarity and proliferation of neural progenitors during CNS development. Dev Cell. 2010;19(1):126-37.

349. Mohun TJ, Leong LM, Weninger WJ, Sparrow DB. The Morphology of Heart Development in Xenopus laevis. Dev Biol. 2000;218(1):74-88.

350. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002;12(6):996-1006.

351. Wolfe D, Dudek S, Ritchie M, Pendergrass S. Visualizing genomic information across chromosomes with PhenoGram. BioData Min. 2013;6(1):18.

352. Team RDC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria2008.

353. Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Research. 2007;35(suppl 2):W43-W6.

354. Holembowski L, Schulz R, Talos F, Scheel A, Wolff S, Dobbelstein M, et al. While p73 is essential, p63 is completely dispensable for the development of the central nervous system. Cell Cycle. 2011;10(4):680-9.

355. Mitchell G, Ballinger ML, Wong S, Hewitt C, James P, Young M-A, et al. High Frequency of Germline <italic>TP53</italic> Mutations in a Prospective Adult-Onset Sarcoma Cohort. PLoS ONE. 2013;8(7):e69026.

356. Chène P, Bechter E. P53 mutants without a functional tetramerisation domain are not oncogenic1. Journal of Molecular Biology. 1999;286(5):1269-74.

357. Kibar Z, Bosoi CM, Kooistra M, Salem S, Finnell RH, De Marco P, et al. Novel mutations in VANGL1 in neural tube defects. Human Mutation. 2009;30(7):E706-E15.

358. Allache R, De Marco P, Merello E, Capra V, Kibar Z. Role of the planar cell polarity gene CELSR1 in neural tube defects and caudal agenesis. Birth Defects Research Part A: Clinical and Molecular Teratology. 2012;94(3):176-81.

359. Merello E, Kibar Z, Allache R, Piatelli G, Cama A, Capra V, et al. Rare missense variants in DVL1, one of the human counterparts of the Drosophila dishevelled gene, do not confer increased risk for neural tube defects. Birth Defects Research Part A: Clinical and Molecular Teratology. 2013;97(7):452-5.

360. Macara IG. Parsing the Polarity Code. Nat Rev Mol Cell Biol. 2004;5(3):220-31.

361. Zak J, Schuster-Boeckler B, Bond G. Cancer Genetics May Aid Diagnostics of Developmental Disorders. Human Mutation. 2016;37(10):989-.

362. Qi H, Dong C, Chung WK, Wang K, Shen Y. Deep Genetic Connection Between Cancer and Developmental Disorders. Human Mutation. 2016;37(10):1042-50.

363. Chen X, Guo J, Lei Y, Zou J, Lu X, Bao Y, et al. Global DNA hypomethylation is associated with NTD-affected pregnancy: A case-control study. Birth defects research Part A, Clinical and molecular teratology. 2010;88(7):575-81.

364. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.

365. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.

366. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

367. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;25(17):2283-5.

368. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

369. March HN, Rust AG, Wright NA, ten Hoeve J, de Ridder J, Eldridge M, et al. Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. Nat Genet. 2011;43(12):1202-9.

370. Starr TK, Allaei R, Silverstein KAT, Staggs RA, Sarver AL, Bergemann TL, et al. A Transposon-Based Genetic Screen in Mice Identifies Genes Altered in Colorectal Cancer. Science. 2009;323(5922):1747-50.

371. Takeda H, Wei Z, Koso H, Rust AG, Yew CCK, Mann MB, et al. Transposon mutagenesis identifies genes and evolutionary forces driving gastrointestinal tract tumor progression. Nat Genet. 2015;47(2):142-50.

372. Takeda H, Rust AG, Ward JM, Yew CCK, Jenkins NA, Copeland NG. Sleeping Beauty transposon mutagenesis identifies genes that cooperate with mutant Smad4 in gastric cancer development. Proceedings of the National Academy of Sciences. 2016;113(14):E2057-E65.

373. Bard-Chapeau EA, Nguyen A-T, Rust AG, Sayadi A, Lee P, Chua BQ, et al. Transposon mutagenesis identifies genes driving hepatocellular carcinoma in a chronic hepatitis B mouse model. Nat Genet. 2014;46(1):24-32.

374. Rahrmann EP, Watson AL, Keng VW, Choi K, Moriarity BS, Beckmann DA, et al. Forward genetic screen for malignant peripheral nerve sheath tumor

formation identifies new genes and pathways driving tumorigenesis. Nat Genet. 2013;45(7):756-66.

375. Perez-Mancera PA, Rust AG, van der Weyden L, Kristiansen G, Li A, Sarver AL, et al. The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. Nature. 2012;486(7402):266-70.

376. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. Science. 2014;343(6166):80.

377. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science. 2014;343(6166):84.

378. Aguirre AJ, Meyers RM, Weir BA, Vazquez F, Zhang C-Z, Ben-David U, et al. Genomic copy number dictates a gene-independent cell response to CRISPR-Cas9 targeting. Cancer Discovery. 2016.

379. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat Chem Biol. 2016;12(2):109-16.

380. Abbott KL, Nyre ET, Abrahante J, Ho Y-Y, Isaksson Vogel R, Starr TK. The Candidate Cancer Gene Database: a database of cancer driver genes from forward genetic screens in mice. Nucleic Acids Research. 2015;43(D1):D844-D8.

381. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. Neoplasia (New York, NY). 2004;6(1):1-6.

382. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Research. 2015;43(D1):D512-D20.

383. Wherry EJ, Ha S-J, Kaech SM, Haining WN, Sarkar S, Kalia V, et al. Molecular Signature of CD8⁺ T Cell Exhaustion during Chronic Viral Infection. Immunity.27(5):824.

384. Crawford A, Angelosanto Jill M, Kao C, Doering Travis A, Odorizzi Pamela M, Barnett Burton E, et al. Molecular and Transcriptional Basis of CD4+ T Cell Dysfunction during Chronic Infection. Immunity. 2014;40(2):289-302.

385. Pircher H, Mak TW, Lang R, Ballhausen W, Rüedi E, Hengartner H, et al. T cell tolerance to Mlsa encoded antigens in T cell receptor V beta 8.1 chain transgenic mice. The EMBO Journal. 1989;8(3):719-27.

386. Boehm U, Guethlein L, Klamp T, Ozbek K, Schaub A, Fütterer A, et al. Two Families of GTPases Dominate the Complex Cellular Response to IFN-γ. The Journal of Immunology. 1998;161(12):6715-23.

387. Pestka S, Kotenko SV, Muthukumaran G, Izotova LS, Cook JR, Garotta G. The interferon gamma (IFN-γ) receptor: a paradigm for the multichain cytokine receptor. Cytokine & Growth Factor Reviews. 1997;8(3):189-206.

388. Pernis A, Gupta S, Gollob K, Garfein E, Coffman R, Schindler C, et al. Lack of interferon gamma receptor beta chain and the prevention of interferon gamma signaling in TH1 cells. Science. 1995;269(5221):245-7.

389. Iwata S, Mikami Y, Sun H-W, Brooks SR, Jankovic D, Hirahara K, et al. The Transcription Factor T-bet Limits Amplification of Type I IFN Transcriptome and Circuitry in T Helper 1 Cells. Immunity. 2017;46(6):983-91.e4.

390. Su LK, Kadesch T. The immunoglobulin heavy-chain enhancer functions as the promoter for I mu sterile transcription. Molecular and Cellular Biology. 1990;10(6):2619-24.

391. Stubbington MJT, Corcoran AE. Non-coding transcription and large-scale nuclear organisation of immunoglobulin recombination. Current Opinion in Genetics & Development. 2013;23(2):81-8.

392. Leong YA, Chen Y, Ong HS, Wu D, Man K, Deleage C, et al. CXCR5+ follicular cytotoxic T cells control viral infection in B cell follicles. Nat Immunol. 2016;17(10):1187-96.

393. Utzschneider DT, Charmoy M, Chennupati V, Pousse L, Ferreira DP, Calderon-Copete S, et al. T Cell Factor 1-Expressing Memory-like CD8+ T Cells Sustain the Immune Response to Chronic Viral Infections. Immunity. 2016;45(2):415-27.

394. Teijaro JR, Ng C, Lee AM, Sullivan BM, Sheehan KCF, Welch M, et al. Persistent LCMV Infection Is Controlled by Blockade of Type I Interferon Signaling. Science. 2013;340(6129):207.

395. Wilson EB, Yamada DH, Elsaesser H, Herskovitz J, Deng J, Cheng G, et al. Blockade of Chronic Type I Interferon Signaling to Control Persistent LCMV Infection. Science (New York, NY). 2013;340(6129):202-7.

396. Leong YA, Chen Y, Ong HS, Wu D, Man K, Deleage C, et al. CXCR5+ follicular cytotoxic T cells control viral infection in B cell follicles. Nat Immunol. 2016.

397. Schoggins JW, Rice CM. Interferon-stimulated genes and their antiviral effector functions. Current opinion in virology. 2011;1(6):519-25.

398. Mostafavi S, Yoshida H, Moodley D, LeBoité H, Rothamel K, Raj T, et al. Parsing the Interferon Transcriptional Network and Its Disease Associations. Cell. 2016;164(3):564-78.

399. Murray PJ. The JAK-STAT Signaling Pathway: Input and Output Integration. The Journal of Immunology. 2007;178(5):2623.

400. Pflanz S, Hibbert L, Mattson J, Rosales R, Vaisberg E, Bazan JF, et al. WSX-1 and Glycoprotein 130 Constitute a Signal-Transducing Receptor for IL-27. The Journal of Immunology. 2004;172(4):2225.

401. Hirahara K, Onodera A, Villarino Alejandro V, Bonelli M, Sciumè G, Laurence A, et al. Asymmetric Action of STAT Transcription Factors Drives Transcriptional Outputs and Cytokine Specificity. Immunity.42(5):877-89.

402. Kim JI, Ho IC, Grusby MJ, Glimcher LH. The Transcription Factor c-Maf Controls the Production of Interleukin-4 but Not Other Th2 Cytokines. Immunity. 1999;10(6):745-51.

403. Pauken KE, Sammons MA, Odorizzi PM, Manne S, Godec J, Khan O, et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. Science. 2016.

404. Malandro N, Budhu S, Kuhn Nicholas F, Liu C, Murphy Judith T, Cortez C, et al. Clonal Abundance of Tumor-Specific CD4+ T Cells Potentiates Efficacy and Alters Susceptibility to Exhaustion. Immunity. 2016;44(1):179-93.

405. Wu T, Ji Y, Moseman EA, Xu HC, Manglani M, Kirby M, et al. The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. Science immunology. 2016;1(6):eaai8593.

406. Ahmed R, Salmi A, Butler LD, Chiller JM, Oldstone MB. Selection of genetic variants of lymphocytic choriomeningitis virus in spleens of persistently infected mice. Role in suppression of cytotoxic T lymphocyte response and viral persistence. The Journal of Experimental Medicine. 1984;160(2):521.

407. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

408. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):1-21.

409. Kolde R. pheatmap: Pretty Heatmaps. . 1.0.8 ed2013.

410. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545-50.

411. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Research. 2013;41(D1):D991-D5.

412. Monier B, Gettings M, Gay G, Mangeat T, Schott S, Guarner A, et al. Apico-basal forces exerted by apoptotic cells drive epithelium folding. Nature. 2015;518(7538):245-8.