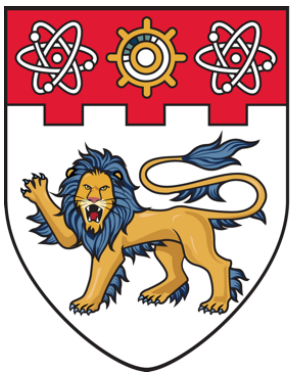


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

FUNCTIONALIZING NOVEL MODULATORS OF LUNG CANCER

LEE YI FEI

SCHOOL OF BIOLOGICAL SCIENCES

2022

FUNCTIONALIZING NOVEL MODULATORS OF LUNG CANCER

LEE YI FEI

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University in partial fulfilment of the requirement for the degree of Doctor of Philosophy

2022

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

14/01/2022

. . . . . .
Date

. . . . . .
LEE YI FEI

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

14/01/2022

. . . . . .
Date

. . . . . .
TAM WAI LEONG

Authorship Attribution Statement

This thesis contains material from one paper published in the following peer-reviewed journal in which I am listed as an author.

Chapters 1 and 3 reference figures published in:

Chen J, Yang H, Teo ASM, Amer LB, Sherbaf FG, Tan CQ, Alvarez JJS, Lu B, Lim JQ, Takano A, Nahar R, Lee YY, Phua CZJ, Chua KP, Suteja L, Chen PJ, Chang MM, Koh TPT, Ong BH, Anantham D, Hsu AAL, Gogna A, Too CW, Aung ZW, **Lee YF**, Wang L, Lim TKH, Wilm A, Choi PS, Ng PY, Toh CK, Lim WT, Ma S, Lim B, Liu J, Tam WL, Skanderup AJ, Yeong JPS, Tan EH, Creasy CL, Tan DSW, Hillmer AM, Zhai W. Genomic landscape of lung adenocarcinoma in East Asians. *Nature Genetics.* 2020 Feb;52(2):177-186. doi: 10.1038/s41588-019-0569-6.

14/01/2022

. . . . . .
Date

. . . . . .
LEE YI FEI

Acknowledgments

Firstly, I would like to express my gratitude to my supervisor Dr. Tam Wai Leong for his guidance and advice over the years. He has given me the freedom to explore and also taught me to ask more focused questions, which have enabled me to grow as a scientist. I am also thankful to my co-supervisor Dr. Andrew Tan for his support and incisive feedback that has been helpful to my project and learning. To my thesis advisory committee members Dr. Koh Cheng Gee and Dr. Newman Sze, thank you for your insightful comments and questions. I always emerge from these meetings with fruitful takeaways.

This work would not have been possible without the efforts of Dr. Jianbin Chen, Dr. Zhai Weiwei and Dr. Daniel Tan, as well as other members of the Lung Open Fund-Large Collaborative Grant team, in driving the recruitment and analysis of the largest Asian lung adenocarcinoma cohort to date. The findings from this prior work were pivotal to the foundation of my project.

Huge thanks go to Cheryl Phua, Dr. Zhang Bin and Dr. Yvonne Tay for their help with the bioinformatics analyses, Dr. Dennis Kappei and Charlene Chan for their help with the mass spectrometry, as well as Dr. Yuan Ju, Edwin Lim, Lim Yuhua and Casslynn Koh for their experimental help.

To Dr. Lee May Yin, Dr. Wang Zhenxun, Dr. Loo Ser Yue and Dr. Wu Zhengwei, thank you for giving me new perspectives, for inspiring me as a scientist, and for always being so generous and readily available with your helpful advice.

PhD life has its many ups and downs, and through it all I am especially thankful to Dr. Yeo Xun Hui, Dr. Toh Li Ping and Shirley Lam for the helpful discussions, companionship, and refuelling in the form of exercise, good food and endorphins.

To the Tam lab members past and present, thank you for walking this journey with me and adding so much more color, laughter, and joy. Dr. Jane Lee, Dr. Elina Pathak, Dr. Lim Jing Shan, Joanna Yeo, Auntie Rani, Valarie Tham,

Kimberly Choo, Rebecca Lim, Jessica Lidster, William Xie, Shathishwaran and Chng Wee Ling, thanks for making lab a happy place and for your help in one way or another.

Furthermore, thank you to Adai Ramasamy, Gokce Oguz and Yang Jian for introducing me to R and guiding me as I wrote my first lines of code. Coding is still tough, but nowhere near as daunting as when I first started.

I would like to thank the Agency for Science, Technology and Research, Singapore (A*STAR) for providing me with funding support and resources for my PhD. I am also thankful to the Genome Institute of Singapore (A*STAR) for the facilities and infrastructure that have enabled my research. I would also like to acknowledge the School of Biological Sciences at Nanyang Technological University for supporting me in my education as well as administratively.

Finally, I am immensely thankful to my parents and my sister for their unwavering love all my life. I am very fortunate to have my supportive family and I would like to dedicate this thesis to them.

Table of Contents

Summary

Lung cancer is the leading cause of cancer mortality globally. The discovery of driver mutations has paved the way for the use of targeted therapy in the management of advanced disease. However, the effectiveness of targeted therapy is limited firstly by the eventual development of treatment resistance, and secondly by the sizeable percentage of cases in which the driver mutations remain unknown. Furthermore, lung adenocarcinoma (LUAD) in Asians is known to exhibit a molecular profile distinct from that in Caucasians, yet the majority of existing cohort studies have been biased towards the latter. To address this gap, we and our collaborators have spearheaded the largest Asian LUAD cohort study to date. Seven potential novel LUAD drivers were discovered from whole exome sequencing data from this cohort. We observed recurrent mutations in one of these genes, PARP4, that had also been observed across different cancer cohorts.

PARP4 levels were observed to be higher in normal lung cells as opposed to lung cancer cells. In LUAD patients, reduced PARP4 expression levels were further correlated with poorer prognosis, suggesting that PARP4 is a putative tumor suppressor. In line with my hypothesis, I observed loss of PARP4 copy number both within our cohort as well as the Caucasian TCGA LUAD cohort. Interestingly, PARP4 mutations were found to be associated with a Chinese ethnic bias when compared to the TCGA dataset, suggesting why PARP4's putative role as a LUAD modulator was not previously discovered. I was able to show that loss of PARP4 increases the *in vitro* and *in vivo* tumorigenicity of lung cancer cell lines, thereby functionally suggesting PARP4's tumor suppressive activity.

Furthermore, through site-directed mutagenesis experiments, I propose that the recurrent PARP4 I1039T mutation is in effect a loss-of-function mutation which may destabilize PARP4 protein to result in increased tumorigenicity. The PARP catalytic domain also appears to be important for PARP4's tumor suppressive activity.

In cells, a fraction of PARP4 resides within the vault complex, of which PARP4's interaction partner MVP forms the core component. Strikingly, loss of MVP does not significantly increase tumorigenicity, suggesting a tumor suppressive role for PARP4 that is independent of the vault complex.

To uncover novel PARP4 interaction partners, a SILAC co-immunoprecipitation mass spectrometry analysis was performed. Following this with targeted validation experiments, I successfully identified hnRNPM, a splicing regulatory protein with pleiotropic functions, as a novel PARP4 binding partner whose loss phenocopies PARP4 loss. PARP4 or hnRNPM loss results in several shared phenotypes that include dysregulated splicing.

I have thus managed to functionally validate a new cancer gene relevant to LUAD based on sequencing data from a clinically relevant cohort. Moreover, PARP4 action may be mediated through a previously undescribed interacting partner, hnRNPM.

List of Figures

List of Tables

Abbreviations

A3SS	Alternative 3’ splice site
A5SS	Alternative 5’ splice site
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BRCA1	Breast cancer type 1 susceptibility protein
BRCA2	Breast cancer type 2 susceptibility protein
BRCT	BRCA1 C Terminus
cDNA	Complementary deoxyribonucleic acid
CN	Copy number
COSMIC	Catalogue Of Somatic Mutations In Cancer
Co-IP	Co-immunoprecipitation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial to Mesenchymal Transition
FC	Fold change
FDA	US Food and Drug Administration
FDR	False discovery rate
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescent protein
gRNA	Guide RNA
hnRNP	Heterogeneous nuclear ribonucleoprotein
hnRNPM	Heterogeneous nuclear ribonucleoprotein M
HRP	Horseradish peroxidase
IC50	Half-maximal inhibitory concentration
IR	Intron retention
KO	Knockout

KRAS	Kirsten rat sarcoma virus
KRT18	Keratin 18
LUAD	Lung adenocarcinoma
MAPK	Mitogen-activated protein kinase
MAR	Mono-ADP-ribose
MES	Multiple exon skipping
mRNA	Messenger RNA
miRNA	MicroRNA
MVP	Major vault protein
NAD	Nicotinamide adenine dinucleotide
NLS	Nuclear localization signal
NSG	NOD *scid* gamma
p53	Tumor protein p53 (protein)
PAR	Poly-ADP-ribose
PARP	Poly (ADP-ribose) polymerase
PARP1	Poly (ADP-ribose) polymerase 1
PARP4	Poly (ADP-ribose) polymerase 4
PCM1	Pericentriolar material 1
PCNT	Pericentrin-1
PCR	Polymerase chain reaction
PHB	Prohibitin
PHB2	Prohibitin-2
PSI	Percent spliced in
PTEN	Phosphatase and Tensin Homolog
RB	Retinoblastoma protein
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RT-qPCR	Quantitative reverse transcription PCR
SES	Single exon skipping
shRNA	Short hairpin RNA

siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
SSO	Splice switching oligonucleotide
SV40LT	Simian virus 40 large T antigen
SILAC	Stable isotope labelling by amino acids in cell culture
TCGA	The Cancer Genome Atlas
TERT	Human telomerase reverse transcriptase
TP53	Tumor protein p53 (gene)
TKI	Tyrosine kinase inhibitor
VIT	Vault protein inter-alpha-trypsin
VWFA	Von Willebrand factor type A
WES	Whole exome sequencing
WGS	Whole genome sequencing

CHAPTER 1: Introduction

1.1 Lung cancer and targeted therapy

Lung cancer consistently ranks amongst the three most frequently occurring cancers and is the leading cause of cancer mortality worldwide (Jemal et al., 2011; Ridge et al., 2013; Siegel et al., 2018). Lung cancer refers to a group of diseases highly heterogeneous in both phenotypic and molecular terms. Among these, lung adenocarcinoma (LUAD) is the most prevalent histologic subtype, accounting for 38.5% of all lung cancer cases (Dela Cruz et al., 2011; Herbst et al., 2008). In spite of advances in diagnosis and treatment, lung cancer prognosis remains poor, with the five-year survival rate having plateaued at 15% over the past four decades (Yang, 2009; Zappa and Mousa, 2016). This underlines the need for better lung cancer management strategies.

The discovery of LUAD driver mutations has paved the way for the use of targeted therapy in the management of advanced LUAD. Not all mutations within cancer cells affect processes relevant to cancer development. Driver mutations refer specifically to gene mutations that confer a fitness advantage to the cell and thus promote cancer initiation or progression (Bozic et al., 2010; Stratton et al., 2009). In contrast, passenger mutations are thought not to functionally alter cell fitness, but occur incidentally due to the high mutation rates and genomic instability inherent to cancer (Haber and Settleman, 2007).

Driver mutations are typically activating mutations in oncogenes and loss of function mutations in tumor suppressors (Pon and Marra, 2015; Stratton et al., 2009). Correspondingly, the genes harboring these mutations are collectively termed driver genes. The dependency of tumor cells on specific oncogenic signaling pathways provides a means of preferentially targeting the cancer via inhibition of these oncogenic pathways while sparing normal cells (Bartholomew et al., 2017).

There are several known drivers of LUAD (Figure 1-1). One driver particularly relevant to LUAD is the epidermal growth factor receptor (EGFR), which is a member of the receptor tyrosine kinase family and signals downstream

to the RAS/RAF/MEK/MAPK and PI3K/AKT/mTOR pathways (Liu et al., 2018; Wee and Wang, 2017). Amongst LUAD cases, EGFR frequently undergoes gene amplification or gains hyperactivating mutations within its kinase domain, leading to elevated signaling to enhance cell proliferation, migration and survival (Forde and Ettinger, 2013; Ladanyi and Pao, 2008).

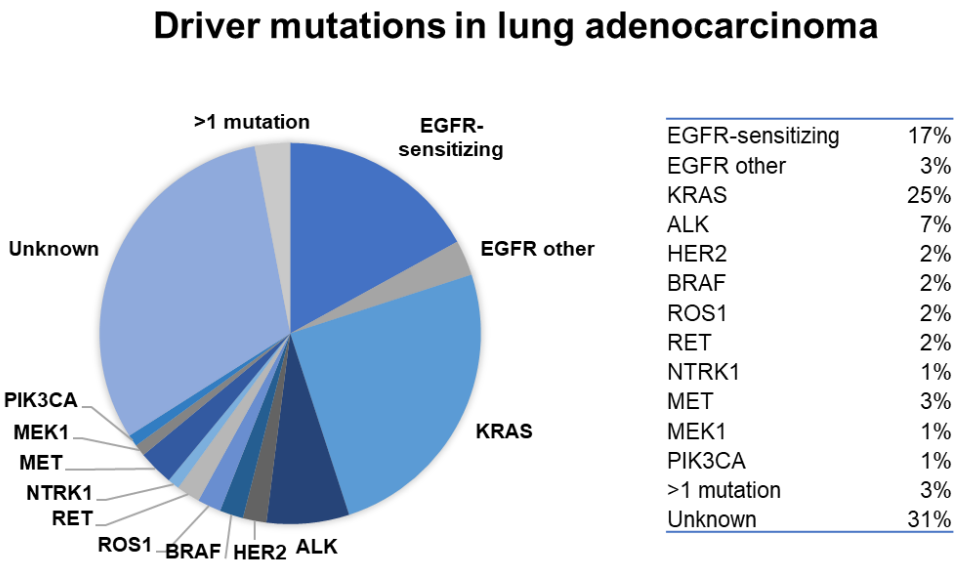


Figure 1-1: Driver mutations and their mutational frequency in lung adenocarcinoma

Data retrieved from (Hirsch et al., 2017).

Gefitinib and erlotinib are FDA-approved small molecule inhibitors of EGFR that act by binding to EGFR's ATP binding pocket within the kinase domain, thus preventing the binding and hydrolysis of ATP required for the phosphorylation of downstream effectors. Due to their demonstrated ability to prolong survival in patients bearing EGFR sensitizing mutations, these tyrosine kinase inhibitors (TKIs) are now routinely prescribed to advanced LUAD patients bearing EGFR Exon 19 deletions or Exon 21 L858R substitutions (Lynch et al., 2004; Pao and Chmielecki, 2010; Sequist et al., 2007; Shigematsu et al., 2005).

Apart from EGFR, other LUAD driver mutations have also been successfully targeted, with a list of their corresponding FDA-approved drugs reflected in Table 1-1.

Table 1-1: List of FDA-approved kinase inhibitors for use in LUAD patients

FDA-approved Kinase Inhibitors (KIs)

EGFR TKIs	Erlotinib Gefitinib	EGFR sensitizing mutations (e.g. L858R, exon19del)
ALK TKIs	Crizotinib Ceritinib	ALK rearrangements
ROS1 TKI	Crizotinib	ROS1 rearrangements
BRAF combination KIs	Dabrafenib Trametinib	BRAF V600E

However, patients who initially respond well eventually acquire resistance to TKIs 6-12 months later, rendering these targeted therapies useless (Gazdar, 2009; Huang and Fu, 2015; Pao et al., 2004; Tomasello et al., 2018; Wang et al., 2016). This points to the need for alternative therapeutic strategies. The identification of additional driver mutations could thus provide further drug targets and potentially allow for more effective combination strategies.

Furthermore, TKI-responsive cases form only a minority of all LUAD cases. In reality, a sizeable percentage of LUAD cases have no identified driver mutations (Figure 1-1), and are thus difficult to target specifically (Pao and Girard, 2011). There is hence a pressing need to identify novel driver mutations in LUAD, which may offer up alternative therapeutic opportunities.

1.2 Driver prediction efforts from sequencing data

The prediction of cancer driver genes from cohort sequencing data has been an active field of research and has been greatly enabled by advances in sequencing technology, reduced costs of sequencing and development of better pipelines (Bailey et al., 2018; Gordon et al., 2020). Several large cohort studies have previously been performed to characterize the genomic landscape of LUAD, serving as the foundation on which driver prediction algorithms could be applied (Campbell et al., 2016; Imielinski et al., 2012; The Cancer Genome Atlas Research Network, 2014).

Approaches to predicting driver genes typically fall within three categories. Frequency-based approaches such as MutSigCV and MuSiC identify genes that have more mutations than expected from the background somatic mutation rate, adjusting for variables such as gene size, base context and gene expression (Dees et al., 2012; Lawrence et al., 2013, 2014; Luo et al., 2019). Function-based methods such as 20/20 and OncodriveFM account for the nature and expected functional impacts of the mutations within a gene, as well as whether the mutations are clustered within particular regions on the gene (Davoli et al., 2013; Gonzalez-Perez and Lopez-Bigas, 2012; Tamborero et al., 2013; Vogelstein et al., 2013). Thirdly, pathway or interaction network analysis methods such as DawnRank identify genes that have known roles in pathways relevant to cancer, or with known effects on transcriptional patterns (Chen et al., 2020a; Collier et al., 2019; Hou and Ma, 2014; Tokheim et al., 2016; Watson et al., 2013). Collectively, these approaches have been used for the prediction of drivers in different cancers.

1.3 Ethnic considerations in LUAD cohort studies

There are notable differences in the nature of LUAD between Caucasian and Asian populations. Caucasian LUAD tend to be dominated by male smokers whereas LUAD in East Asians are enriched with female non-smokers (Chen et al., 2020c; Sun et al., 2007; Tseng et al., 2019). Notably, LUAD in East Asians tend to be early onset, especially in non-smokers (Kawaguchi et al., 2010). Furthermore, EGFR mutations are present in 40-60% of Asian LUAD cases but make up only 7-10% of Caucasian LUAD, among which KRAS mutations are more prevalent (Ha et al., 2015; Kohno et al., 2015; Li et al., 2011; Shi et al., 2014; Yang et al., 2020). Other epidemiological differences between Caucasian and Asian LUAD take the form of differing incidence rates, risk factors, response to targeted therapies and prognosis (Zhou and Christiani, 2011).

At the time of conceptualization of this study, the number of sequencing studies focused on Asian LUAD were far and few, with the majority still heavily biased toward Caucasian data to date (Table 1-2). The disparities between Asian and Caucasian LUAD meant that potential driver genes relevant to the Asian clinical context may be missed from studies based primarily on Caucasian cohorts.

Table 1-2: Summary of existing LUAD cohort studies with paired tumor and normal

Ethnicity	Study Description	No. of LUAD Cases	Staging	Reference
Caucasian	Copy number analysis using microarrays	371	Stage I-IV	(Weir et al., 2007)
Caucasian	Somatic mutation analysis by SNP arrays followed by targeted re-sequencing	188	Stage I-IV	(Ding et al., 2008)
Not specified. Samples from Washington University	Whole genome sequencing (WGS) and transcriptome sequencing (RNA-seq)	16	Not specified	(Govindan et al., 2012)
Caucasian	WGS and whole exome sequencing (WES)	183	Stage I-IV	(Imielinski et al., 2012)
Caucasian	WES and RNA-seq	230	Stage I-IV	(The Cancer Genome Atlas Research Network, 2014)
Caucasian	WES (274 new, combined with Imielinski et al., 2012 and TCGA, 2014)	660	Stage I-IV	(Campbell et al., 2016)
Caucasian	WES	108	Stage I-III	(Kadara et al., 2017)
Caucasian (8% Asian)	Hybridization capture-based sequencing assay for 341 key cancer genes	860	Recurrent or metastatic disease	(Jordan et al., 2017)
Caucasian	WGS (never smoker)	232	Stage I-III	(Zhang et al., 2021)
Caucasian	WGS of samples from (Campbell et al., 2016) that lack RTK/RAS/RAF pathway alterations	85	Stage I-IV	(Carrot-Zhang et al., 2021)
Asian	WGS and WES	101	Stage I-IV	(Wu et al., 2015)

Asian	Screening for known EGFR, KRAS, ALK, ROS1, and RET mutations, WGS for 33 samples lacking mutations in the 5 genes. Focus on female nonsmokers.	271	Stage I-IV	(Li et al., 2016)
Asian	WGS and WES	57 (WGS) 27 (WES)	Stage I-IV	(Wang et al., 2018b)
Asian	WGS (never-smoker)	36	Stage I-IV	(Luo et al., 2018)
Asian	WES and RNA-seq	131	Stage I-IV	(Zhang et al., 2019b)
56% Asian 31% Caucasian	WGS (49 new, combined with Imielinski et al., 2012, TCGA, 2014, Wu et al., 2015)	138	Stage I-IV	(Lee et al., 2019)
Asian	**WES (n=302) and RNA-seq (n=181) (213 new, combined with Wu et al., 2015)**	**305**	**Stage I-IV**	**(Chen et al., 2020b)**

1.4 Asian LUAD cohort study

These differences formed the rationale for our own large Asian LUAD cohort study. In this study spearheaded by Dr. Jianbin Chen, Dr. Weiwei Zhai and Dr. Daniel Tan (Chen et al., 2020b), 213 Singaporean LUAD patients of Chinese ethnicity were recruited from the National Cancer Center of Singapore (NCCS). Whole exome sequencing (WES) was performed for 210 of these patients while RNA sequencing was conducted for 181 patients. The WES data was combined with that from 92 Chinese patients from an independently published study (Wu et al., 2015) for analysis. Collectively, this cohort represents the largest known Asian LUAD clinical dataset, and could allow the prediction of novel drivers relevant to Asian LUAD.

1.4.1 Seven potential novel drivers discovered from whole exome sequencing of an Asian LUAD cohort

Two computational approaches – MutSigCV (Lawrence et al., 2013) and 20/20+ (Tokheim et al., 2016) - were applied to the whole-exome sequencing data from the LUAD cohort to discover potential new LUAD drivers. These two algorithms function on different basic principles and may pick up different potential drivers. The first of these – MutSigCV – is a recurrence-based algorithm which identifies genes bearing mutation rates higher than background as drivers (Lawrence et al., 2013). The second method (20/20+) is a function-based algorithm that employs machine learning approaches to identify drivers based on a compendium of mutational patterns. Comparative analyses of multiple tools had shown that these two methods perform better, identifying drivers in a more stringent manner (Tokheim et al., 2016).

Using these methods, 27 driver genes were identified at a false discovery rate of less than 0.1 (Figure 1-2). As expected, the most frequent drivers were EGFR (47%), TP53 (36%), and KRAS (11%). While the majority of these genes correspond to known LUAD drivers or have been implicated in other cancers, seven other genes showed up as putative novel drivers (Figure 1-2, orange boxes). From literature and earlier cohort studies, these seven genes have not been previously reported to drive LUAD. A summary of the seven candidate novel drivers and their functions can be found in Table 1-3.

Of these, PARP4 was mutated at the highest frequency of 6% (Figure 1-2B), placing it among the ranks of other well-known LUAD drivers such as KEAP1 and STK11 (both at 5%) (Cardnell et al., 2015; Kadara et al., 2017).

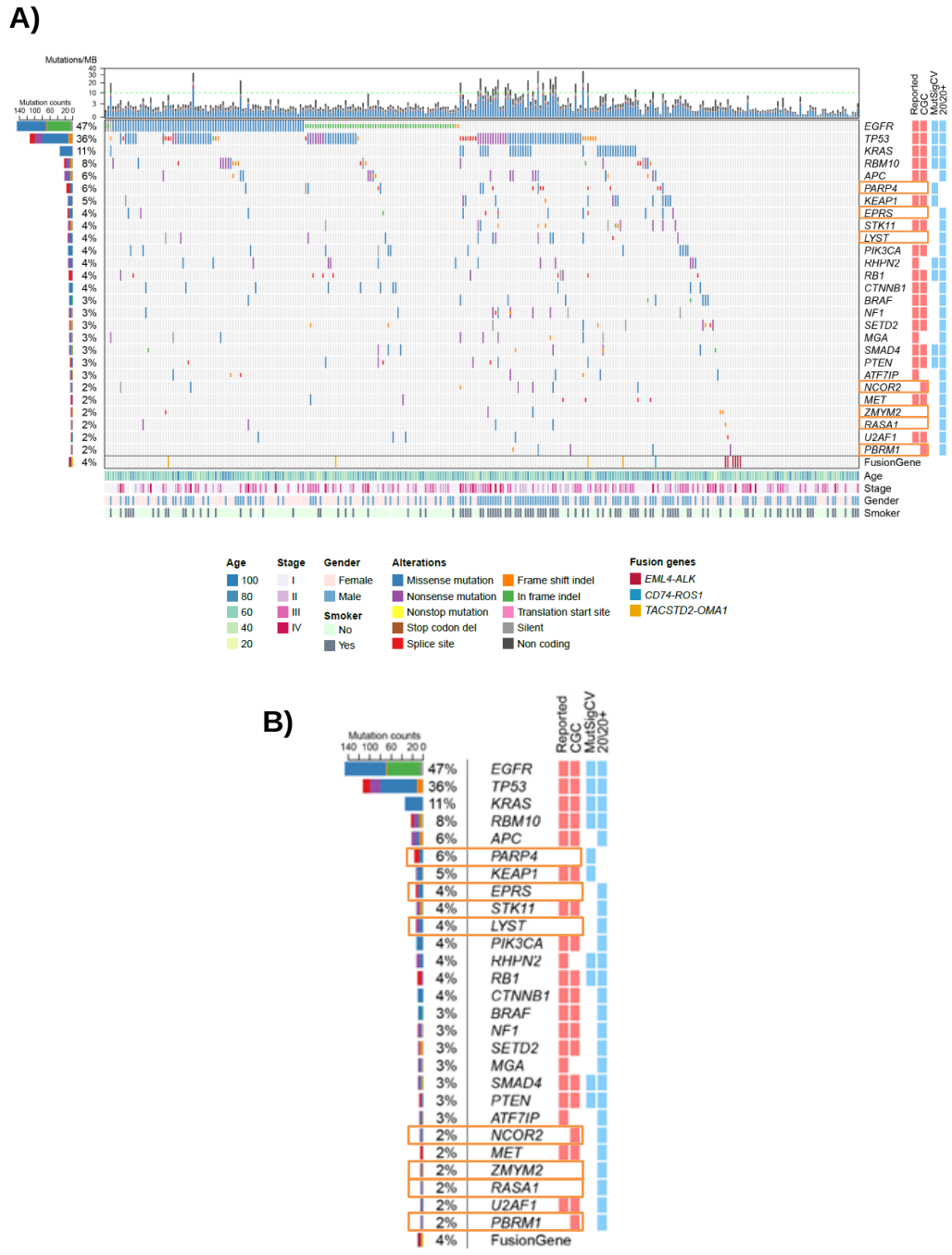


Figure 1-2: Oncoprint of the majority Singaporean LUAD cohort

A) Each bar represents a single patient, with details of gender, age, smoking status and cancer stage indicated on the X-axis. Driver genes are displayed along the Y-axis, in order of decreasing mutation frequency. The seven genes highlighted in orange are putative novel drivers which have not previously been implicated in LUAD. Analysis performed by Dr. Jianbin Chen. Figure reproduced from (Chen et al., 2020b). **B)** Enlargement of predicted driver genes and their mutation frequencies from **A)**.

Table 1-3: List of candidate novel LUAD drivers alongside their associated functions.

Genes are listed in decreasing order of mutation frequency within the cohort.

Candidate Driver and Mutation Frequency	Brief Description of Functions	References
PARP4 (6%)	(ADP)-ribosyl transferase; component of the vault complex); downregulated in more aggressive colorectal cancer cell lines; PARP4 loss promotes carcinogen-induced tumorigenesis in mice; germline variants associated with primary breast and thyroid cancers	(Ikeda et al., 2016; Long et al., 2016; Raval-Fernandes et al., 2005)
EPRS (4%)	Glutamyl-prolyl-tRNA synthetase; upregulated in estrogen receptor positive breast tumors; regulates cell cycle and estrogen response genes	(Katsyv et al., 2016)
LYST (4%)	Lysosomal trafficking regulator; LYST gene silencing inhibits cell proliferation and induces apoptosis in myeloma cells; LYST is inactivated at high frequency in chordoma	(Bong et al., 2016; Tarpey et al., 2017)
NCOR2 (2%)	Nuclear receptor co-repressor: member of thyroid hormone- and retinoic acid receptor-associated co-repressors; upregulated in oral squamous cell carcinoma, endometrial carcinoma and breast cancer	(Kershah et al., 2004; Osathanon et al., 2016; Smith et al., 2012)
ZMYM2 (2%)	Zinc Finger, MYM-type 2; possible transcription factor or member of BHC histone deacetylase complex; ZMYM2 inhibition resulted in sorafenib resistance in hepatocellular carcinoma cell lines	(Nam et al., 2014)
RASA1 (2%)	RAS p21 protein activator 1; GAP1 family of GTPase-activating proteins; RASA1 downregulation associated with poor prognosis in hepatocellular carcinoma and breast invasive ductal carcinoma	(Chen et al., 2017; Liu et al., 2015)
PBRM1 (2%)	Chromatin targeting subunit of the PBAF SWI/SNF chromatin remodeling complex; regulator of p21 expression for cell cycle arrest and senescence	(Xia et al., 2008)

1.5 Introduction to PARP4

1.5.1 PARP4 and the vault complex

Poly (ADP-ribose) polymerase 4 (PARP4) is an (ADP)-ribosyl transferase that was first discovered as a component of the vault complex, which is a barrel-shaped ribonucleoprotein assembly that derives its name from its resemblance to the vaulted ceilings of gothic cathedrals (Kickhoefer et al., 1999a). Following its initial sighting by electron microscopy within preparations of clathrin-coated vesicles isolated from rat liver, the vault complex was later found to be highly conserved among many eukaryotes (Kedersha and Rome, 1986; Kedersha et al., 1990). Gradient centrifugation methods were subsequently used to purify these complexes, while biochemical methods were employed to study their components and properties (van Zon et al., 2003a).

1.5.2 Vault complex structure

The vault complex comprises an external shell with 8-fold symmetry formed by 78 copies of the major vault protein (MVP) (Figure 1-3A), which as its name suggests, is the main component of the vault complex accounting for more than 70% of its mass (Kedersha et al., 1991). Several components are located inside the hollow lumen of the vault: PARP4, telomerase-associated protein-1 (TEP1) as well as vault RNAs, which are small untranslated RNA molecules 88-141 bases in length (Figure 1-3B) (Kickhoefer et al., 1999a, 1999b). Altogether, the vault complex is a large assembly approximately 13 megadaltons (MDa) in size (Kedersha et al., 1991). For reference, the human ribosome is only about 4.3 MDa in size (Sulima et al., 2017).

MVP is thought to be the only component necessary for the vault structure, as overexpression of MVP alone in insect or bacterial cells lacking endogenous vault complexes led to its self-assembly into vault-like structures when examined by electron microscopy (Figure 1-3C) (Stephen et al., 2001; Zheng et al., 2004). Furthermore, vaults purified from the livers of mice genetically deficient in either PARP4 or TEP1 remained structurally intact (Kickhoefer et al., 2001; Liu et al., 2004). In the latter however, vault RNAs did not co-purify with the vault complex,

suggesting that TEP1 is required for vault RNA sequestration within the complex (Kickhoefer et al., 2001).

1.5.3 Vault complex localization and function

While vault complexes predominantly localize to the cytoplasm, a subset has been observed at the nuclear periphery, adjacent to the nuclear pore complex (Chugani et al., 1993). Together with the observation that a fraction of MVP co-localized with cytoskeletal components, these findings prompted suggestions of the vault complex's involvement in nucleocytoplasmic transport of cargo (Hamill and Suprenant, 1997; Herrmann et al., 1999; Slesina et al., 2006; van Zon et al., 2002). However, another study using GFP-tagged MVP found no evidence of GFP-labelled vaults in the nucleus, and depletion of MVP did not affect nuclear import or export kinetics, thus casting doubt on this hypothesis (van Zon et al., 2006).

Apart from nucleocytoplasmic shuttling, the vault complex has also been proposed to mediate drug resistance. This stemmed from observations that MVP expression levels were upregulated in several cancers and were positively correlated with the multidrug resistance phenotype (Berger et al., 2001; Hu et al., 2002; Kickhoefer et al., 1998; Komarov et al., 1998; Meijer et al., 1999; Scheffer et al., 1995; Scheper et al., 1993; Stein et al., 2005). However, siRNA depletion of MVP from multidrug resistant lung cancer cells did not sensitize them to the chemotherapeutic doxorubicin. MVP overexpression in the parental cells also could not confer resistance (Huffman and Corey, 2005). Furthermore, MVP knockout mice responded similarly to wildtype mice to doxorubicin treatment, and cells derived from both groups of mice were equally sensitive to *in vitro* treatment with several cytostatic agents (Mossink et al., 2002). Hence, the role of vaults and MVP in mediating drug resistance still requires further clarification.

The vault complex was additionally proposed to serve as a protein scaffold for signal transduction. In particular, MVP was observed to enhance interactions between the BRAF and MEK kinases to promote activation of the mitogen-activated protein kinase (MAPK) signaling pathway (Liu et al., 2019).

Furthermore, co-immunoprecipitation experiments revealed that MVP interacts with Phosphatase and Tensin Homolog **(**PTEN), which is known to dephosphorylate phosphoinositides in the regulation of the Protein kinase B (AKT) signaling pathway (Chen et al., 2018). PTEN was also co-purified with vault complexes from gradient centrifugation performed on HeLa cervical cancer cells (Yu et al., 2002). In breast cancer cells, MVP had also been reported to form a complex with SH2 domain-containing protein tyrosine phosphatase-2 (SHP2) and extracellular signal-regulated kinases (ERKs), suggesting a possible role as a scaffold protein (Kolli et al., 2004).

While the vault complex has been proposed to play a number of different roles, several of these studies used levels of MVP as a proxy for that of the vault complex, and it would be important to study the contribution of the other minor vault components to these functions.

1.5.4 Vault-independent roles for components of the vault complex

While MVP is mostly incorporated into the vault complex, non-vault-associated roles have been proposed for the other components (van Zon et al., 2003b).

Notably, TEP1 is shared between the vault complex and telomerase (Figure 1-3D) (Harrington et al., 1997; Kickhoefer et al., 1999b). However, TEP1 is likely not functionally important for telomere extension, as telomerase activity or telomere length were not affected in TEP1-deficient mice (Beattie et al., 1998; Kickhoefer et al., 1999b).

Vault RNAs have been reported to be further processed into smaller regulatory RNAs, which function similarly to microRNAs to negatively regulate gene expression (Persson et al., 2009), Specifically, vault RNAs have been implicated in the regulation of epidermal differentiation and autophagy (Horos et al., 2019; Hussain et al., 2013; Sajini et al., 2019).

In the case of PARP4, fractionation and immunofluorescence experiments revealed that PARP4's subcellular distribution overlaps only

partially with that of MVP (Kickhoefer et al., 1998, 1999a). In fact, non-vault-associated fractions of PARP4 have been reported in both the cytoplasm and nucleus, although a later study suggested that the nuclear localization could have been an artefact (Figure 1-3E, F) (Schroeijers et al., 2000; Vyas et al., 2013; van Zon et al., 2003b). PARP4 had also been observed to localize to the mitotic spindle of dividing cells (Kickhoefer et al., 1999a). However, the functional significance of PARP4 remains to be understood.

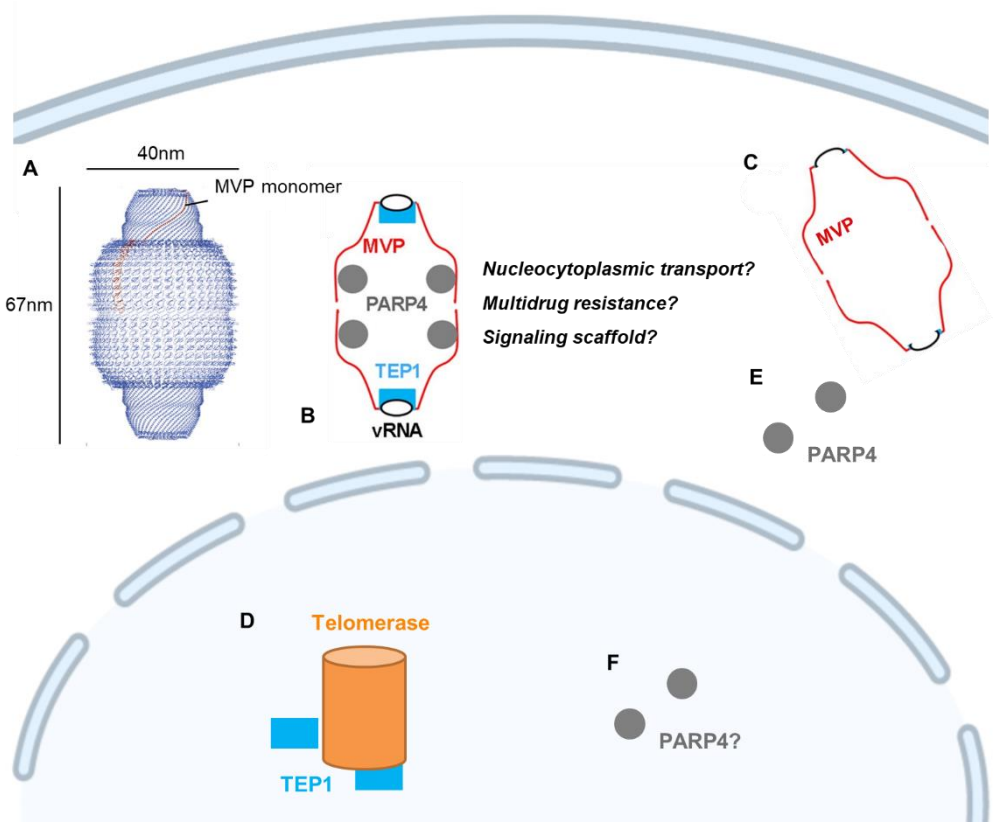


Figure 1-3: Schematic diagram of the vault complex and its proposed functions

A) Diagram representing the vault complex outer shell formed by 78 copies of the MVP monomer (in orange) and its dimensions. **B)** Schematic diagram of vault complex constituents and proposed roles. **C)** MVP alone is sufficient for the formation of vault structures. **D)** TEP1 is shared with the telomerase. **E, F)** Vault-independent fractions of PARP4 have been observed in the **E)** cytoplasm and **F)** nucleus. Figure adapted from (Tanaka et al., 2009; Zhang et al., 2015) and edited using Biorender.

1.5.5 PARP4 and its PARP family members

As a member of the poly (ADP-ribose) polymerase (PARP) family, which comprises 17 members involved in a diverse range of cellular processes, PARP4 possesses the PARP catalytic domain and is capable of ADP-ribosylation (Amé et al., 2004).

ADP-ribosylation is a reversible post-translational modification where nicotinamide adenine dinucleotide (NAD+) is used as a source of ADP-ribose groups for transfer onto target proteins via N-, O-, or S-glycosidic linkages, leaving nicotinamide as a by-product (Figure 1-4) (Bai, 2015; Cohen and Chang, 2018). Lysine, arginine, aspartate, glutamate, serine, tyrosine and cysteine residues have been identified as amino acid acceptors of ADP-ribose groups (Daniels et al., 2015; Leidecker et al., 2016; Leslie Pedrioli et al., 2018; Martello et al., 2016). More recently, DNA and RNA have also been reported as substrates of ADP-ribosylation (Munnur et al., 2019; Zarkovic et al., 2018).

ADP-ribose groups can be added onto target molecules either as single units (mono-ADP-ribosylation) or in chains of multiple units (poly-ADP-ribosylation), the latter of which can be linear or branched (Figure 1-4). These different modifications have distinct structural properties which in turn affect their recognition by other proteins (Gupte et al., 2017). Finally, ADP-ribosylation can be reversed by the family of ADP-ribosyl hydrolases (Rack et al., 2020).

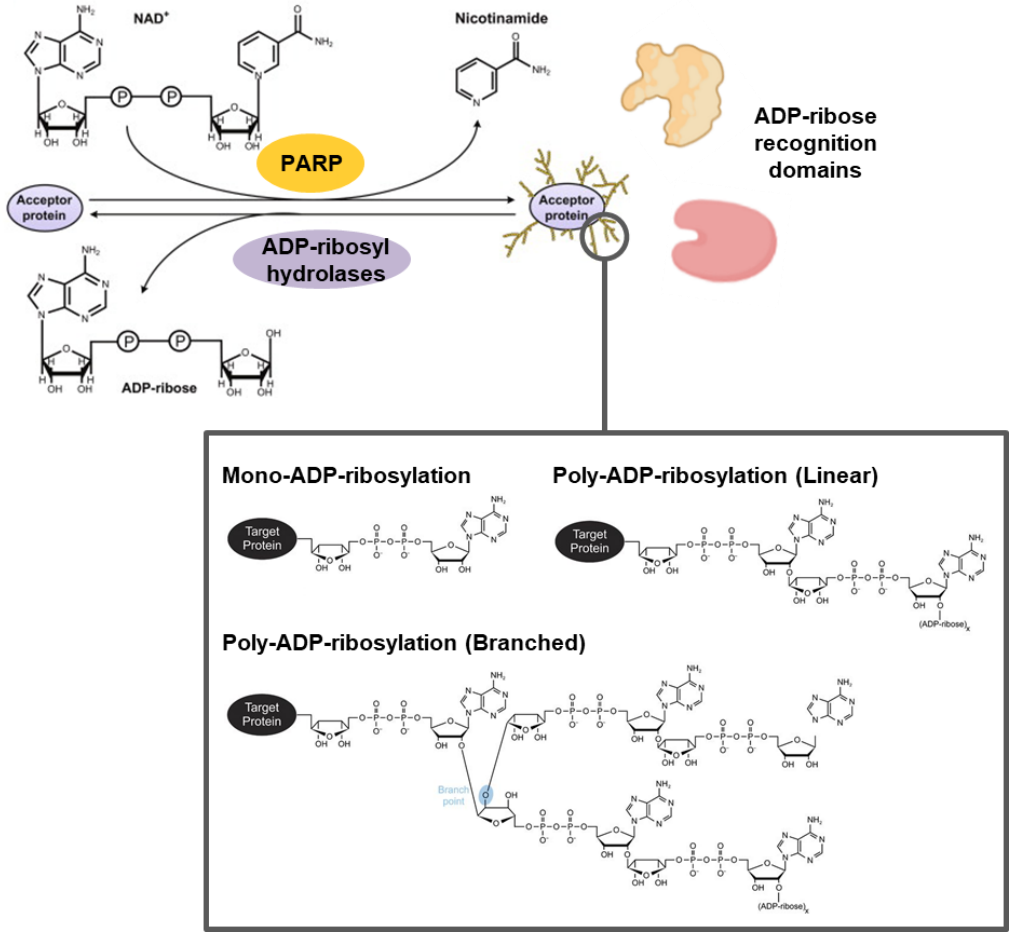


Figure 1-4: Schematic diagram of reversible ADP-ribosylation

PARP proteins use nicotinamide adenine dinucleotide (NAD+) as a source of ADP-ribose groups for transfer onto target proteins, whereas this process is reversed by ADP-ribosyl hydrolases. The ADP-ribose groups can be added singly or in chains, which are differentially recognized by proteins with ADP-ribose recognition domains. Figure adapted from (Feng and Koh, 2013; Leung, 2014) and edited using Biorender.

The type of ADP-ribosylation reaction that a PARP protein can mediate is thought to be determined by the key active site residues within the PARP domain. Specifically, a catalytic triad of histidine, tyrosine and glutamate (H-Y-E) found in PARP's 1-4 appears to be associated with poly-ADP-ribosylation (PARylation) (Scarpa et al., 2013).

In vitro radiolabeling experiments conducted on purified vault complexes revealed that PARP4 was indeed capable of PARylation, as radiolabeled poly-

ADP-ribose groups were detected on both PARP4 and MVP (Kickhoefer et al., 1999). However, in a more recent study involving immunoprecipitation of PARP4 followed by *in vitro* ADP-ribosylation, PARP4 was suggested to auto-modify itself via mono-ADP-ribosylation (MARylation) instead (Vyas et al., 2014). A possible account for these differing observations is that PARP4's PARylation or MARylation activity could be dependent on its cellular context. More studies are also needed to understand the functional relevance of PARP4's ADP-ribosylation activity.

ADP-ribosylation has been implicated in a number of processes relevant to cancer. For example, PARP1-mediated ADP-ribosylation is crucial in DNA repair. PARP1 is rapidly recruited to and activated by DNA damage sites, where it PARylates itself, histones and other DNA repair proteins to promote the recruitment of repair machinery to resolve the damage (Azarm and Smith, 2020; Ray Chaudhuri and Nussenzweig, 2017; Weaver and Yang, 2013). Apart from its well-known roles in DNA repair, PARP1 was also proposed to regulate centrosome function, as PARP1 was shown to PARylate centrosomal proteins while use of a PARP inhibitor led to centrosome hyperamplification (Kanai et al., 2003). In ovarian cancer cells, MARylation of the ribosome by PARP16 has been reported to regulate levels of protein synthesis to prevent the accumulation of protein aggregates that would otherwise lead to proteotoxic stress and reduced cancer cell viability (Challa et al., 2021).

Like PARP1, PARP4 possesses a BRCA1 C Terminus (BRCT) domain that is also present in several other DNA repair proteins, and was thus proposed to similarly mediate response to DNA damage (Figure 1-5, orange unit) (Bork et al., 1997; Callebaut and Mornon, 1997; Wu et al., 1996). However, UV-induced DNA damage did not alter the cellular distribution of PARP4 and the vault complex nor activate PARP4's ADP-ribosylation activity, suggesting that it might not be functionally involved in DNA repair (Kickhoefer et al., 1999a).

As the largest member of the PARP family, PARP4 has other protein domains that could conceivably lend it different functionalities (Figure 1-5, red box), although this would need to be further explored.

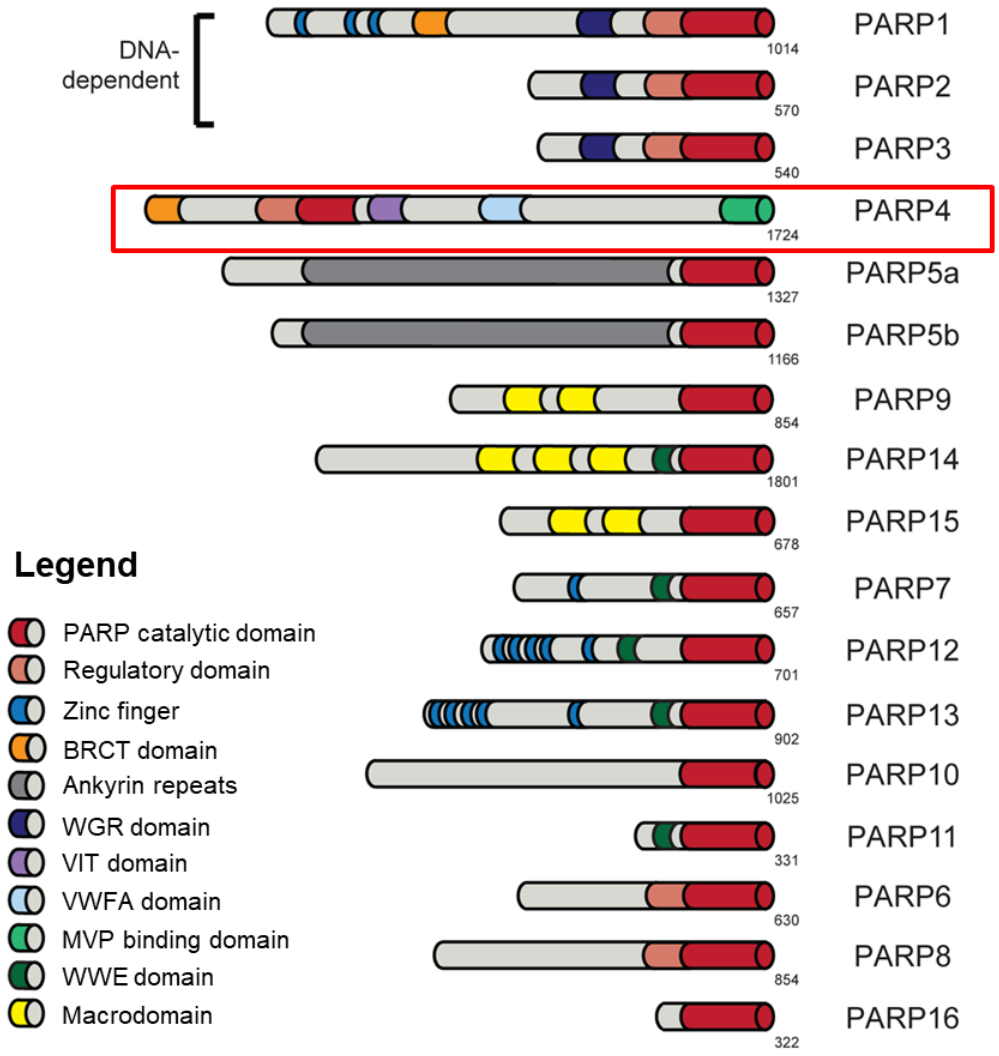


Figure 1-5: The PARP family members and their domain structures

Schematic diagram of the 17 PARP family members and their domain structures. The PARP catalytic domain in red is shared by all members. PARP4 (highlighted by the red box) is the largest member of the PARP family, with a number of unique domains not found in the other PARPs. These include the VIT (purple), VWFA (light blue) and MVP binding (light green) domains. Figure adapted from (Daugherty et al., 2014).

1.5.6 PARP4 and cancer

Mice in which PARP4 had been genetically ablated remained viable and fertile, indicating that PARP4 was probably dispensable in the normal context (Liu et al., 2004). However, when challenged with the chemical carcinogens dimethylhydrazine and urethane, PARP4-deficient mice were more likely than wildtype mice to develop colon and lung tumors, pointing to a potential tumor suppressive role (Raval-Fernandes et al., 2005).

In recent years, a few studies reported germline PARP4 mutations in cancer patients, and PARP4 was proposed as a candidate cancer susceptibility gene (Alimirzaie et al., 2018; Cirello et al., 2019; Ikeda et al., 2016; Prawira et al., 2019). Notably, in a cohort of 14 patients with independent thyroid and breast cancers who not only had a family history of cancer but were also null for germline mutations in the cancer susceptibility gene PTEN, six possessed germline PARP4 variants (T1170I, G496V) (Ikeda et al., 2016). PARP4 was also identified as a prospective biomarker for colorectal cancer metastasis, as cell lines that were highly aggressive were observed to have lower PARP4 expression than less metastatic cell lines (Long et al., 2016). However, most of these studies did not functionally validate PARP4's proposed effects on tumorigenicity, and a conclusive link between PARP4 and cancer has never been shown.

1.6 Rationale and key aims of PhD project

Given that targeted therapy against LUAD has been limited firstly by the lack of durability in patient response to existing TKIs due to treatment resistance and secondly by an incomplete identification of LUAD drivers to target, we propose to identify novel modulators of LUAD tumorigenicity that are relevant to the Asian setting.

Firstly, the role of novel candidate drivers identified from our Asian LUAD cohort will be examined. Specifically, by studying the mutational profiles of these candidate drivers and performing data mining, we will shortlist candidates to pursue. This will be followed by *in vitro* and *in vivo* tumorigenicity assays to validate the involvement of shortlisted candidate drivers in regulating LUAD tumorigenicity. The significance of key mutations identified within validated drivers will also be assessed by systematically generating the mutations and evaluating their outcomes. Subsequently, the mechanistic contribution of validated drivers to LUAD tumorigenesis will be characterized via broad-based profiling approaches including mass-spectrometry analysis of binding partners and transcriptomic analyses. Finally, drug screening will be performed to evaluate therapeutic opportunities in targeting lung cancer cells harbouring alterations in the candidate drivers.

The specific aims of this study are:

1. To functionally validate novel LUAD driver(s) discovered in Asian cohort data
2. To understand how mutations in putative driver genes, especially PARP4, drive or modulate LUAD tumorigenesis
3. To evaluate the downstream mechanisms of PARP4 in controlling tumorigenicity

The identification of novel LUAD modulators will open up opportunities for the discovery of new actionable mutations or biomarkers and targeted therapy to improve patient outcomes. This work will provide a better understanding of pathways deregulated in cancer, as well as the step-wise progression of tumorigenesis.

CHAPTER 2: Materials and Methods

2.1 Cell lines and culture conditions

A549, HEK293T and Platinum-A (Plat-A) cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS) (Gibco), 1% (v/v) Penicillin-Streptomycin (Gibco), 2mM L-glutamine (Gibco), and 1mM sodium pyruvate (Gibco). PC-9 cell line was maintained in Roswell Park Memorial Institute 1640 media (RPMI) (Gibco) supplemented with 10% (v/v) FBS (Gibco), 1% (v/v) Penicillin-Streptomycin (Gibco), 1% (v/v) Minimum Essential Medium non-essential amino acids (MEM NEAA), 2mM L-glutamine (Gibco) and 1mM sodium pyruvate (Gibco). SAEC and normal patient-derived lung cells were maintained on irradiated NIH-3T3 feeders using a proprietary media formulation. SAECTS and SAECK cells were maintained in the same media formulation. Cells were grown in a humidified incubator at 37^{o}C with 7.5% CO_2 and passaged when 80%-90% confluent using 0.25% trypsin-EDTA (Gibco) for PC-9 and 0.05% trypin-EDTA (Gibco) for all other cell lines.

2.2 Generation of SAECTS and SAECK

SAECTS cells were generated by the sequential transduction of SAEC cells cultured on NIH-3T3 feeders with retroviral supernatant generated from the following plasmid constructs: pBABE-puro SV40 LT (Addgene #13970) and pBABE-hygro-hTERT (Addgene #1773). Cells were selected with puromycin and hygromycin and subsequently taken off feeders.

SAECK cells were generated by the further transduction of SAECTS with retroviral supernatant generated from the pBABE puro K-Ras V12 (Addgene #9052) plasmid construct. The generation of viral supernatants is described in section 2.7. SAECTS and SAECK cells were generated by Dr. Yuan Ju.

2.3 Cell proliferation assay

500-1,000 cells were seeded in triplicate in a total volume of 100µL per well, on 96-well Greiner flat-bottomed white plates (Sigma-Aldrich). Cell viability was measured using the CellTiter-Glo Luminescent Cell Viability Assay (Promega) according to manufacturer's instructions at 24- or 48-hour intervals over a duration of 4 to 6 days.

2.4 Soft agar assay

Each well of a 12-well plate was first coated with a base layer comprising 1mL of 1% (w/v) low-melt agarose (Bio-Rad) dissolved in DMEM/F12 medium supplemented with 10% (v/v) FBS, 1% (v/v) Penicillin-Streptomycin, 2mM L-glutamine, and 1mM sodium pyruvate (soft agar medium). 2,500 cells were suspended in 1mL of 0.35% (w/v) low melt agarose dissolved in soft agar medium, carefully layered atop the base layer and allowed to solidify before the addition of 1mL of the respective cell line culture medium. After 3 weeks of growth, colonies were fixed and stained with crystal violet. The plates were imaged using a Gel Doc machine (Bio-Rad) and positively-stained colonies that were visible by eye were counted.

2.5 Tumor xenografts

All animal protocols and experiments were approved and carried out in accordance with the Agency for Science, Technology, and Research, Singapore – Biological Resource Centre (A*STAR – BRC) Institutional Animal Care and Use Committee (IACUC) guidelines (Protocol number 171286). Cells were resuspended in 100µL media containing 50% Matrigel Matrix (Corning) and injected subcutaneously into both flanks of NOD *scid* gamma (NSG) mice (The Jackson Laboratory) at either 1 million SAECK cells or 400,000 A549 cells per site. Mice were examined for tumor formation on a weekly basis, and palpable tumors were measured by vernier caliper. Tumor volume was determined as follows: ½ x tumor length x (tumor width)^2. Mice were euthanized when tumors reached 15mm in diameter.

2.6 Cloning

pLKO-EGFP was generated by replacing the puromycin cassette within pLKO.1 puro (Addgene #8453) with the EGFP cassette from LV-GFP (Addgene ##25999). Briefly, plasmids were digested using BamHI and KpnI restriction enzymes in NEB Cutsmart Buffer (New England Biolabs) at 37°C for 3 hours to release the puromycin and EGFP cassette respectively. The linearized pLKO.1 puro backbone and EGFP cassette from LV-GFP fragments were extracted from the agarose gel using Qiagen Gel Extraction Kit. The two fragments were ligated using Thermo Fisher Scientific Rapid Ligation Kit according to manufacturer's instructions to create pLKO-EGFP.

The shRNA sequences in Table 2-1 were subsequently cloned into pLKO-EGFP linearized by MluI and EcoRI (New England Biolabs).

Table 2-1: shRNA target sequences

Gene	shRNA Name	Target Sequence
Non-target control	shControl	CAACAAGATGAAGAGCACCAA
PARP4	shPARP4 #1	CCTGGGACTATTGGCTAAGAA
PARP4	shPARP4 #2	GCATTCAATCTCTAGGTGTAA
PARP4	shPARP4 #3	GCTCAGTACAAGTATCAAGTA
MVP	shMVP #1	CCCATCAACCTCTTCAACACA
MVP	shMVP #3	CACTTTCGATGACTTCCATAA
hnRNPM	shhnRNPM	CTGTGCAAGCTATATCTATGT
RASA1	shRASA1 #1	CCTGGCGATTATTCACTTTAT
RASA1	shRASA1 #2	CCCTACATGGAAGGTGTCAAT
RASA1	shRASA1 #3	CCTGACATCAATAGATTTGAA

The gRNA sequences in Table 2-2 were cloned into LentiCRISPRv2GFP (Addgene #82416) linearized by BsmBI (New England Biolabs).

Table 2-2: gRNA target sequences

Gene	gRNA Name	Target Sequence
Luciferase	gLuc	ACAACTTTACCGACCGCGCC
PARP4	gPARP4 #2	CTGGGTTTGCAATATGAACG
PARP4	gPARP4 #3	AGGAGGTGGTGTGATGTCCA
MVP	gMVP #1	CATGGATATAGTGGTATGGG
MVP	gMVP #2	GGCATCCCGAGACACAGGGT

PARP4 open reading frame (ORF) was purchased from Bio Basic Asia Pacific Pte Ltd and subcloned into pCDH-T2A-mCherry backbone (Addgene # 214632). PCR-based site-directed mutagenesis was performed using KAPA HiFi PCR Kit (Roche) according to manufacturer's instructions. Primer sequences used are in Table 2-3.

Table 2-3: Primer sequences for site-directed mutagenesis of PARP4 overexpression constructs

Gene	Forward Primer	Reverse Primer
BRCT domain del	CATGGTGTATAAGCCCCTGGACATCACACCA	GCTTATACACCATGGTGGCTTAATTAAGAATTCGCTAGCTCTAG
E547K	GAGGATGATAAGTTTGTTGTCTATAAAACCAATCAGG	CAACAAACTTATCATCCTCAAAGTCTGTGGTG
I1039T	AACAGACAGAAGACCAAATGACCAGGCTATGTTCTC	GGTCTTCTGTCTGTTTTCTCCAACTATGCTTGGATTTTGC
MVP domain del	ACACTGGCTCCATTACAGTCAAGGCGC	AATGGAGCCAGTGTTGTATGCACACTATTTCATCCTC
PARP domain del	ACCCAACTTTCATCCTAGTGATCATACTGAATTAGAGG	GATGAAAGTTGGGTTTGGACAAATTAGTTTCACAGAC
VIT domain del	AACTTCCGAACTCAGCATCCTGGGCA	TGAGTTCGGAAGTTTTGGCATCTGGTAACTGG
VWFA domain del	TGAGAGCTGTTCTCCGAGTTGCCACTC	GAGAACAGCTCTCACTGGCTAGGTCAGG

2.7 Transfection and generation of stable cell lines

HEK293T cells were transfected with the respective lentiviral constructs as well as the pCMV-delta R8.2 and pCMV-VSV-G packaging plasmids (Addgene) using the FuGENE 6 Transfection Reagent (Promega) for lentivirus production. Plat-A cells were transfected with the respective retroviral constructs alongside the pUMVC and pCMV-VSV-G packaging plasmids (Addgene) in the same manner. Viral supernatant was collected 48 hours and 72 hours post-transfection and filtered before use. Cells were transduced with the viral supernatant using polybrene (Sigma-Aldrich) at a final concentration of 8ng/µL.

2.8 Fluorescence-activated cell sorting (FACS)

Cells transduced with pLKO-EGFP shRNA, LentiCRISPRv2GFP gRNA or pCDH-PARP4-T2A-mCherry overexpression constructs were trypsinized and washed once with PBS before resuspension in a solution containing 2% FBS in PBS and DAPI at a final concentration 0.25µg/mL. The cell suspensions were filtered through a 35µm strainer into polypropylene tubes for analysis and sorting on the BD FACSAria^{TM} Fusion Cell Sorter (BD Biosciences) to isolate GFP+ or mCherry+ cells. FACS sorting was performed with the kind help of Ms. Salanne Lee and Ms. Ong Jia Min at the Singapore Immunology Network (SIgN) Flow Cytometry Platform.

2.9 Generation of clonal PARP4 knockout lines

SAECK cells were first transduced with gPARP4 #2 or gPARP4 #3. FACS sorting of single GFP+ cells into each well of several 96-well plates was performed. Single colonies were expanded, followed by protein extraction and immunoblotting (described in sections 2.11 and 2.13) to check for loss of PARP4. Genomic DNA was then extracted from successful clones using the DNeasy Blood and Tissue Kit (Qiagen) as per manufacturer's instructions. PCR amplification of PARP4 exon 3 was performed using the Phusion HF Master Mix (Thermo Fisher Scientific) according to manufacturer's instructions with the following primer pair:

PARP4 Exon 3 Forward Primer: 5’ ACCTTTGAACCTCCCTTTCCA 3’

PARP4 Exon 3 Reverse Primer: 5’ ACCCAAGCGTGTACTATGGC 3’

PCR products were sent for Sanger sequencing to check for homozygous frameshift mutations at the gRNA target site.

2.10 RNA Isolation and Quantitation via Real Time Quantitative Polymerase Chain Reaction (RT-qPCR)

RNA extraction was performed using the phenol-chloroform method with TRIzol® Reagent (Thermo Fisher Scientific) and further purified using the RNeasy Mini Kit (Qiagen). 500µg of total RNA was then converted to cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) following manufacturer’s instructions. The cDNA was diluted five-fold for use in quantitative real-time PCR (qPCR) analysis. Specifically, the BlitzAmp Hotstart qPCR Master Mix (MiRXES) was used according to manufacturer’s instructions. Transcript levels were determined using gene-specific primers (Table 2-4) and reactions were carried out using the QuantStudio^{TM} 5 Real-Time PCR System (Applied Biosystems). Samples were assayed in triplicate and GAPDH was used as the internal normalization control. Relative gene expression was determined using the ΔΔCT method.

Table 2-4: Primer sequences for RT-qPCR analysis

Gene	Forward Primer	Reverse Primer
GAPDH	TGGCAAATTCCATGGCACCG	CGCCCCACTTGATTTTGGAGG
hnRNPM	GAGGCCATGCTCCTGGG	TTTAGCATCTTCCATGTGAAATCG
KRAS	TGTACTGGTCCCTCATTGCA	CCCTCATTGCACTGTACTCC
MVP	AACTCCCAAGCCCCACCC	GGGGAGCATCTAGAAGTGCAG
PARP4	CAGACGTCGGAAACCTTGGA	AACTCCATGCACACTGTCGT

RASA1	GTTTATGATGGGAGGCCGGT	TCCTTGCCATCCACTGTGTC
SV40LT	CTGACTTTGGAGGCTTCTGG	GGAAAGTCCTTGGGGTCTTC
TERT	TGTTTCTGGATTTGCAGGTG	GTTCTTGGCTTTCAGGATGG

2.11 Protein Extraction

Cells were first rinsed with PBS to remove residual media and then lysed in ice-cold RIPA Lysis and Extraction Buffer (Thermo Fisher Scientific) supplemented with 1X Halt™ Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific). Samples were incubated on ice and vortexed every 10 minutes for a total of 30 minutes. Following this, sonication was performed at 5 cycles of 30s ON, 30s OFF at high setting on the Bioruptor Plus (Diagenode). Samples were then centrifuged at 21,000g for 20 minutes at 4°C to remove cell debris. The supernatant was then collected as the lysates. Protein concentrations of the lysates were determined via Bradford Assay using the Coomassie Plus™ Protein Assay Reagent (Thermo Fisher Scientific), with sample absorbance values measured at 595nm.

2.12 Cell Fractionation

Cell fractionation was performed using the PARIS™ kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Briefly, cells were first rinsed with PBS to remove residual media and then harvested by scraping in PBS while on ice. Cells were pelleted by centrifugation at 250g for five minutes at 4°C, and gently resuspended in ice-cold Cell Fractionation Buffer. After 10 minutes of incubation on ice, the samples were centrifuged at 500g for five minutes at 4°C. The supernatant was collected as the cytoplasmic fraction while the nuclear pellet was washed twice with Cell Fractionation Buffer to minimize contamination with the cytoplasmic fraction. The cytoplasmic fraction was further clarified by centrifugation at 21,000g for 20 minutes at 4°C. The nuclear pellet was lysed in ice-cold Cell Disruption Buffer on ice and vortexed every 10 minutes for a total of 30 minutes. Following this, the samples were centrifuged

at 21,000g for 20 minutes at 4°C. The supernatant was collected as the nuclear fraction. Protein concentrations of both fractions were determined via Bradford Assay using the Coomassie Plus™ Protein Assay Reagent (Thermo Fisher Scientific), with sample absorbance values measured at 595nm.

2.13 Immunoblotting

30-50µg of total protein was mixed with NuPAGE™ LDS Sample Buffer and NuPAGE™ Sample Reducing Agent (Thermo Fisher Scientific) before incubation at 70°C for 10 minutes. Protein samples were resolved by electrophoresis using 4-12% Bis-Tris precast gels (Thermo Fisher Scientific) and transferred onto PVDF membranes using the Trans-Blot® Turbo™ Transfer System (Bio-Rad) according to manufacturer's instructions. Membranes were blocked in 5% (w/v) Blotting Grade Blocker Non-Fat Dry Milk (Bio-Rad) or 5% (w/v) bovine serum albumin (BSA) (Hyclone) in Tris-buffered saline containing 0.1% Tween-20 (TBS-T) for an hour at room temperature. Membranes were incubated in primary antibody at 4°C overnight and incubated in secondary antibody at room temperature for an hour. Protein bands were visualized via chemiluminescence using the SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific). Images were captured using the ChemiDoc MP Imaging System (Bio-Rad).

Table 2-5: Details of primary and secondary antibodies

Primary Antibody	Company	Catalog Number
Anti-B actin	Santa Cruz Biotechnology	sc-47778
Anti-GAPDH	Santa Cruz Biotechnology	sc-47724
Anti-H3	Abcam	ab1791
Anti-hnRNPM	Santa Cruz Biotechnology	sc-20002
Anti-KRAS	Santa Cruz Biotechnology	sc-30
Anti-KRT18	Santa Cruz Biotechnology	sc-6259
Anti-mono(ADP-ribose)	Merck	MABE1076
Anti-MVP	Santa Cruz Biotechnology	sc-23916
Anti-PARP4	Santa Cruz Biotechnology	sc-515898
Anti-PCM1	Santa Cruz Biotechnology	sc-398365
Anti-PCNT	Santa Cruz Biotechnology	sc-376111

Anti-PHB	Santa Cruz Biotechnology	sc-377037
Anti-PHB2	Santa Cruz Biotechnology	sc-133094
Anti-poly(ADP-ribose)	Santa Cruz Biotechnology	sc-56198
Anti-RASA1	Santa Cruz Biotechnology	sc-63
Anti-SV40 T Antigen	Santa Cruz Biotechnology	sc-55461
Anti-TERT	Santa Cruz Biotechnology	sc-393013
Anti-Ubiquitin	Cell Signaling	CST3933S
Anti-VIM	Santa Cruz Biotechnology	sc-32322
Secondary Antibody	**Company**	**Catalog Number**
Anti-mouse IgG, HRP-linked	Cell Signaling	CST7076S
Anti-rabbit IgG, HRP-linked	Cell Signaling	CST7074S
Streptavidin, HRP-linked	Abcam	ab7403

2.14 Immunoprecipitation

Cells were first rinsed with PBS to remove residual media and then lysed in ice-cold immunoprecipitation lysis buffer containing 10mM Tris-HCl (pH 7.5), 150mM NaCl, 0.5% (v/v) NP-40, 0.25% (w/v) sodium deoxycholate, 0.5mM EDTA (pH 8.0) (co-IP buffer) supplemented with 1X Halt™ Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific), 2mM $MgCl_2$, and 50 units/mL benzonase (Sigma-Aldrich). Samples were incubated on rotation at 4°C for 1 hour and then centrifuged at 21,000g for 20 minutes at 4°C to remove cell debris. The supernatant was then collected as the lysates. Protein concentrations of the lysates were determined via Bradford Assay using the Coomassie Plus™ Protein Assay Reagent (Thermo Fisher Scientific), with sample absorbance values measured at 595nm.

1mg of total protein from co-IP lysates, cytoplasmic or nuclear fractions were pre-cleared for 60 minutes using 50µl of ChIP-grade Protein A/G Dynabeads (Thermo Fisher Scientific) before overnight incubation with 4µg of the immunoprecipitating antibody at 4°C on rotation. 50µl of fresh Protein A/G Dynabeads were similarly blocked overnight in 1% (v/v) BSA (Hyclone) in co-IP buffer. The antibody-lysate mixtures were incubated with blocked Dynabeads for 90 minutes. The unbound fraction was kept while the beads were washed 3

times with co-IP buffer and bound proteins were eluted in 2x Laemlli buffer (Bio-Rad) by incubating at 95°C for 5 minutes. The various fractions were subsequently analyzed by immunoblotting as described above.

Table 2-6: Antibodies used for immunoprecipitation (IP)

IP Antibody	Company	Catalog Number
Anti-hnRNPM	Santa Cruz Biotechnology	sc-20002
Anti-KRT18	Santa Cruz Biotechnology	sc-6259
Anti-MVP	Santa Cruz Biotechnology	sc-23916
Anti-PARP4	Santa Cruz Biotechnology	sc-515898
Anti-PCM1	Santa Cruz Biotechnology	sc-398365
Anti-PCNT	Santa Cruz Biotechnology	sc-376111
Anti-PHB	Santa Cruz Biotechnology	sc-377037
Anti-PHB2	Santa Cruz Biotechnology	sc-133094
Anti-VIM	Santa Cruz Biotechnology	sc-32322
Mouse IgG1 Isotype Control	Cell Signaling	5415S

2.15 Stable Isotope Labelling by Amino Acids in Cell Culture (SILAC)

DMEM lacking L-arginine and L-lysine was supplemented with 10% dialysed FBS (Gibco), and either with [13C6 15N4]- L-arginine and [13C6 15N2]-L-lysine (Sigma-Aldrich) to form the “heavy” labelled SILAC medium, or with non-labelled L-arginine and L-lysine to form the “light” unlabelled SILAC medium. Cells were grown in the respective “heavy” and “light” medium for at least eight passages for complete labelling before they were used for co-immunoprecipitation experiments.

Thereafter, samples were submitted to the Cancer Science Institute (CSI) Quantitative Proteomics Core where further sample processing, mass spectrometry and data analysis was kindly performed by Dr. Dennis Kappei and Ms. Charlene Chan.

2.16 Nanopore sequencing

Total RNA of respective cell lines was extracted in duplicate using the RNeasy Mini Kit (Qiagen) as described in section 2.10, with an additional on-column DNA digestion step performed according to manufacturer's instructions. RNA quality was assessed using the Agilent 2100 Bioanalyzer Instrument (Agilent). RNA samples were subsequently sent to the Next Generation Sequencing Platform at the Genome Institute of Singapore (GIS-NGSP), where mRNA enrichment and cDNA synthesis was performed. The cDNA samples were then sequenced using the GridION device from Oxford Nanopore Technologies, with one sample per MinION flow cell. Samples were sequenced in high accuracy mode with a total run-time of 72h.

2.17 PCR validation of splice events

PCR amplification of cDNA samples was performed using the Phusion HF Master Mix (Thermo Fisher Scientific) using the primers in Table 2-7. PCR products were then separated by gel electrophoresis. The resultant bands were imaged on the Gel Doc system (Bio-Rad) and quantitated using the Image Lab software (Bio-Rad).

Table 2-7: Primer sequences for splicing validation

Gene	Forward Primer	Reverse Primer
ASB1	GGCTGGATGAACGTGTGAC	TCAAACTGGAAACCACGAGC
FIP1L1	TGCCCTTCCATCTACAAAAGC	CTGTTCTCATTTGCCCGTCG
MIF4GD	TCTCTGCAGGACTGTGTGTT	TCCTCCTTGCTCAAACTGTCT
PRPF4B	ACGAATTAGCATCAACCAGGC	ACTGACTACACCAAGGGCAT
SEC24C	CGAAGTCTAACCTGGATCTGG	CTGGGTAGATGGGCTGGG
SLC19A1	GGTTCCTCTCTCCCACCCTA	CCTGAGATCCGGCAACATCA
TMEM107	GTGGCCCTGTCCTTCTTCAT	CGTCCTTAGGTTCCCGTCAT

2.18 High throughput screening of metabolic and anti-cancer small molecule compound libraries

The anti-cancer drug library (SelleckChem) comprised a total of 317 drug compounds whereas the metabolic library was assembled from 303 different agonists and inhibitors targeting major metabolic pathways (Med Chem Express).

Stock plates of these compound libraries were prepared by the Centre for High Throughput Phenomics, previously at the Genome Institute of Singapore (CHiP-GIS). The compounds were reconstituted in DMSO and plated into a 384-well format.

Cells were seeded at a density of 250 cells per well onto Corning™ 384-well flat-bottomed white plates using the MultiFlo™ FX multi-mode dispenser (BioTek) and allowed to attach to the plate overnight. Drugs from the respective compound libraries were then added to a final concentration of 1µM, with each well having a total volume of 50µL, with the kind help of Mr. Matan Thangavelu Thangavelu (CHiP-GIS). 72 hours later, cell viability was measured using the CellTiter-Glo Luminescent Cell Viability Assay (Promega) and normalized to the DMSO control.

2.19 Statistical Analyses

Statistical analyses were performed using GraphPad Prism. Two-tailed unpaired student's t-test was performed to evaluate the significance of differences between two sample groups. For comparisons between multiple groups, ordinary one-way ANOVA followed by either the Tukey test to correct for multiple comparisons between groups, or the Dunnett test to correct for multiple comparisons to a control group, was performed. The multiplicity adjusted p-value was then reported. A p-value of less than 0.05 was defined to be statistically significant. Significance values are indicated as follows – ns: not significant $p \geq 0.05$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ and ****$p < 0.0001$.

CHAPTER 3: Results

3.1 Selection of novel candidate drivers for validation and investigation

The driver analysis of WES data from our Asian LUAD cohort had earlier revealed seven candidate driver genes novel to LUAD (Chen et al., 2020b). In order of decreasing mutational frequency, the seven genes are PARP4 (6%), EPRS (4%), LYST (4%), NCOR2 (2%), ZMYM2 (2%), RASA1 (2%) and PBRM1 (2%) (Figure 1-2).

I began by studying the profiles of mutations occurring within these genes to determine which of them I should prioritize. Specifically, the presence of recurrent mutations could indicate a hotspot mutation with likely functional importance, whereas the localized clustering of mutations within key protein domains could suggest perturbation of protein functions (Chen et al., 2016; Kamburov et al., 2015; Miller et al., 2015).

Previously, Dr. Jianbin Chen had generated lollipop plots representing the mutational profiles of each of the 27 candidate drivers and these were published in (Chen et al., 2020b). Here, I have reproduced the lollipop plots of the three most frequently mutated drivers in the Asian LUAD cohort – EGFR, TP53 and KRAS (Figure 3-1A-C), as well as those of the seven candidate drivers (Figure 3-1D-J). For each driver, the lollipop plots in Figure 3-1 describe the distribution of individual mutations along the protein sequence, as well as their identity and frequency.

Matching what is already known, most EGFR mutations reported in the cohort localize to the kinase domain of EGFR (Figure 3-1A), with the exon 21 L858R mutation being of highest prevalence, followed by deletions in exon 19 (Gazdar, 2009; Sharma et al., 2007; Siegelin and Borczuk, 2014). In the case of TP53, most mutations are clustered within its core DNA-binding domain (Figure 3-1B), consistent with previous reports of close to 80% of TP53 mutations across cancers occurring within this region (Hainaut and Pfeifer, 2016; Olivier et al., 2010). As for KRAS (Figure 3-1C), all mutations occur at the G12, G13 and Q61

loci, all of which localize to KRAS' nucleotide-binding pocket and are known oncogenic mutations (Haigis, 2017; Prior et al., 2012).

In the case of the seven candidate drivers, most of them have mutations which occur singly and are scattered throughout the protein (Figure 3-1F-J). However, the mutational profiles of two genes in particular caught my attention.

The first of these, PARP4, is the only candidate driver that has a recurrent mutation (Figure 3-1D). Notably, more than a third of the 17 PARP4-mutant cases shared this I1039T mutation, which is located within PARP4's von Willebrand factor type A (VWFA) domain.

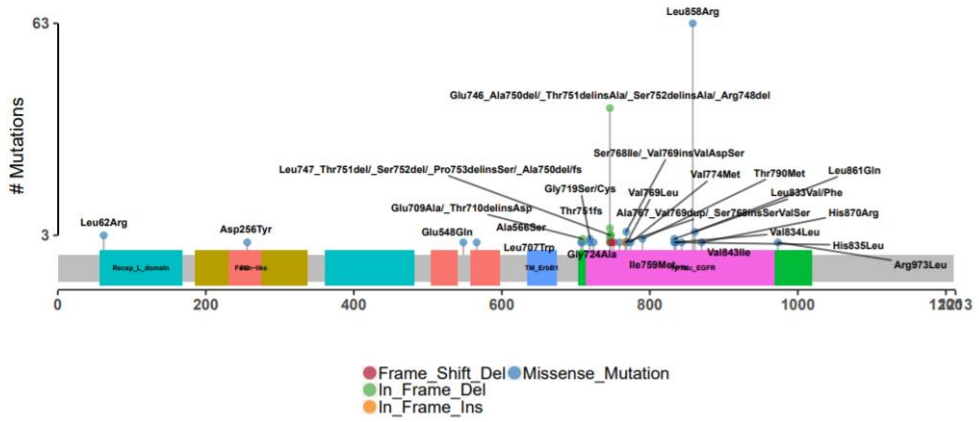
Furthermore, the recurrent I1039T mutation, the function of which remains unknown, had previously been reported at 2.35% frequency in a predominantly squamous Korean lung cancer cohort (n=170) as well as in other cancers (Table 3-1). Thus, it could be functionally important.

Interestingly, PARP4 mutations are associated with a Chinese ethnic bias when compared to a TCGA dataset comprising 230 Caucasian LUAD patients (Chen et al., 2020b; The Cancer Genome Atlas Research Network, 2014) (Figure 3-2). This could suggest why PARP4's putative role as a modulator of LUAD tumorigenesis was never previously picked up from Caucasian cohort data, highlighting the value of this study in identifying novel LUAD modulators.

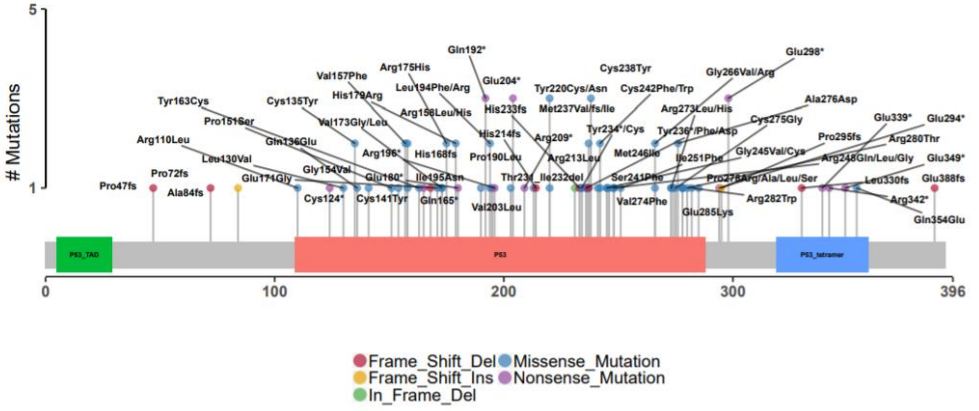
The second candidate, RASA1, was not as frequently mutated compared to some of the other candidates. However, all five mutations localized to the same Ras GTPase-activating (RasGAP) domain (Figure 3-1E). Furthermore, three of these are truncating mutations, with one instance of a frameshift deletion and two instances of nonsense mutations. Thus, these various mutations could share a common outcome of disrupting or altering RASA1's RasGAP function.

Hence, I decided to conduct further analyses on these two genes.

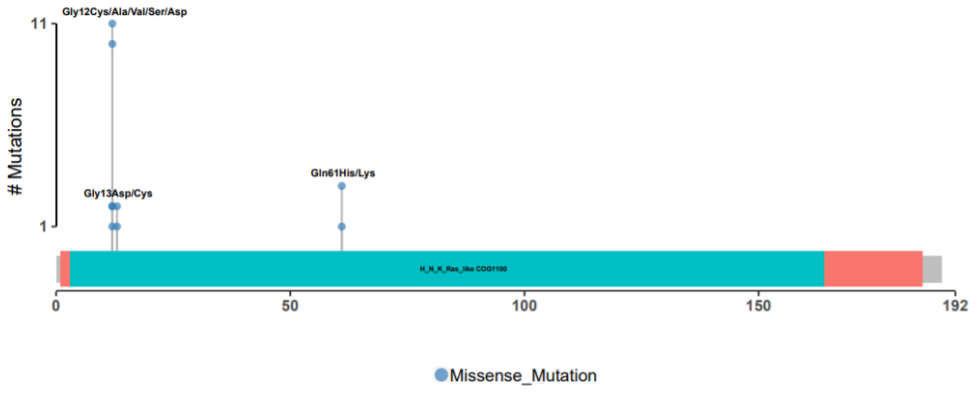
A)


EGFR: [Somatic Mutation Rate: 56.4%]
NM_005228
Mutations
63
3
Leu858Arg
Glu746_Ala750del/_Thr751delinsAla/_Ser752delinsAla/_Arg748del
Ser768Ile/_Val769insValAspSer
Leu747_Thr751del/_Ser752del/_Pro753delinsSer/_Ala750del/fs
Val774Met
Thr790Met
Leu861Gln
Gly719Ser/Cys
Val769Leu
Leu833Val/Phe
Glu709Ala/_Thr710delinsAsp
Thr751fs
His870Arg
Leu62Arg
Asp256Tyr
Ala566Ser
Val834Leu
Glu548Gln
Leu707Trp
Gly724Ala
Val843Ile
His835Leu
Ile759Met
Arg973Leu
0
200
400
600
800
1000
Frame_Shift_Del
Missense_Mutation
In_Frame_Del
In_Frame_Ins

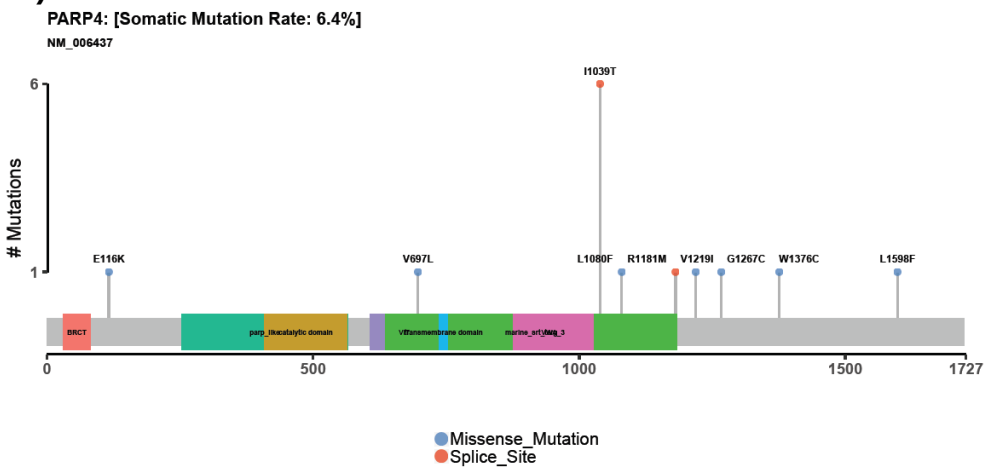
B)


TP53: [Somatic Mutation Rate: 43.2%]
NM_000546
Mutations
5
1
Gln192*
Glu298*
Arg175His
Cys238Tyr
Val157Fhe
Gly266Val/Arg
Glu204*
Leu194Phe/Arg
Tyr220Cys/Asn
Cys242Phe/Trp
Tyr163Cys
Cys135Tyr
His179Arg
Ala276Asp
His233fs
Met237Val/fs/Ile
Arg273Leu/His
Pro151Ser
Val173Gly/Leu
Arg156Leu/His
Cys275Gly
Glu339*
Glu294*
Arg110Leu
Gln136Glu
His214fs
Arg209*
Tyr234*/Cys
Tyr236*/Phe/Asp
Pro295fs
Arg280Thr
Leu130Val
Arg196*
His168fs
Pro190Leu
Arg213Leu
Met246Ile
Ile251Phe
Gly245Val/Cys
Arg248Gln/Leu/Gly
Glu349*
Pro72fs
Gly154Val
Glu171Gly
Glu180*
Ile195Asn
Thr23_Ile232del
Ser241Phe
Pro278Arg/Ala/Leu/Ser
Leu330fs
Glu388fs
Pro47fs
Ala84fs
Cys124*
Cys141Tyr
Gln165*
Val203Leu
Val274Phe
Glu285Lys
Arg282Trp
Arg342*
Gln354Glu
P53_TAD
P53
P53_tetramer
0
100
200
300
396
Frame_Shift_Del
Missense_Mutation
Frame_Shift_Ins
Nonsense_Mutation
In_Frame_Del

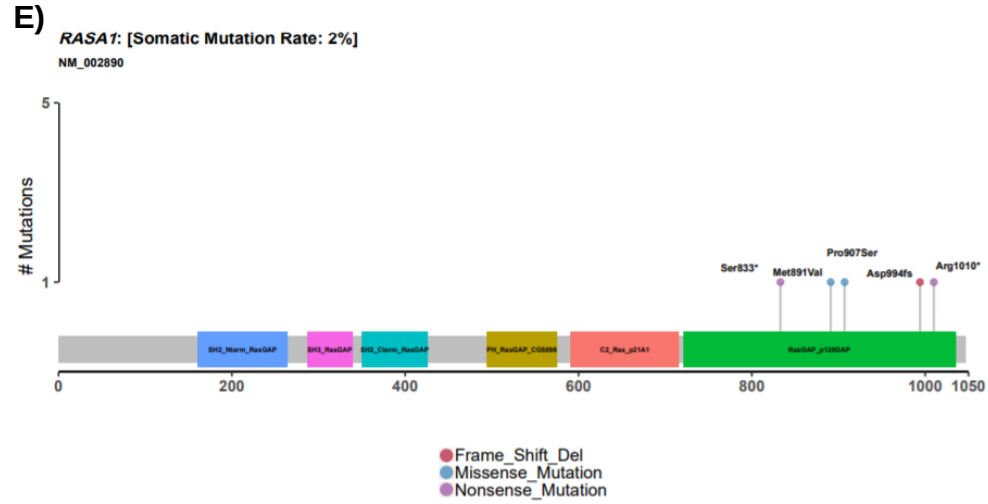
C)


KRAS: [Somatic Mutation Rate: 13.2%]
NM_033360
Mutations
11
1
Gly12Cys/Ala/Val/Ser/Asp
Gln61His/Lys
Gly13Asp/Cys
H_N_K_Ras_like COG1100
0
50
100
150
192
Missense_Mutation

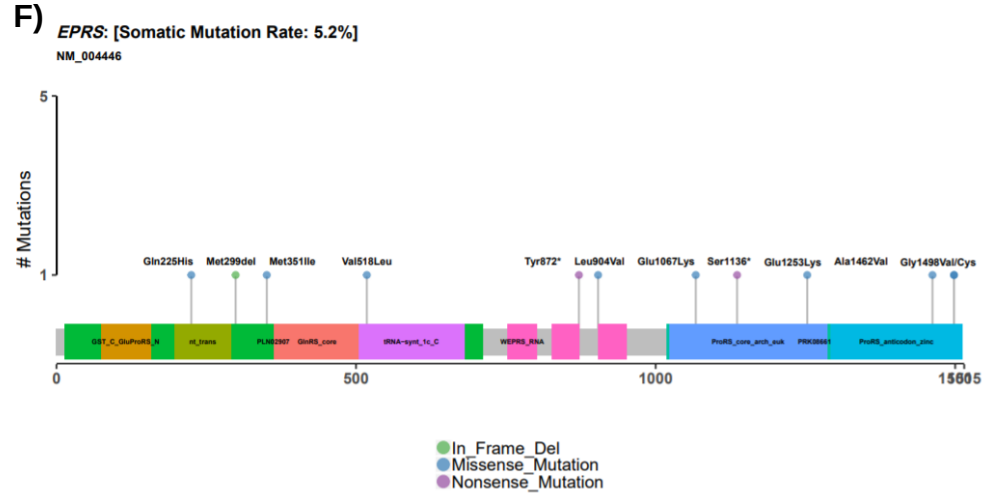
D)


PARP4: [Somatic Mutation Rate: 6.4%]
NM_006437
Mutations
6
1
I1039T
E116K
V697L
L1080F
R1181M
V1219I
G1267C
W1376C
L1598F
BRCT
parp_like catalytic domain
Transmembrane domain
0
500
1000
1500
1727
Missense_Mutation
Splice_Site

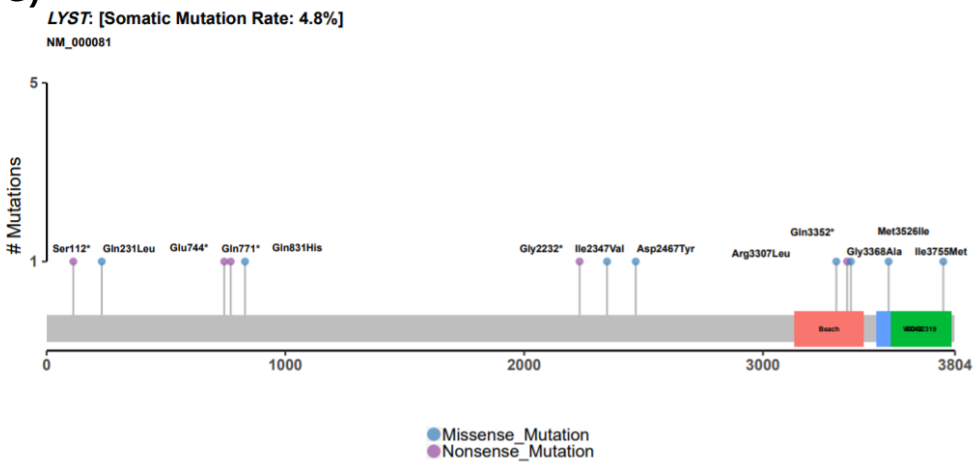
E)


RASA1: [Somatic Mutation Rate: 2%]
NM_002890
Mutations
5
1
Pro907Ser
Ser833*
Met891Val
Asp994fs
Arg1010*
SH2_Nterm_RasGAP
SH3_RasGAP
SH2_Cterm_RasGAP
PH_RasGAP_C
C2_Ras_p21A1
RasGAP_p120GAP
0
200
400
600
800
1000
1050
Frame_Shift_Del
Missense_Mutation
Nonsense_Mutation

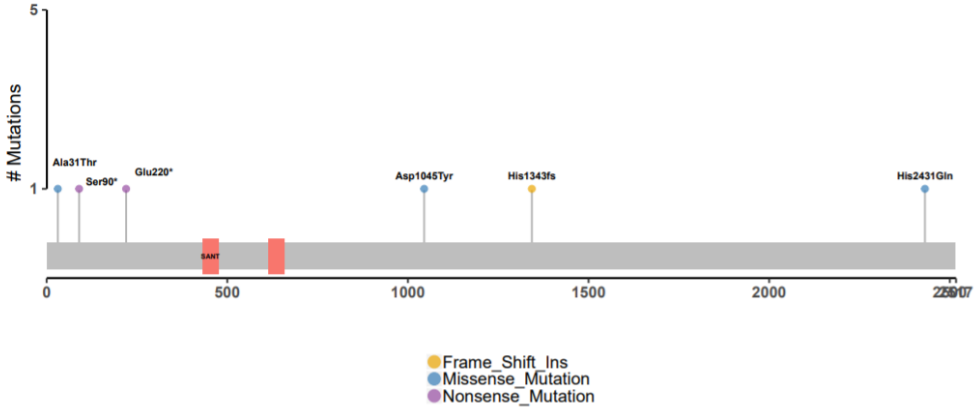
F)


EPRS: [Somatic Mutation Rate: 5.2%]
NM_004446
Mutations
5
1
Gln225His
Met299del
Met351Ile
Val518Leu
Tyr872*
Leu904Val
Glu1067Lys
Ser1136*
Glu1253Lys
Ala1462Val
Gly1498Val/Cys
GST_C_GluProRS_N
nt_trans
PLN02907
GlnRS_core
tRNA-synt_1c_C
WEPRS_RNA
ProRS_core_arch_euk
PRK08661
ProRS_anticodon_zinc
0
500
1000
1500
In_Frame_Del
Missense_Mutation
Nonsense_Mutation

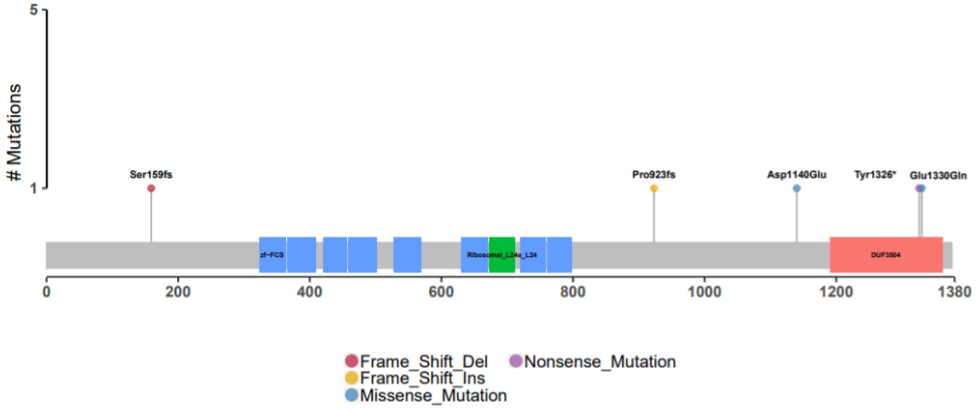
G)


LYST: [Somatic Mutation Rate: 4.8%]
NM_000081
Mutations
5
1
Ser112*
Gln231Leu
Glu744*
Gln771*
Gln831His
Gly2232*
Ile2347Val
Asp2467Tyr
Arg3307Leu
Gln3352*
Gly3368Ala
Met3526Ile
Ile3755Met
Beach
0
1000
2000
3000
3804
Missense_Mutation
Nonsense_Mutation

H)


NCOR2: [Somatic Mutation Rate: 2.4%]
NM_006312
Mutations
5
1
Ala31Thr
Ser90*
Glu220*
Asp1045Tyr
His1343fs
His2431Gln
SANT
0
500
1000
1500
2000
Frame_Shift_Ins
Missense_Mutation
Nonsense_Mutation

I)


ZMYM2: [Somatic Mutation Rate: 2%]
NM_003453
Mutations
5
1
Ser159fs
Pro923fs
Asp1140Glu
Tyr1326*
Glu1330Gln
zf-FCS
Ribosomal_L24e_L24
DUF3504
0
200
400
600
800
1000
1200
1380
Frame_Shift_Del
Frame_Shift_Ins
Missense_Mutation
Nonsense_Mutation

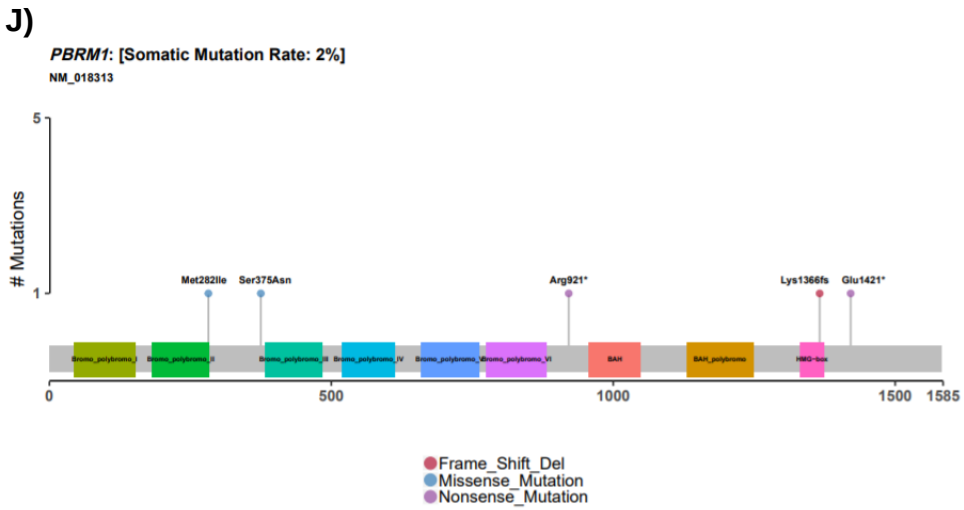


Figure 3-1: Mutational profiles of EGFR, TP53, KRAS and seven candidate drivers

A-J) Lollipop plots reflecting the distribution of individual mutations reported in the Asian LUAD cohort for **A)** EGFR, **B)** TP53, **C)** KRAS, **D)** PARP4, **E)** RASA1, **F)** EPRS, **G)** LYST, **H)** NCOR2, **I)** ZMYM2 and **J)** PBRM1. The X-axis represents the protein sequence and the position of each line along the X-axis reflects the amino acid residue affected by the indicated mutation. The height of the line describes how frequently the mutation was observed, whereas the color of the circle indicates the nature of the mutation. Analysis performed by Dr. Jianbin Chen. Figure reproduced from (Chen et al., 2020b).

Table 3-1: Distribution of PARP4 I1039T mutation across different cancer cohorts

Data retrieved from the International Cancer Genome Consortium (ICGC) Data Portal (Zhang et al., 2019a)

Project	Tumor Subtype	No. of PARP4 I1039T cases	Cohort size	Frequency of PARP4 I1039T (%)
Colorectal Cancer (China)	Adenocarcinoma, non-Western	9	321	2.80
Lung Cancer (Korea)	Adenocarcinoma, Squamous cell carcinoma	4	170	2.35
Skin Cancer (Australia)	Melanoma	3	183	1.64
Acute Myeloid Leukemia (Korea)	Acute myeloid leukemia	2	205	0.98
Gastric Cancer (China)	Intestinal- and diffuse-type	1	123	0.81
Liver Cancer (China)	Hepatocellular carcinoma HBV-associated	3	402	0.75
Cervical Squamous Cell Carcinoma (TCGA, US)	Cervical squamous cell carcinoma	1	194	0.52
Esophageal Cancer (China)	Squamous carcinoma	1	332	0.30
Kidney Renal Clear Cell Carcinoma (TCGA, US)	Clear cell carcinoma	1	408	0.25

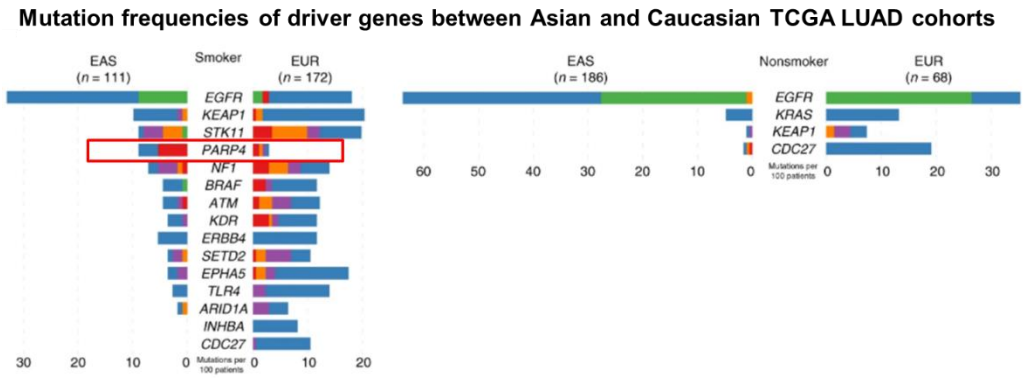


Figure 3-2: PARP4 mutations are associated with a Chinese ethnic bias

Comparison of driver mutational frequencies between the Asian (EAS) and Caucasian (EUR) TCGA cohorts among smokers (left) and non-smokers (right), reveals significant ethnicity differences. All genes displayed in the figure have significant differences (FDR < 0.1) as calculated by two-sided Fisher's exact test. PARP4 is highlighted in the red box. Analysis performed by Dr. Jianbin Chen. Figure reproduced from (Chen et al., 2020b).

3.2 Lower PARP4 or RASA1 expression is correlated with poorer overall survival in LUAD

Having identified PARP4 and RASA1 as interesting candidate drivers to follow up on, I first sought to investigate whether they might play a tumor suppressive or oncogenic role in the context of LUAD. To do so, I examined our Asian LUAD dataset (Chen et al., 2020b), the Caucasian TCGA LUAD dataset (The Cancer Genome Atlas Research Network, 2014), as well as a racially mixed collection of microarray datasets from the KM Plotter database (Győrffy et al., 2013).

3.2.1 PARP4

I first determined the effect of PARP4 expression on LUAD patient overall survival, as one would expect lower expression of a tumor suppressor and higher expression of an oncogene to be associated with poorer prognosis. Using microarray data from the KM Plotter database, LUAD patients (n=720) were segregated into a low and high expression group, with overall survival then compared between the two groups (Győrffy et al., 2013). Consistent with a

potential tumor suppressive role, lower levels of PARP4 were correlated with poorer overall survival (Figure 3-3).

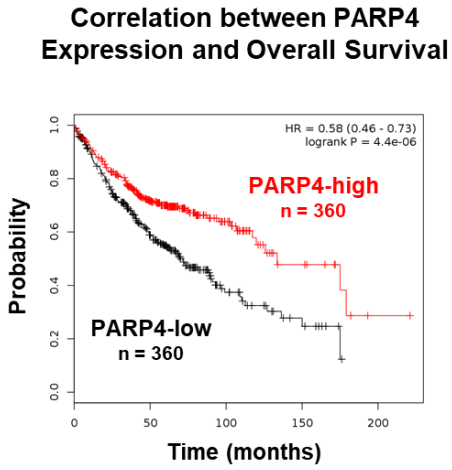


Figure 3-3: Lower PARP4 expression in LUAD is correlated with poorer overall survival

Kaplan-Meier plot generated using LUAD microarray data (n=720, Affymetrix ID 202239_at for PARP4) from the KM Plotter database (Győrffy et al., 2013). The patient group with high gene expression (n=360) is displayed in red while the low expression group (n=360) is marked in black.

Although PARP4 somatic mutations were reported in only 6% of patients within the Asian LUAD cohort, I wanted to determine whether there could be other patients affected by copy number alterations. Having observed PARP4 expression to be inversely correlated with survival (Figure 3-3), I hypothesized that alterations in PARP4 copy number could be another means of perturbing PARP4 gene expression levels, thereby giving rise to phenotypic differences even in the absence of somatic mutations.

In fact, the relationship between copy number and gene expression has been previously examined. In one study where molecular subtyping and clustering analyses integrating both copy number and gene expression data were performed on multiple independent cancer cohorts, it was revealed that certain gene expression patterns may not be driven by copy number (Ramazzotti et al., 2018). However, several studies have also reported a concordance between copy

number status and relative gene expression levels (Graham et al., 2017; Ohashi et al., 2019; Paolella et al., 2017; Zhao and Zhao, 2016). I thus sought to examine this within the Asian LUAD cohort as well as the Caucasian TCGA LUAD cohort.

Details of copy number alterations within the Asian and TCGA LUAD cohorts were obtained from the OncoSG or cBio portals (Cerami et al., 2012; Gao et al., 2013) respectively, where samples had been stratified into thresholded copy number status using the GISTIC/GISTIC2.0 algorithm (Beroukhim et al., 2007; Mermel et al., 2011) as in Table 3-2.

Table 3-2: Summary of copy number status as described by GISTIC/GISTIC2.0 (Beroukhim et al., 2007; Cerami et al., 2012; Mermel et al., 2011)

Effect on copy number	Copy number status	Thresholded copy number	Functional implication
Loss	Deep deletion	-2	Possible homozygous deletion
	Shallow deletion	-1	Possible heterozygous deletion
Neutral	Diploid	0	Diploid
Gain	Gain	1	Low-level gain
	Amplification	2	High-level gain of more copies

More than half (51.7%, n=156) of all cases within the Asian LUAD cohort (n=302) had PARP4 copy number loss, with the vast majority of these having shallow deletions (50%, n=151). In contrast, gain in PARP4 copy number was only observed in 8.3% of patients (gain: n=24, amplification: n=1) (Figure 3-4A).

To determine the impact of PARP4 copy number status on gene expression levels, I focused on the subset of patients for which RNA sequencing data was available. Patients with shallow or deep deletions were grouped in the same category of copy number loss, while patients with gains or amplifications were classified together as copy number gain (Table 3-2). The normalized RNA

expression of PARP4 was then compared among the copy number loss, diploid and gain groups (Figure 3-4B).

Normalized PARP4 expression was indeed significantly lower in patients with PARP4 copy number loss as compared to patients that were diploid or had gains in PARP4 copy number (Figure 3-4B).

Similarly, in the TCGA LUAD cohort (n=230), the majority (53.0%, n=121) of samples had a loss in PARP4 copy number that was also associated with lower PARP4 expression (Figure 3-4A, C).

Given that the patients from both cohorts predominantly had PARP4 copy number deletions which were accompanied by reduced PARP4 mRNA expression, I hypothesized that PARP4 could be acting as a tumor suppressor where copy number loss leads to the loss of PARP4 expression and function.

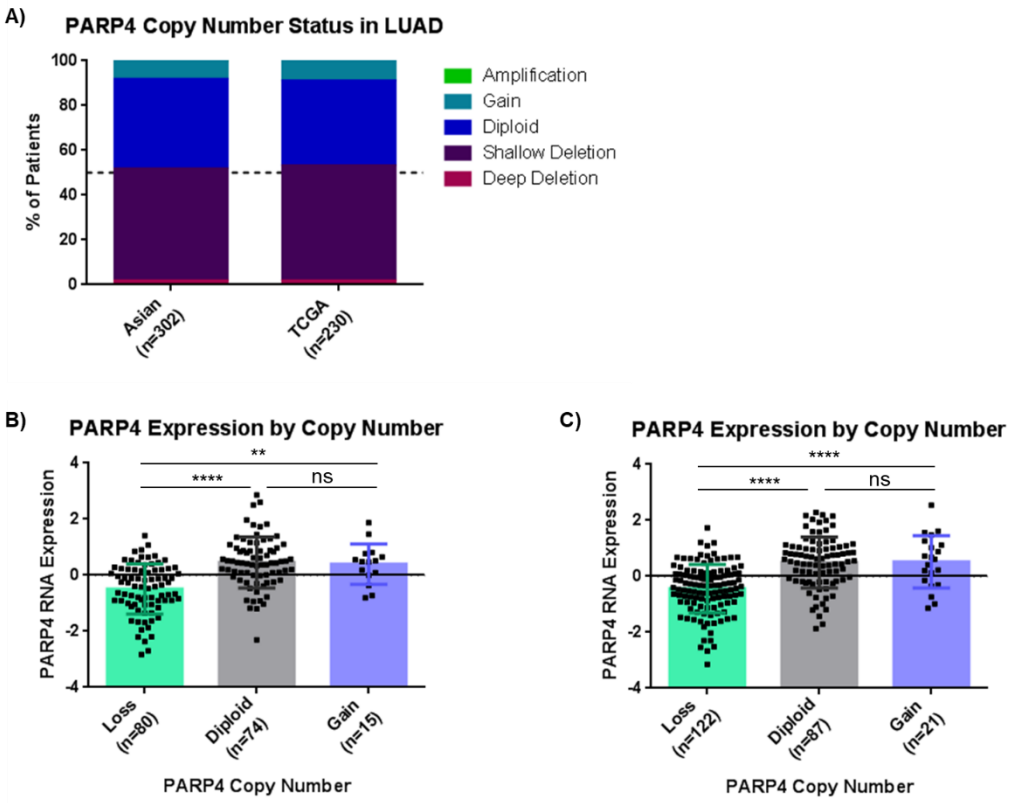


Figure 3-4: Correlation between PARP4 copy number status and RNA expression

A) Bar graph showing the distribution of PARP4 copy number status within the Asian LUAD cohort (n=302) and TCGA LUAD cohort (n=230). **B, C)** Bar graphs showing PARP4 RNA expression z-scores grouped by PARP4 copy number status in **B)** Asian LUAD patients (n=169) and **C)** TCGA LUAD patients (n=230). Data represent the mean ± s.d.; **p<0.01, ****p<0.0001, ns: not significant, as determined by ordinary one-way ANOVA followed by the Tukey test to correct for multiple comparisons. The multiplicity adjusted p-value was reported.

EGFR (47%) and KRAS (11%) are the major oncogenic drivers in the LUAD cohort, and there are cases of PARP4 copy number loss that concurrently had either EGFR or KRAS mutations (Figure 3-5), which are known to be mutually exclusive in LUAD (Chen et al., 2020b; Li et al., 2016; Unni et al., 2015). To determine if there was any association between EGFR or KRAS mutation status as well as PARP4 copy number status, I performed the Fisher's exact test and found that EGFR ($p<0.0001$) or KRAS ($p<0.05$) mutations were indeed more frequently found in cases with PARP4 copy number loss (Table 3-3, Table 3-4). Thus, it would be important to study PARP4 loss in the context of EGFR or KRAS mutations.

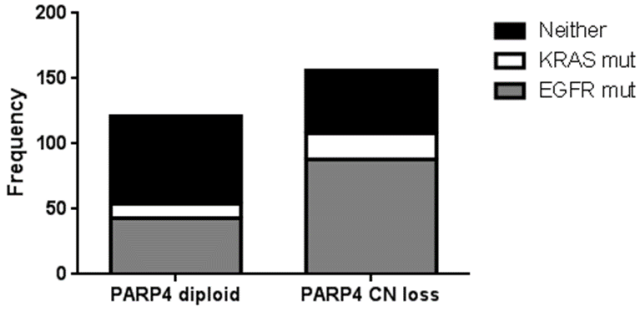


Figure 3-5: Relationship between PARP4 copy number status and EGFR or KRAS mutation status

Data represent the frequency of PARP4 diploid or copy number loss patients with a concurrent mutation in EGFR, KRAS or neither, within the Asian LUAD cohort.

Table 3-3: Contingency table displaying PARP4 copy number and EGFR mutation status

	EGFR mut	*Neither*	*Total*
PARP4 diploid	43	67	110
PARP4 CN loss	88	48	136
Total	131	115	246
Fisher's exact test P-value	< 0.0001		

Table 3-4: Contingency table displaying PARP4 copy number and KRAS mutation status

	KRAS mut	*Neither*	*Total*
PARP4 diploid	11	67	78
PARP4 CN loss	20	48	68
Total	31	115	146
Fisher's exact test P-value	0.0271		

3.2.2 RASA1

Similar to what was observed for PARP4, lower RASA1 expression was significantly associated with poorer overall survival in LUAD (Figure 3-6A).

Unlike in the case of PARP4, slightly over half of the Asian LUAD patients (52.0%, n=157) retained their copy number diploid status, while the remainder were evenly distributed between RASA1 copy number gain (25.2%, n=76) or loss (22.9%, n=69) (Figure 3-6B). Normalized RASA1 expression in the copy number gain patients was significantly higher than patients bearing diploid or loss of RASA1 copy number, although there was no significant difference between the latter two groups (Figure 3-6C).

Interestingly, compared to Asian LUAD, the TCGA LUAD cohort had more cases of RASA1 copy number loss (41.3%, n=95) that was accompanied by significantly lower RASA1 expression (Figure 3-6B, D). Thus, the initial data mining pointed to a potential tumor suppressive role for RASA1 that would be subsequently investigated.

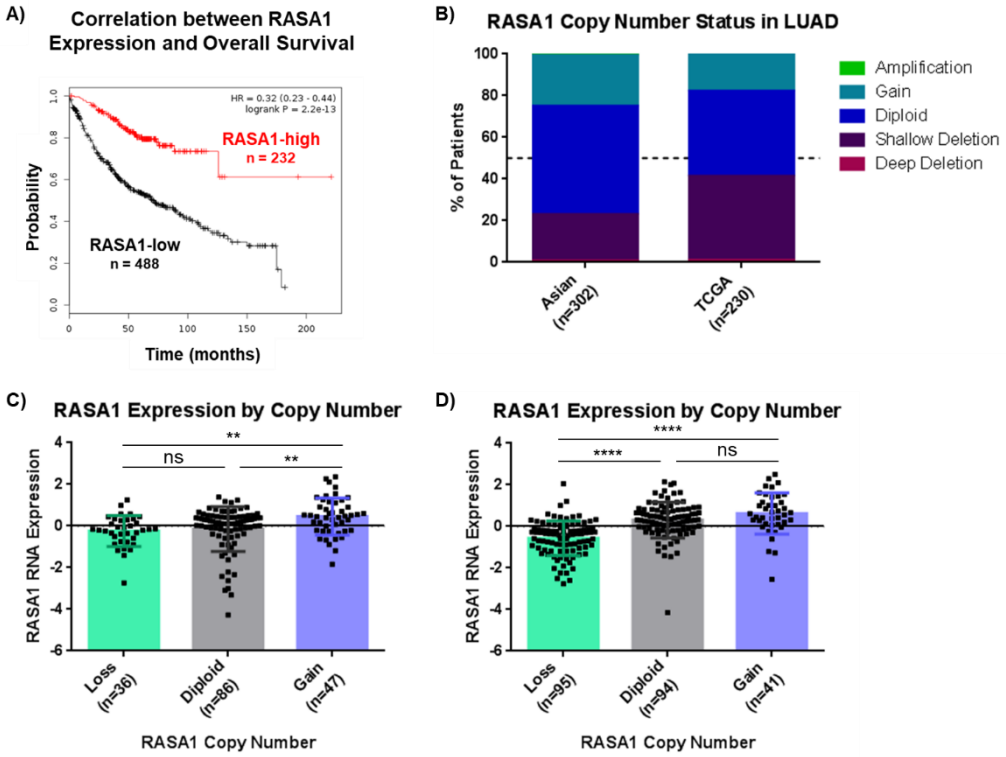


Figure 3-6: RASA1 expression, survival and copy number data

A) Kaplan-Meier plot generated using LUAD microarray data (n=720, Affymetrix ID 202677_at for RASA1) from the KM Plotter database (Győrffy et al., 2013). The patient group with high gene expression (n=232) is displayed in red while the low expression group (n=488) is marked in black. **B)** Bar graph showing number of patients with each RASA1 copy number status within the Asian LUAD cohort (n=302) and TCGA LUAD cohort (n=230). **C, D)** Bar graphs showing normalized RASA1 RNA expression z-scores grouped by RASA1 copy number status in **C)** Asian LUAD patients (n=169) and **D)** TCGA LUAD patients (n=230). Data represent the mean ± s.d.; **p<0.01, ****p<0.0001, ns: not significant, as determined by ordinary one-way ANOVA followed by the Tukey test to correct for multiple comparisons. The multiplicity adjusted p-value was reported. Data was retrieved from OncoSG **(B, C)** and cBio portal (Cerami et al., 2012; Gao et al., 2013) **(B, D)**.

3.3 Genetically defined transformed primary small airway epithelial cells as a model system

To investigate the functions of PARP4 and RASA1 in the LUAD context, I first determined a cell line model that would be relevant to LUAD. I chose human lung primary cells – specifically small airway epithelial cells (SAEC), as LUAD was previously observed to arise from SAEC cells in the distal regions of the lung (Figure 3-7A) (Sainz de Aja et al., 2021; Sutherland and Berns, 2010; Sutherland et al., 2014; Xu et al., 2012). SAEC cells transformed with defined genetic elements mimics step-wise tumorigenesis and would facilitate the characterization of our candidate driver genes as they have a relatively well-defined genetic background with few passenger mutations.

Previous studies have shown that conversion of normal primary human cells towards malignancy requires the acquisition of a minimal set of genetic alterations, which include expression of the telomerase catalytic subunit (TERT) together with two cooperating oncoproteins (Dolma et al., 2003; Hahn and Weinberg, 2002; Hahn et al., 1999, 2002; Pütz et al., 2010). Specifically, TERT overexpression prevents telomere erosion, whereas overexpression of the viral oncoprotein SV40 large T antigen (SV40LT) inhibits RB and p53 to bypass cell cycle checkpoints, collectively enabling primary cells to not only overcome the barriers of senescence and crisis but also tolerate the further overexpression of oncogenic KRAS (Figure 3-7B) (Hahn and Weinberg, 2002).

Our lab had previously generated a series of SAEC lines by the serial addition of TERT and SV40LT (SAECTS), as well as a separate cell line with further addition of mutant $KRAS^{G12V}$ (SAECK) (Figure 3-7D). SAECK cells overexpress TERT, SV40LT and mutant KRAS at high levels compared to parental SAEC cells (Figure 3-7C, E).

Parental primary SAEC cells require a feeder layer to be maintained in culture, whereas SAECTS can be maintained feeder-free (Figure 3-7F, top panel). Consistent with previous reports, SAECTS cells fail to form soft agar colonies or

tumor xenografts in immunodeficient mice, which are classical assays for *in vitro* and *in vivo* tumorigenicity respectively (Figure 3-7F, middle panel).

Similar to SAECTS, SAECK cells can also be maintained feeder-free (Figure 3-7F, top panel). Moreover, and importantly, SAECK cells are capable of both forming soft agar colonies and engrafting in immunodeficient mice (Figure 3-7F, middle and bottom panels). Thus, overexpression of all three genetic elements is essential to convert the SAEC cells into a tumorigenic state.

Both the SAECTS and SAECK cell lines were subsequently used to test candidate driver gene function.

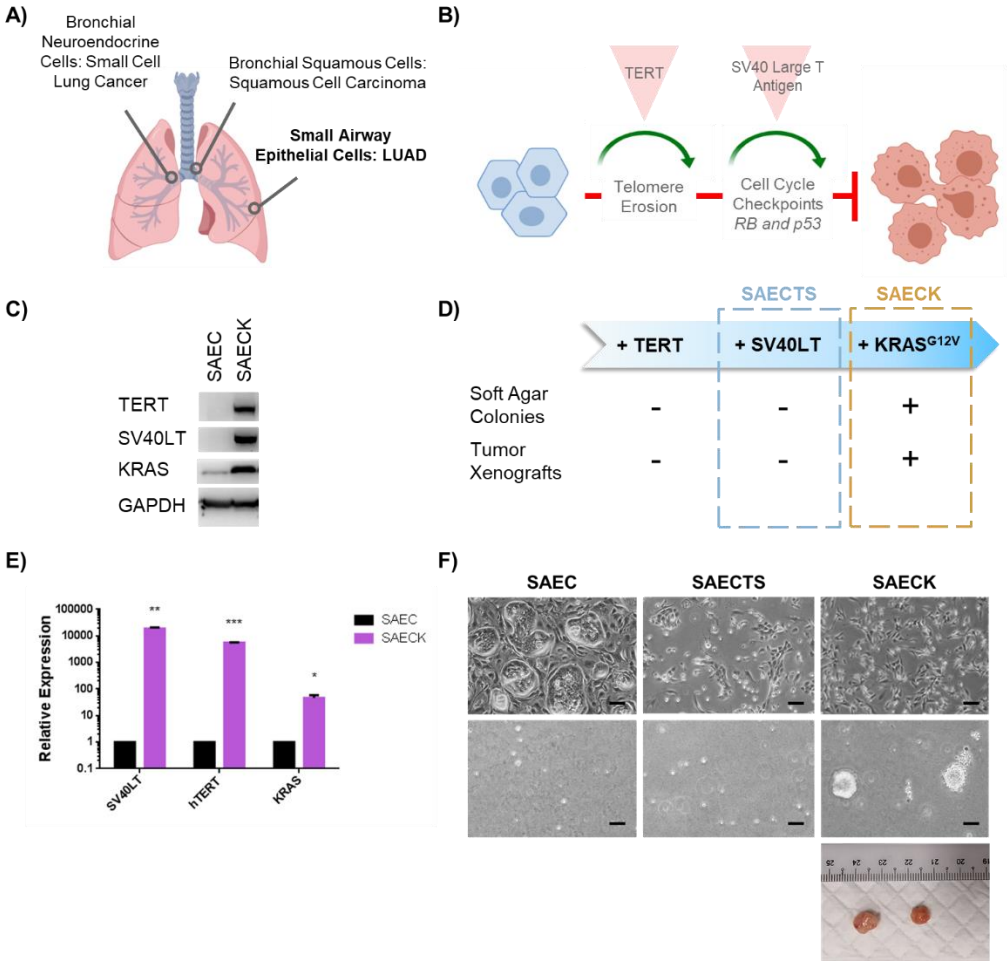


Figure 3-7: Generation of transformed primary small airway epithelial cells

A) Schematic diagram of the human lung and the different cellular origins of lung cancer. **B)** Schematic diagram outlining the barriers to cell proliferation and how overexpression of TERT and SV40 large T antigen can overcome them. **C)** Immunoblot analysis of TERT, SV40LT and KRAS expression in SAECK versus SAEC cells. GAPDH was used as a loading control. **D)** Schematic diagram describing the gain in tumorigenic properties of primary small airway epithelial cells (SAEC) following cumulative addition of genetic elements. The SAECK cell line was generated by Dr. Yuan Ju. **E)** RT-qPCR analysis of transcript levels of the same genes as in **C)**. Data represent the mean ± s.e.m., n = 3; *p<0.05, **p<0.01, ***p<0.001, as determined by unpaired two-tailed t-test. **F)** Representative brightfield images of SAEC, SAECTS and SAECK maintained in culture (top). Representative brightfield images of soft agar colonies formed after 2 weeks. SAEC and SAECTS remained as single cells even after two months (middle). Images taken at 4x magnification. Scale bar represents 100μm. Representative images of tumors formed by subcutaneous injection of 1 million SAECK cells after 12 weeks (bottom).

3.4 RASA1 loss did not alter tumorigenicity of SAECK cells

Based on the nature of RASA1 mutations observed in our Asian LUAD cohort as well as the inverse correlation between RASA1 expression and overall survival, I hypothesized that RASA1 could be a tumor suppressor in LUAD.

To test this, I used three independent shRNAs to generate RASA1 knockdown lines from parental SAECK cells. Of the three shRNAs, only two consistently depleted RASA1 protein and mRNA (Figure 3-8A, B). These two RASA1 knockdown cell lines were then used in soft agar assays, but there was no significant change in the number and size of resultant colonies compared to control cells (Figure 3-8C, D). As loss of RASA1 did not appear to increase *in vitro* tumorigenicity of the SAECK cells, I did not continue to pursue RASA1 as a candidate driver. Nevertheless, it remains possible that only upon transplantation into immunodeficient mice will any differences in their growth rates become apparent.

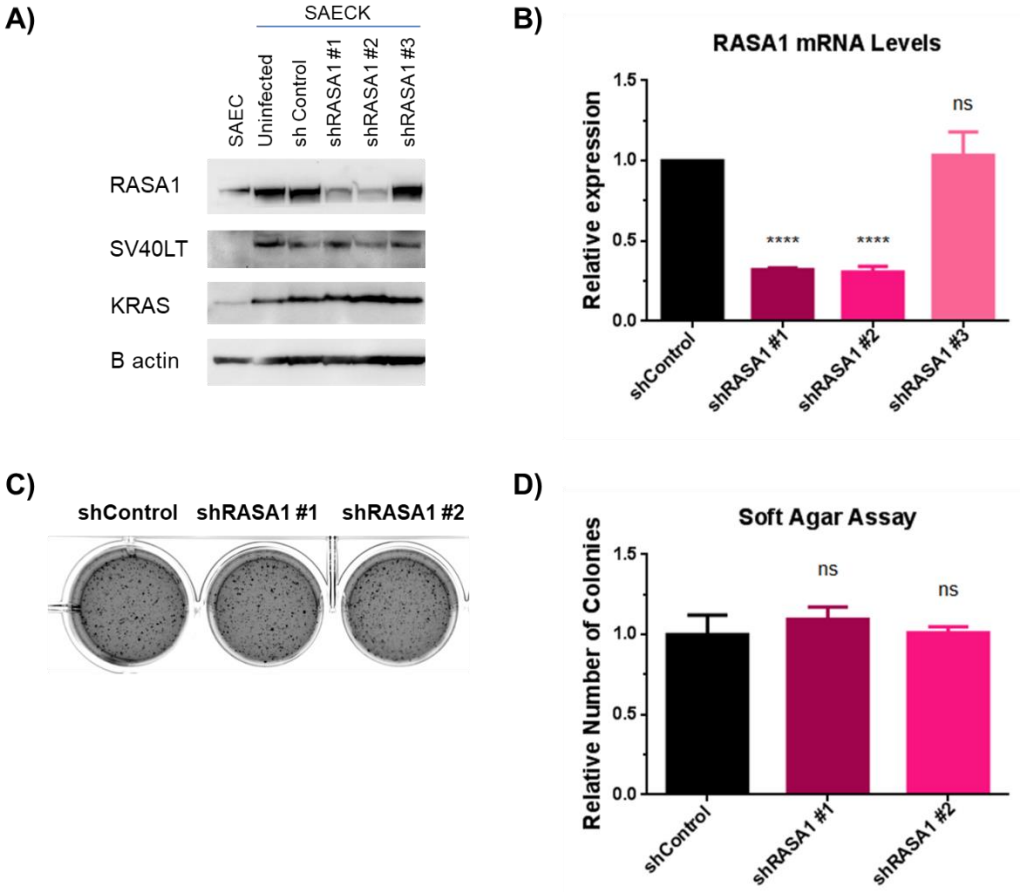


Figure 3-8: RASA1 knockdown did not alter SAECK tumorigenicity

A) Immunoblot showing reduction in RASA1 protein levels upon RASA1 knockdown with shRNA #1 and shRNA #2 in SAECK cells. Beta-actin (B actin) was used as a loading control. **B)** RT-qPCR analysis of RASA1 transcript levels in SAECK control and RASA1 knockdown cells. **C)** Representative images of soft agar colonies stained by crystal violet after 3 weeks of growth. **D)** Quantitation of soft agar colonies. Colonies were counted manually and quantified relative to control. Data represent the mean ± s.e.m., n = 3; ****p<0.0001, ns: not significant, as determined by ordinary one-way ANOVA followed by the Dunnett test to correct for multiple comparisons. The multiplicity adjusted p-value was reported.

3.5 PARP4 is a novel tumor suppressor in LUAD

3.5.1 PARP4 loss increases tumorigenicity of SAECK but not SAECTS

As PARP4 copy number loss was prevalent in the LUAD cohort and was associated with lower expression levels, which were in turn associated with poor prognosis, I hypothesized that PARP4 has a tumor suppressive role. To test this, I proceeded to generate stable PARP4 knockdown SAECK lines using three independent shRNAs. Significant reduction of PARP4 at both the protein and mRNA levels was observed for all three shRNAs, and these cell lines were used in subsequent experiments (Figure 3-9A, B).

While the control and PARP4 knockdown lines did not have significant differences in their proliferation rates when maintained in 2D culture (Figure 3-9C), the PARP4 knockdown lines consistently formed a greater number of soft agar colonies that were also of a larger size compared to the control line (Figure 3-9D-F), indicating their heightened ability for anchorage-independent growth.

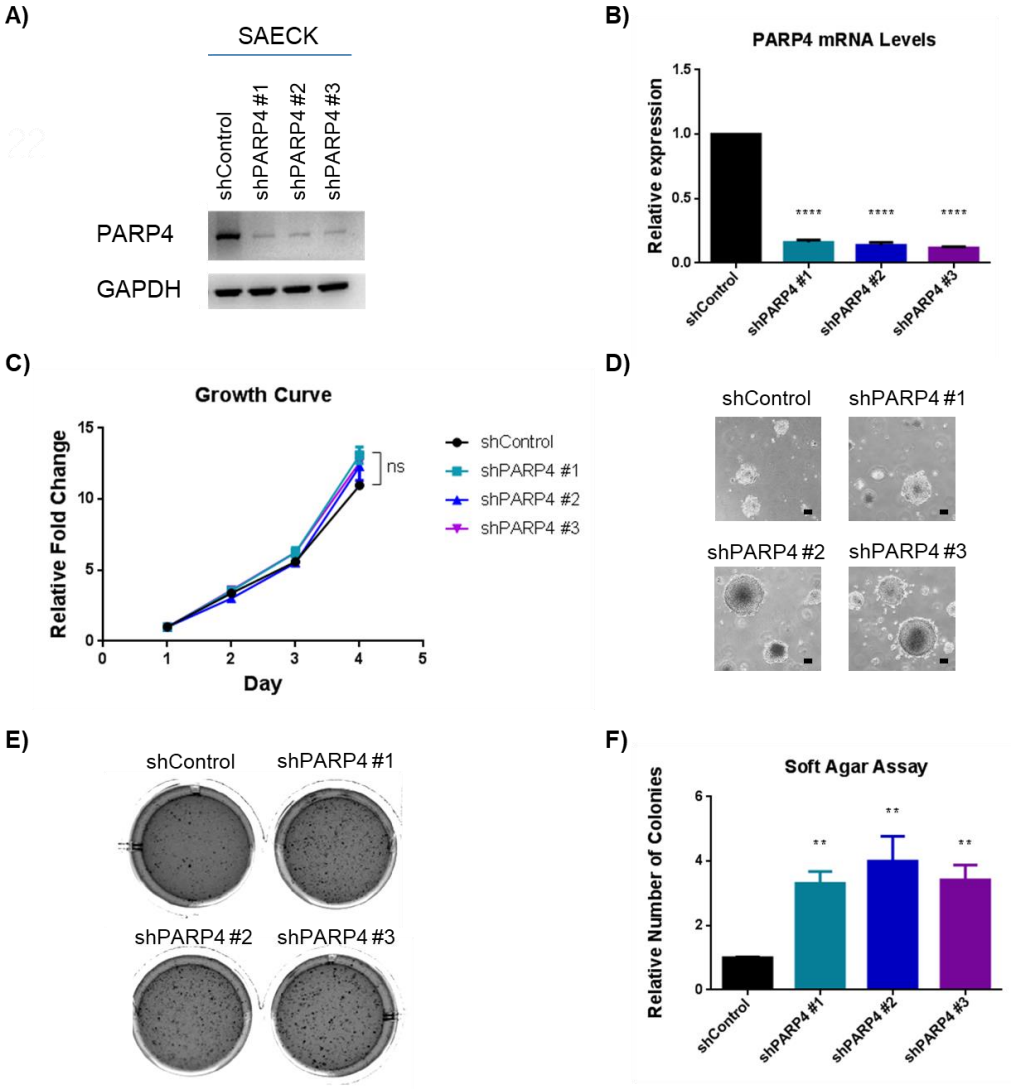


Figure 3-9: Knockdown of PARP4 in SAECK increases *in vitro* tumorigenicity

A) Immunoblot indicating reduction in PARP4 levels upon shRNA knockdown in SAECK cells. GAPDH was used as a loading control. **B)** RT-qPCR analysis of PARP4 transcript levels. **C)** Growth curves comparing proliferative capacity of SAECK control and PARP4 knockdown cells as measured by the Cell-Titer Glo assay. **D)** Representative brightfield images of soft agar colonies. Images were taken at 4x magnification. Scale bar represents 100µm. **E)** Representative images of soft agar colonies stained by crystal violet after 3 weeks of growth. **F)** Quantitation of soft agar colonies. Colonies were counted manually and quantified relative to control. Data represent the mean ± s.e.m., n = 6; **p<0.01, ****p<0.0001, ns: not significant, as determined by ordinary one-way ANOVA followed by the Dunnett test to correct for multiple comparisons. The multiplicity adjusted p-value was reported.

I later selected shPARP4 #1, which consistently had the best knockdown efficiency, for transplantation into immunodeficient NSG mice and observed that the cells formed significantly larger tumors than the control line (Figure 3-10A-C).

To ascertain that the increase in tumorigenicity was indeed a result of PARP4 loss, I proceeded with an orthogonal method of depleting PARP4 from SAECK cells. Whereas shRNA knockdowns were performed in the first set of experiments, I next used guide RNAs (gRNAs) for CRISPR-mediated knockout of PARP4 and was able to successfully reduce PARP4 levels in SAECK cells using two independent gRNAs (Figure 3-10D). As before, the pooled PARP4 knockout cells indeed formed a significantly greater number of soft agar colonies than control cells (Figure 3-10E, F), thus strengthening the evidence for the tumor suppressive properties of PARP4.

Taken together, the loss of PARP4 indeed converts SAECK cells into a more tumorigenic state, and these results are consistent with PARP4 having a tumor suppressive role.

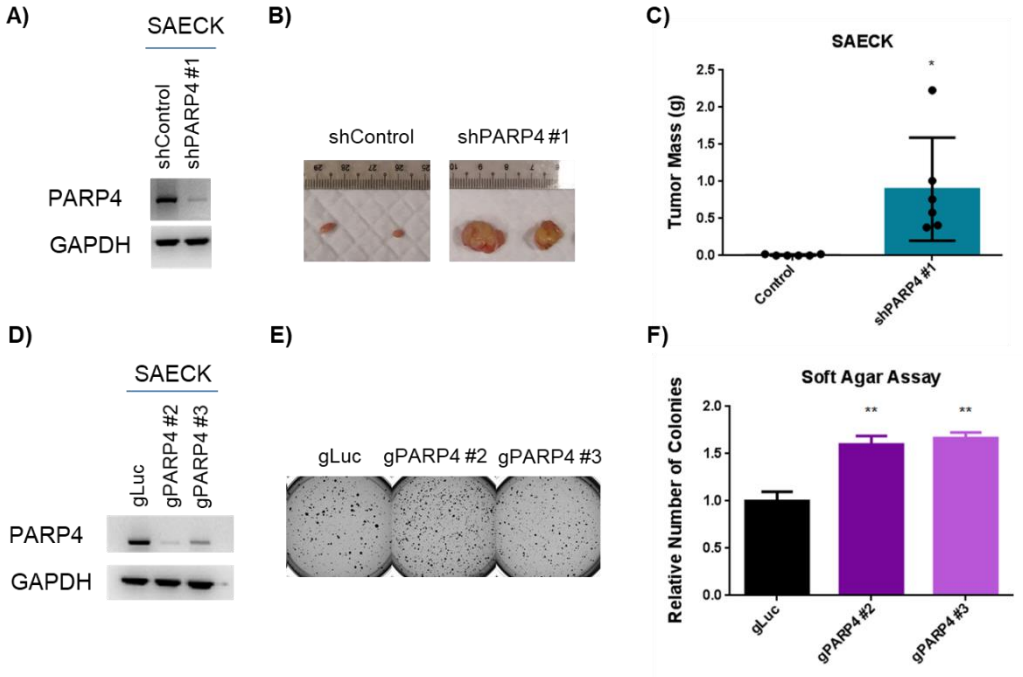


Figure 3-10: Loss of PARP4 increases SAECK tumorigenicity in additional experiments

A) Immunoblot confirming PARP4 knockdown in SAECK cells **B)** Representative images of tumors formed from the subcutaneous injection of 1 million cells per flank of NSG mice. **C)** Bar graph comparing mass of tumors formed after 8 weeks. Data represent the mean ± s.d., n = 6; *p<0.05, as determined by unpaired two-tailed t-test. **D)** Immunoblot confirming reduction in PARP4 levels using the pooled CRISPR knockout method. **E)** Representative images of soft agar colonies stained with crystal violet. **F)** Quantification of soft agar colonies. Data represent the mean ± s.e.m., n = 3; **p<0.01, as determined by ordinary one-way ANOVA followed by the Dunnett test to correct for multiple comparisons. The multiplicity adjusted p-value was reported.

Apart from the tumorigenic SAECK cells, I wanted to determine whether PARP4 loss in the transformed but non-tumorigenic SAECTS cells could convert them into a fully tumorigenic state. Using the same three shRNAs, PARP4 knockdown efficiency of 60%-80% could similarly be achieved in SAECTS cells (Figure 3-11A, B). However, these PARP4 knockdown cells remained unable to form soft agar colonies (Figure 3-11C), indicating that PARP4 loss is insufficient for tumorigenesis and requires a further oncogenic insult.

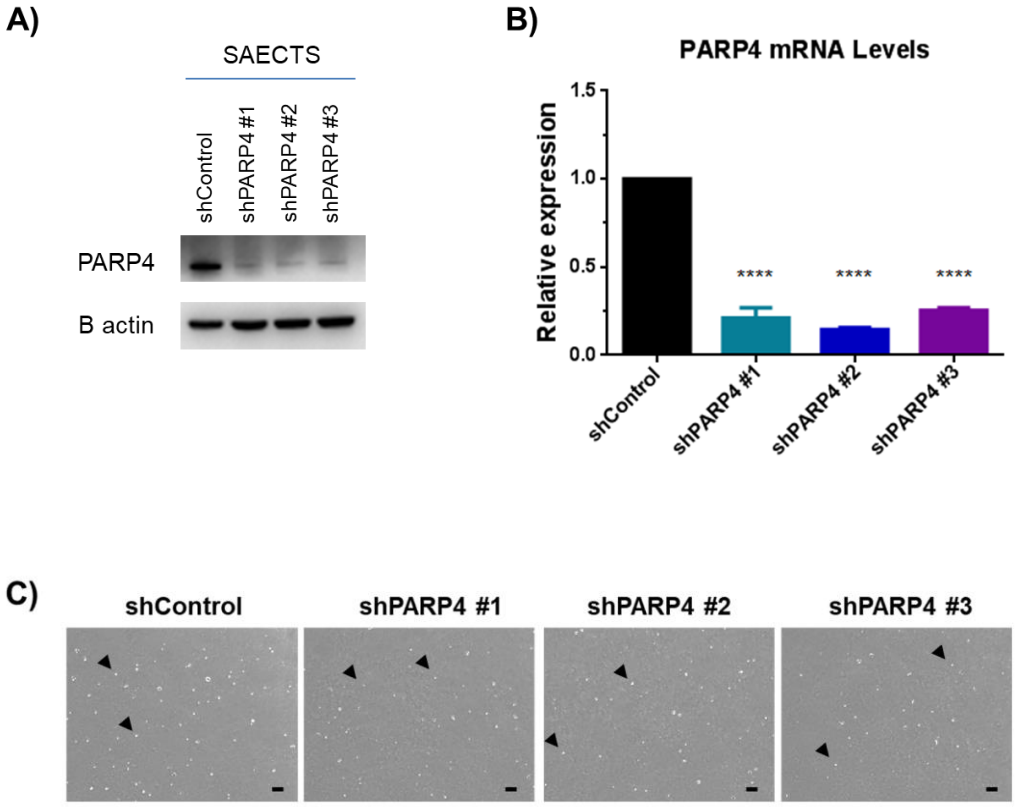


Figure 3-11: PARP4 loss does not convert SAECTS cells into a tumorigenic state

A) Immunoblot indicating knockdown of PARP4 in a primary lung line expressing TERT and SV40 LT (SAECTS). **B)** RT-qPCR analysis indicating a significant reduction in PARP4 transcript levels in knockdown cells. **C)** Representative brightfield images of cells in soft agar at 8 weeks, taken at 4x magnification. Scale bar represents 100µm. Control and PARP4 knockdown cells remained as single cells (examples indicated by black arrowheads) and did not proliferate under anchorage-independent conditions. Data represent the mean ± s.e.m., n = 3; ****$p<0.0001$, as determined by ordinary one-way ANOVA followed by the Dunnett test to correct for multiple comparisons. The multiplicity adjusted p-value was reported.

3.5.2 PARP4 expression is lost in cancer

As the expression levels of many tumor suppressors have been reported to be higher in normal tissues but reduced or lost upon the transition to cancer (Cangemi et al., 2008; Muir and Nunney, 2015; Wang et al., 2018c), I wanted to determine if this was true of PARP4 in the lung context.

Using publicly-available lung cancer datasets, I compared PARP4 expression levels between normal versus lung cancer samples. Indeed, PARP4 was more highly expressed in normal lung cells (Figure 3-12), further supporting its role as a tumor suppressor.

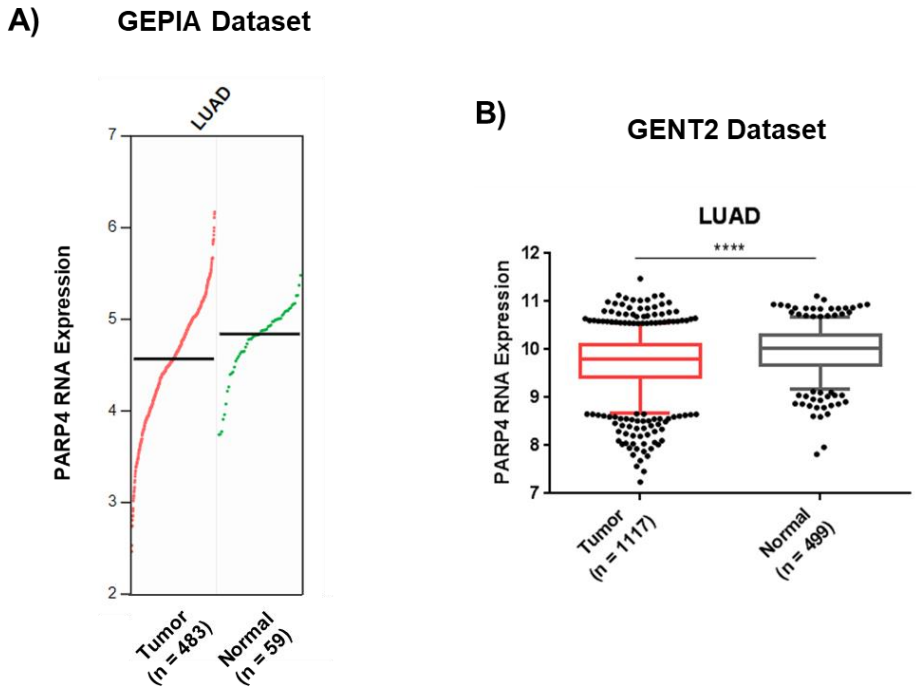


Figure 3-12: PARP4 expression between lung tumor and normal tissue

A) Comparison of PARP4 expression between matched TCGA tumor and normal samples. Y-axis represents $\log_2$(TPM + 1) transformed expression data, where TPM represents transcripts per million. The black line indicates median expression. Data was retrieved from the Gene Expression Profiling Interactive Analysis (GEPIA) database (Tang et al., 2017). **B)** Box-and-whisker plot comparing PARP4 transcript levels between LUAD tumor (n=1117) versus normal tissue (n=499). The 25th, 50th and 75th percentiles are indicated by the boxes, while the whiskers extend to the 5th and 95th percentiles. Beyond, individual points are displayed. ****$p<0.0001$, as determined by unpaired two-tailed t-test. Data was retrieved from the GENT2 database (Park et al., 2019).

To further validate my *in silico* analysis, I proceeded to study endogenous PARP4 expression in a variety of normal and tumor lung cell lines. Our lab had previously developed a method to derive lung cell lines from patient tissue by plating dissociated cells on irradiated NIH-3T3 fibroblasts using a proprietary media formulation. Using this method, we could generate tumor cell lines from surgically resected lung tumor tissue or needle biopsies, as well as normal cell lines from the adjacent normal lung tissue.

From our lab's collection of patient-derived lung cell lines, I observed that PARP4 protein levels were higher in the normal cells relative to the tumor cells (Figure 3-13A). Extending this to commercially available lung cancer cell lines, I similarly observed lower PARP4 levels in the cancer compared to the normal cell lines (Figure 3-13B).

Collectively, these suggest a role for the loss of PARP4 in the progression towards cancer.

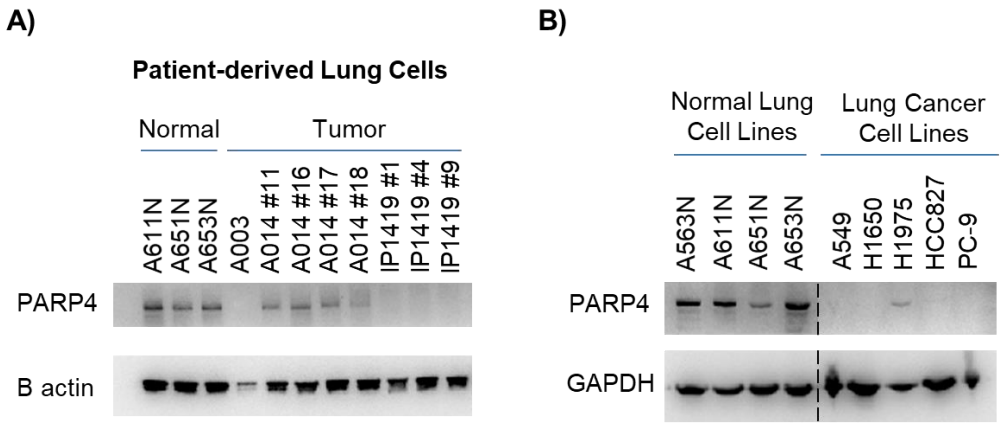


Figure 3-13: PARP4 expression levels are lower in lung cancer cells

A, B) Immunoblot analysis of PARP4 expression in normal patient-derived lung cells compared to **A)** patient-derived tumor cells and **B)** commercial lung cancer cell lines. Beta actin (B actin) and GAPDH are used as loading controls.

3.5.3 PARP4 loss increases tumorigenicity of additional lung cancer cell lines

While PARP4's tumor suppressive activity was evident in the SAECK model system, it was also important to investigate whether these tumor suppressive effects were more universal. As PARP4 copy number loss was earlier shown to be significantly associated with EGFR and KRAS mutations within the Asian LUAD cohort (Figure 3-5), I found it pertinent to determine whether PARP4 modulates tumorigenicity in the same manner in mutant EGFR versus mutant KRAS cells.

From a panel of commercial lung cancer lines, I selected cell lines that had comparatively higher levels of PARP4 from which to further deplete PARP4. PC-9 is one such cell line with a mutant EGFR background (Figure 3-14A). Specifically, PC-9 has a deletion of amino acids 746 to 750 (exon 19 deletion) that renders EGFR constitutively active (Blanco et al., 2009; Kawahara et al., 2010). PARP4-targeting shRNAs were stably introduced into PC-9 cells and were able to successfully reduce PARP4 levels (Figure 3-14B). PC-9 cells with reduced PARP4 expression went on to form a significantly larger number of soft agar colonies (Figure 3-14C-D), indicating their increased *in vitro* tumorigenicity.

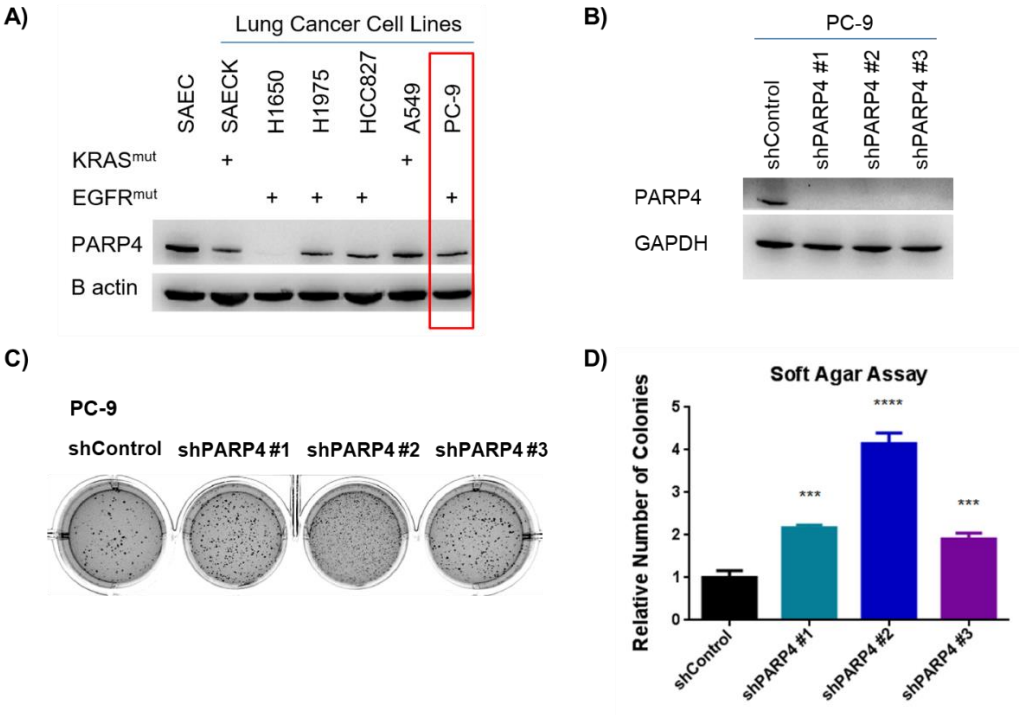


Figure 3-14: PARP4 knockdown in EGFR mutant PC-9 cells increases *in vitro* tumorigenicity

A) Immunoblot analysis of PARP4 levels in a panel of lung cell lines, with KRAS and EGFR mutation status indicated (Blanco et al., 2009; Helfrich et al., 2006). PC-9 is highlighted by the red box. **B)** Immunoblot validation of PARP4 shRNA knockdown in PC-9 cells. **C)** Representative images of soft agar colonies stained by crystal violet after 3 weeks of growth. **D)** Quantitation of soft agar colonies. Colonies were counted manually and quantified relative to control. Data represent the mean ± s.e.m., n = 3; ***p<0.001, ****p<0.0001, as determined by ordinary one-way ANOVA followed by the Dunnett test to correct for multiple comparisons. The multiplicity adjusted p-value was reported.

PARP4-targeting gRNAs were used to deplete PARP4 levels via the orthogonal method of CRISPR-mediated PARP4 knockout in A549 cells, which harbour constitutively active $KRAS^{G12S}$ (Blanco et al., 2009; Helfrich et al., 2006) (Figure 3-15A). After confirming the successful reduction in PARP4 levels via immunoblotting (Figure 3-15B), the cells were subcutaneously injected into immunodeficient mice. The PARP4-depleted A549 cells formed larger tumors than the control (Figure 3-15C, D), indicating that PARP4 loss indeed results in increased *in vivo* tumorigenicity.

Taken together, these experiments demonstrate the relevance of PARP4's tumor suppressive roles in additional lung cancer cell lines and suggest that PARP4 loss increases tumorigenicity of both mutant EGFR- and KRAS-driven cells.

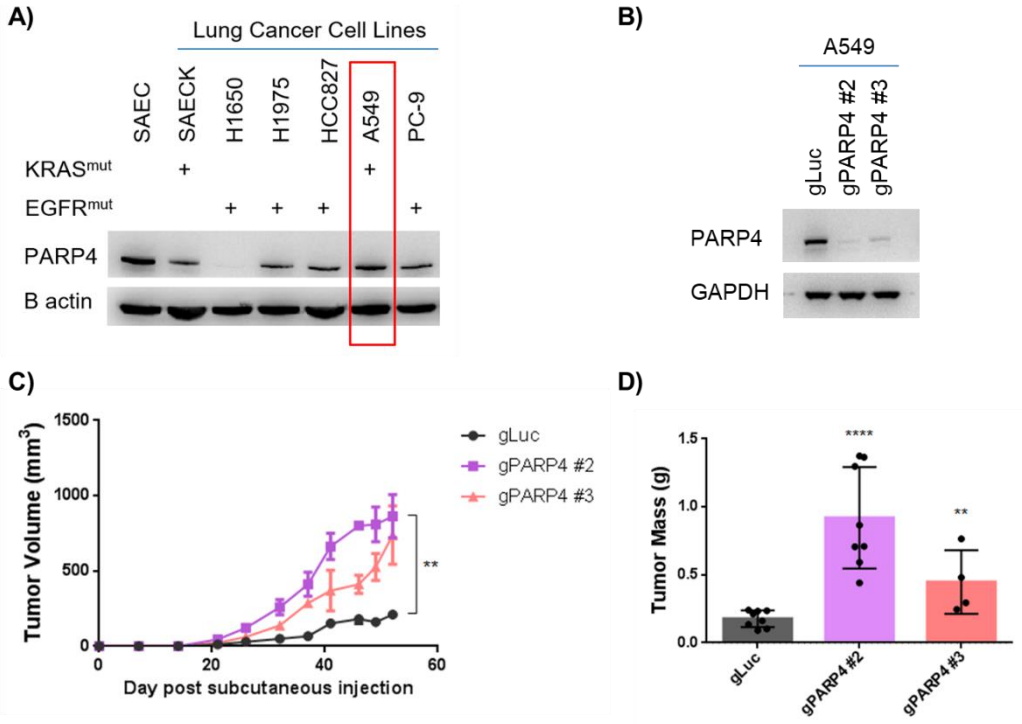


Figure 3-15: PARP4 loss in KRAS-mutant A549 cells increases *in vivo* tumorigenicity

A) Immunoblot analysis of PARP4 levels in a panel of lung cell lines, with KRAS and EGFR mutation status indicated (Blanco et al., 2009; Helfrich et al., 2006). A549 is

highlighted by the red box. **B)** Immunoblot validation of reduction in PARP4 levels following PARP4 pooled CRISPR knockout in A549 cells using two independent gRNAs. **C)** Growth curve of tumors formed from the subcutaneous injection of 400,000 cells per flank of NSG mice. Tumor volume was calculated as 0.5 x length x width^2. **D)** Bar graph comparing mass of tumors formed after 8 weeks. Data represent the mean ± s.d., n ≥ 4; **p<0.01, ****p<0.0001, as determined by ordinary one-way ANOVA followed by the Dunnett test to correct for multiple comparisons. The multiplicity adjusted p-value was reported.

3.5.4 Recurrent I1039T mutation in PARP4 contributes to tumorigenicity

Having accounted for the potential outcome of PARP4 copy number loss, I next wanted to determine the functional significance of the recurrent I1039T mutation observed in the Asian LUAD cohort.

To explore this, a clonal PARP4 knockout (KO) cell line was first generated from SAECK cells using a gRNA targeting exon 3 within the PARP4 gene. The clonal KO was subsequently validated by Sanger sequencing to have a homozygous frameshift mutation at the gRNA target site (Figure 3-16A) and was used for overexpression of either wildtype PARP4 (PARP4^{WT}) or mutant PARP4^{I1039T}.

Because of PARP4's size, the length of the overexpression construct (12.4 kilobases) was close to the limit for efficient production of lentiviral particles and adversely affected transduction efficiency (Kumar et al., 2001; Sweeney and Vink, 2021). It was generally difficult to attain high overexpression levels of PARP4, with the parental SAECK line still having overall higher PARP4 levels (Figure 3-16B).

However, despite repeated lentiviral transduction and fluorescence-activated cell sorting to increase the levels of PARP4 protein expression, the mutant I1039T line always had lower PARP4 protein compared to the wildtype (Figure 3-16B). This was in spite of comparable PARP4 overexpression at the transcript level between the two cell lines (Figure 3-16C).

To determine whether the mutation in its endogenous context was associated with any expression changes, I compared PARP4 transcript levels between I1039T mutant and wildtype cases in the LUAD cohort data and found no significant difference (Figure 3-16D). Unfortunately, there was no proteome data available, for it would have been illuminating to determine if the I1039T mutation was correlated with lower PARP4 protein expression in patients.

Instead, I examined the possible functional outcome of the I1039T mutation using a web-based tool, PolyPhen-2, which predicts the impact of amino acid substitutions on protein structure and function. PolyPhen-2 considers the structural attributes and evolutionary conservation of the protein and amino acid residue in question, and employs machine-learning approaches to estimate the probability of a mutation having a deleterious effect on protein function (Adzhubei et al., 2013, 2010). PolyPhen-2 then assigns a probability score on the scale of 0 to 1, based on increasing likelihood that the mutation will negatively affect protein function. In the case of PARP4, the I1039T mutation received a score of 0.856 and was predicted to be "possibly damaging" (Figure 3-16E), indicating that it could potentially have a deleterious effect.

Having verified that PARP4 mRNA levels were comparable between the two cell lines, I transplanted them subcutaneously into NSG mice to determine the effect of the I1039T mutation on tumorigenicity. As the mutant I1039T line formed larger tumors than the wildtype line (Figure 3-16F, G), I concluded that the recurrent I1039T mutation functionally contributes to tumorigenicity.

Given the prediction of a deleterious effect, as well as the lower protein levels in spite of matched RNA levels, I speculated that the I1039T mutation may have destabilized the protein, resulting in effectively a loss-of-function phenotype where PARP4 protein levels are lowered, and thereby increasing tumor formation. This hypothesis could be further ascertained and studied using cycloheximide and MG-132 to inhibit protein synthesis and degradation respectively.

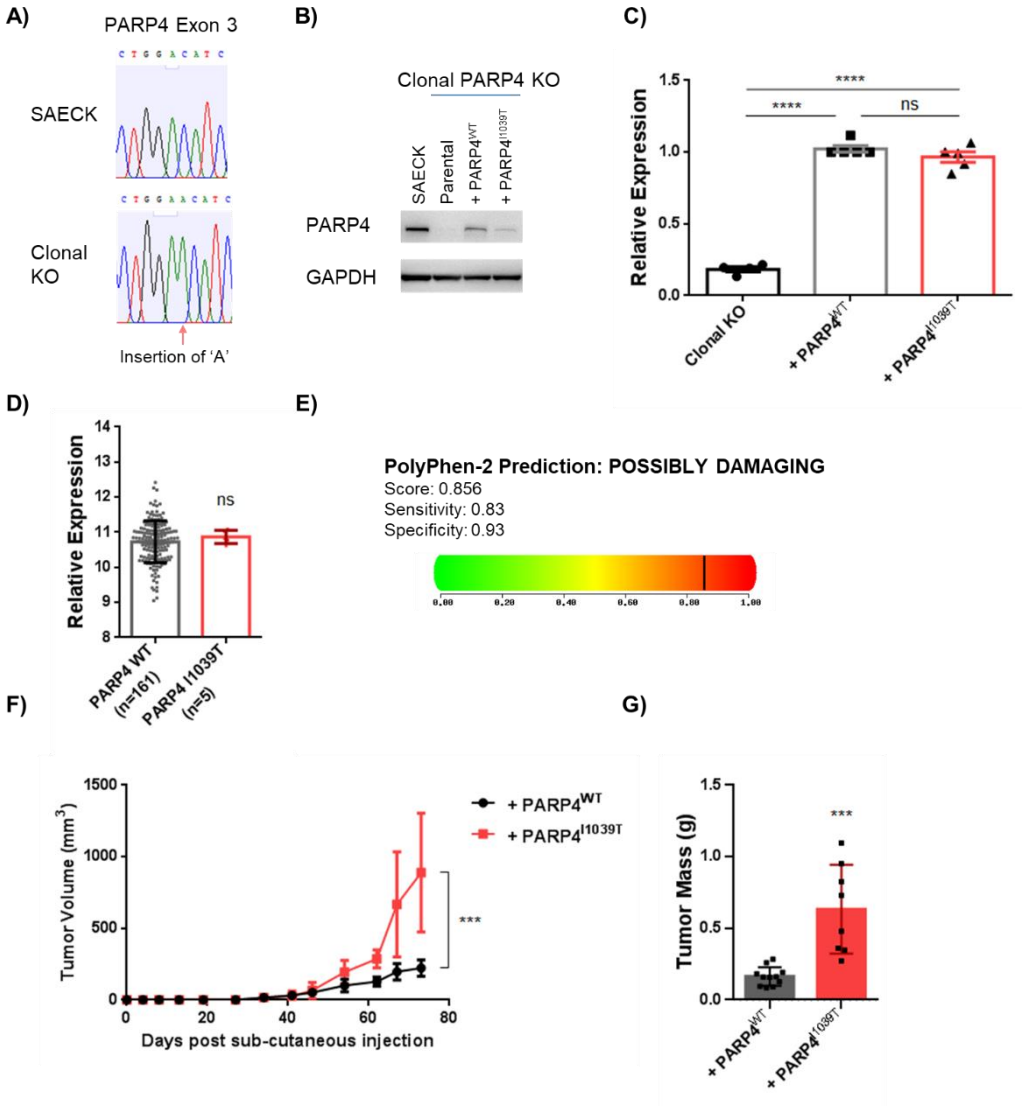


Figure 3-16: PARP4 I1039T results in reduced PARP4 protein and increases tumorigenicity

A) Sanger sequencing chromatogram depicting frameshift mutation within PARP4 exon 3 in the PARP4 clonal KO line. **B)** Immunoblot analysis indicating lower PARP4 expression in clonal PARP4 KO overexpressing $PARP4^{WT}$ compared to SAECK, and even lower PARP4 expression in clonal PARP4 KO overexpressing $PARP4^{I1039T}$. **C)** RT-qPCR analysis indicating similar PARP4 transcript levels between clonal PARP4 KO overexpressing $PARP4^{WT}$ or $PARP4^{I1039T}$. Data represent the mean ± s.e.m., n = 5; ****p<0.0001, ns: not significant, as determined by ordinary one-way ANOVA followed by the Tukey test to correct for multiple comparisons. The multiplicity adjusted p-value was reported. **D)** Relative PARP4 expression in $PARP4^{WT}$ and $PARP4^{I1039T}$ patients from the Asian LUAD cohort. Data represent the mean ± s.d., n ≥ 5; ns: not significant, as determined by unpaired two-tailed t-test. **E)** Results from PolyPhen-2 analysis of I1039T mutation. **F)** Growth curve of tumors formed from the subcutaneous injection of 1 million cells per flank of NSG mice. Tumor volume was calculated as 0.5

x length x width^2. **G)** Bar graph comparing mass of tumors formed after 8 weeks. Data represent the mean ± s.d., n ≥ 8; ***p<0.001, as determined by unpaired two-tailed t-test.

3.5.5 PARP domain is crucial for PARP4's tumor suppressive activity

PARP4 is known to possess ADP-ribosylation activity – *in vitro* assays have revealed that PARP4 is capable of auto-ADP-ribosylation (Kickhoefer et al., 1999a; Vyas et al., 2014). In the rat, PARP4 was also observed to ADP-ribosylate its fellow vault complex member MVP (Kickhoefer et al., 1999a). However, the functional significance of this catalytic activity remains to be understood. Hence, I wanted to determine if PARP4's ADP-ribosylation activity is important for its tumor suppressive function.

To test this, I overexpressed mutant PARP4 lacking its PARP catalytic domain in the same PARP4 clonal knockout SAECK line as before and performed *in vivo* tumorigenicity assays (Figure 3-17A, B). Compared to wildtype PARP4, the PARP domain deletion resulted in larger tumors (Figure 3-17C, D), suggesting that the PARP domain is indeed crucial for PARP4's tumor suppressive activity.

Apart from its ADP-ribosylation activity, the PARP domain in another PARP family member PARP1 was previously reported to bind p53 and histones, suggesting that the domain may have additional roles in mediating protein interactions (Kumari et al., 1998; Pinnola et al., 2007; Wacker et al., 2007).

Given that PARP4's PARP domain appears to be functionally important for tumorigenicity, I intend to dissect the two potential functionalities of the PARP domain – ADP-ribosylation and protein interaction, by mutating key active site residues to generate a catalytically inactive PARP4 mutant that retains the overall PARP domain structure. Comparing the tumorigenicity of this variant relative to the PARP domain-deleted and wildtype PARP4 would give an indication of the degree by which ADP-ribosylation activity contributes to PARP4's tumor suppression.

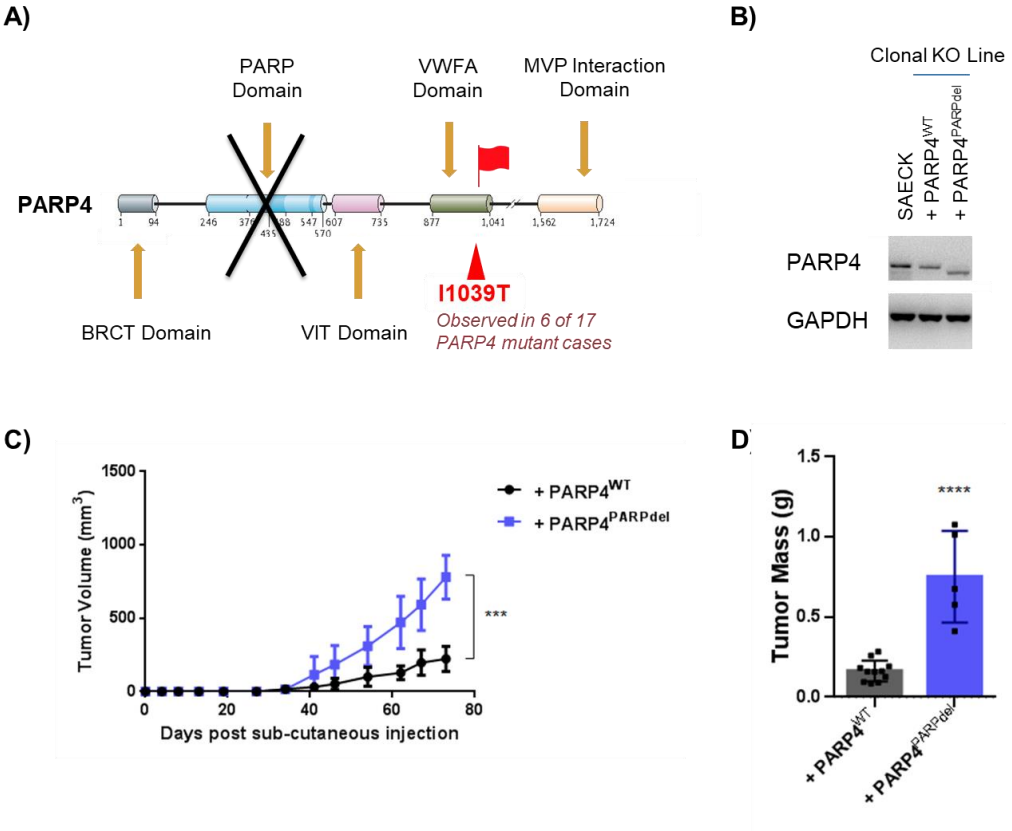


Figure 3-17: Loss of PARP domain increases tumorigenicity

A) Schematic diagram of PARP4 protein structure, with the respective domains and location of the I1039T mutation indicated. Figure was adapted from (Schreiber et al., 2006). **B)** Immunoblot analysis of PARP4. GAPDH was used as a loading control. **C)** Growth curve of tumors formed from the subcutaneous injection of 1 million cells per flank of NSG mice. Tumor volume was calculated as 0.5 x length x width^2. **D)** Bar graph comparing mass of tumors formed after 8 weeks. Data represent the mean ± s.d., n ≥ 5; ***p<0.001, ****p<0.0001, as determined by unpaired two-tailed t-test.

3.5.6 Graphical summary

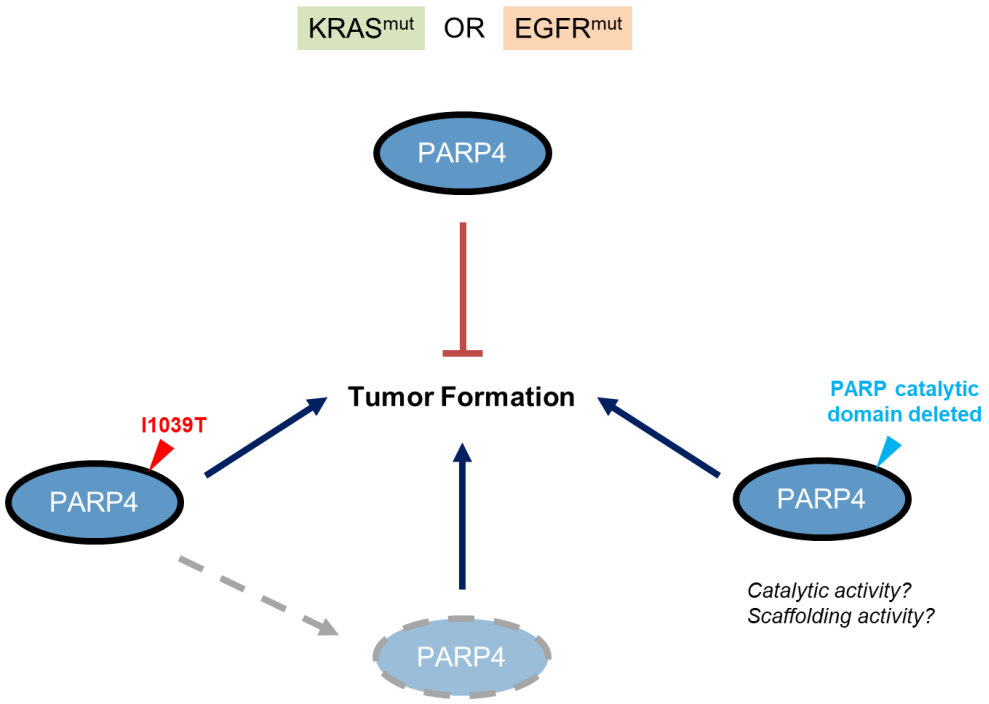


Figure 3-18: Schematic diagram summarizing effects of PARP4 loss or mutations on tumorigenicity

In lung cancer cells that are driven by mutant KRAS or mutant EGFR, loss of PARP4 promotes tumorigenicity. The recurrent I1039T mutation in PARP4 leads to the same phenotype, possibly by reducing levels of PARP4 protein. Deletion of the PARP catalytic domain similarly increases tumor formation, thus suggesting the involvement of PARP catalytic or scaffolding activity in PARP4-mediated tumor suppression.

3.6 PARP4's tumor suppressive activity is independent of the vault complex

3.6.1 MVP loss does not alter tumorigenicity

Given that much of the literature surrounding PARP4 focuses on its association with the vault complex, I asked if the vault complex could be implicated in PARP4's tumor suppressive activity.

Firstly, I determined the effect of expression levels of the key structural component of the vault complex, MVP, on LUAD patient overall survival. Using the same set of LUAD microarray and survival data from the PARP4 and RASA1 comparisons earlier (Győrffy et al., 2013), I found to my surprise that unlike in the case of PARP4, MVP expression levels had no significant impact on overall survival (Figure 3-19A), suggesting that MVP might not be important in LUAD tumorigenicity.

I next set out to test this by depleting MVP from SAECK cells through parallel means of shRNA knockdown and gRNA knockout (Figure 3-19B, E), which would effectively result in the loss of the vault complex structure.

MVP loss did not phenocopy PARP4 loss, as there were no statistically significant effects on the number of soft agar colonies formed (Figure 3-19C, D) nor the size of resulting tumor xenografts (Figure 3-19F, G). Thus, loss of MVP neither increases *in vitro* nor *in vivo* tumorigenicity, in spite of the observation that PARP4 levels were slightly reduced upon MVP loss (Figure 3-19E). One possible explanation could be that the reduction of PARP4 in the MVP-depleted cells did not meet a threshold necessary to observe a significant phenotype.

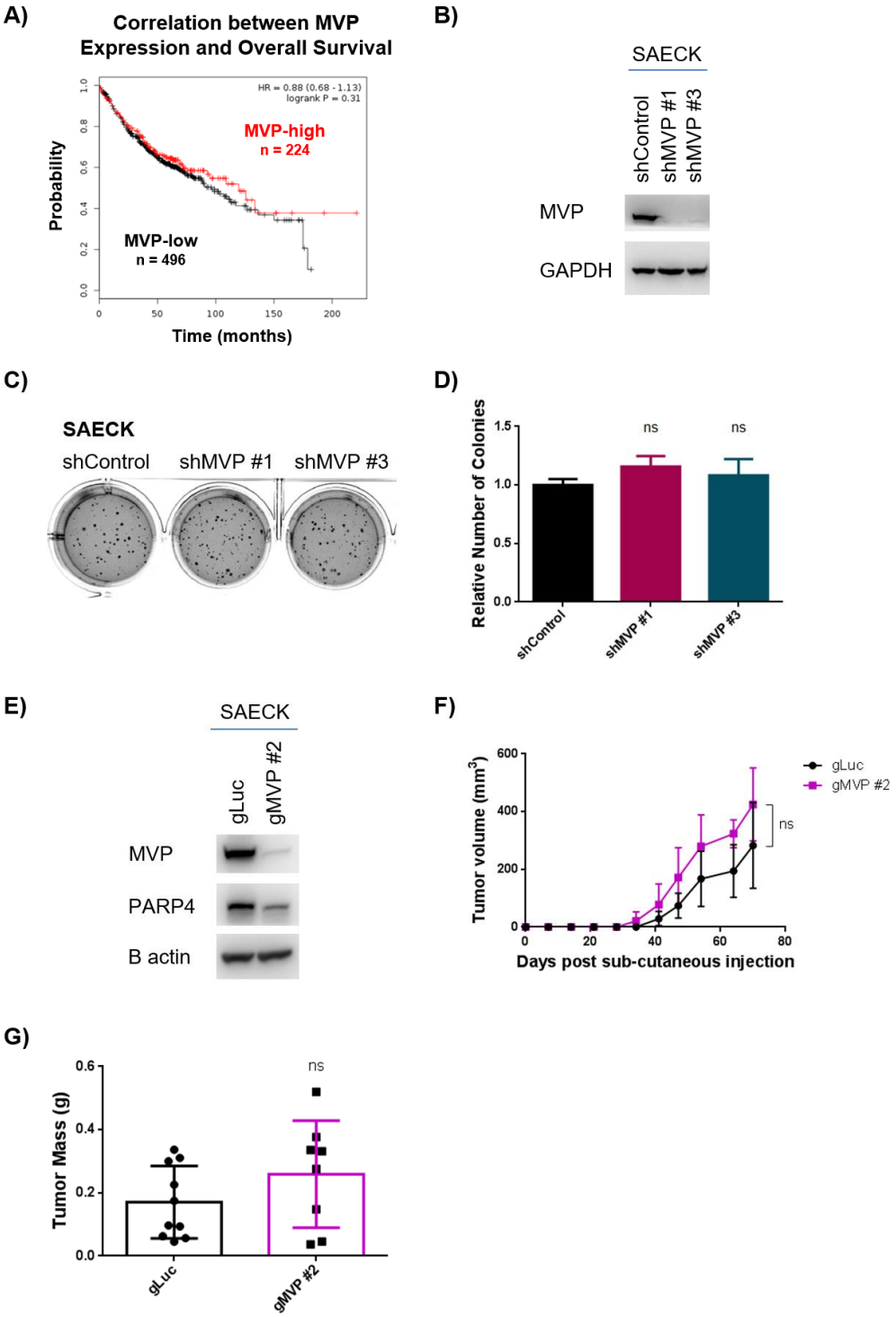


Figure 3-19: MVP loss has no significant effect on tumorigenicity

A) Kaplan-Meier plot generated using LUAD microarray data (n=720, 202180_s_at for MVP) from the KM Plotter database (Győrffy et al., 2013). The patient group with high MVP gene expression (n=224) is displayed in red while the low expression group (n=496) is marked in black. **B)** Immunoblot validation of MVP knockdown in SAECK cells.

C) Representative images of soft agar colonies stained by crystal violet after 3 weeks of growth. **D)** Quantitation of soft agar colonies. Colonies were counted manually and quantified relative to control. Data represent the mean ± s.e.m., n = 6; ns: not significant, as determined by ordinary one-way ANOVA followed by the Dunnett test to correct for multiple comparisons. The multiplicity adjusted p-value was reported. **E)** Immunoblot validation of MVP depletion via pooled CRISPR knockout in SAECK cells. **F)** Growth curve of tumors formed from the subcutaneous injection of 1 million cells per flank of NSG mice. Tumor volume was calculated as 0.5 x length x width^2. **G)** Bar graph comparing mass of tumors harvested after 9 weeks. Data represent the mean ± s.d., n ≥8; ns: not significant, as determined by unpaired two-tailed t-test.

3.6.2 MVP-independent PARP4 may confer greater tumor suppressive activity than MVP-bound PARP4

Interestingly, in addition to the reduction in PARP4 levels observed in MVP knockdown cells (Figure 3-20A), cell lines from our panel of normal and cancer lung lines with reduced levels of MVP also had lower PARP4 expression (Figure 3-20B, C). However, the converse was not true – PARP4-low lines were not necessarily always MVP-low. These seemingly confounding observations raise the possibility that PARP4 levels and regulation may differ between its MVP-associated or MVP-independent forms, which have been described in other model systems (Kickhoefer et al., 1999a; Liu et al., 2004; Schroeijers et al., 2000; van Zon et al., 2003b).

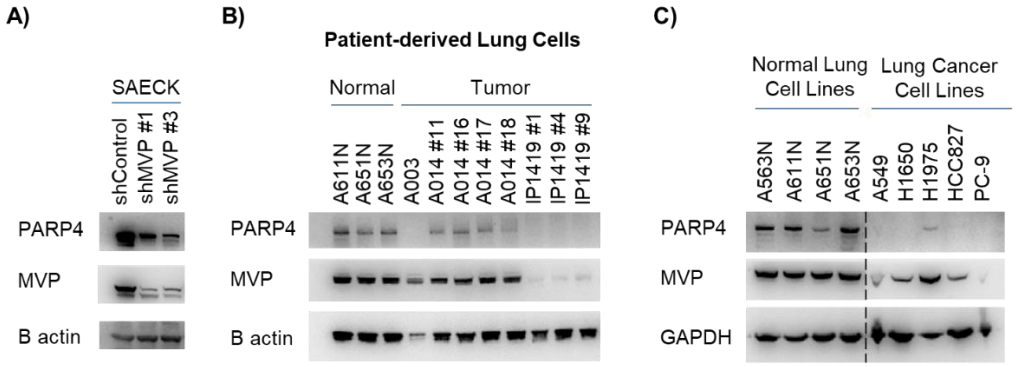


Figure 3-20: Low MVP protein expression is associated with reduced PARP4 protein

A) Immunoblot validation of successful MVP knockdown, with concomitant reduction in PARP4 levels in SAECK cells. **B, C)** Immunoblot analysis of PARP4 and MVP expression levels in a panel of **B)** patient-derived normal and tumor lung lines as well as **C)** commercially available lung cancer cell lines.

Previous studies suggest that PARP4 protein is stabilized by its incorporation within the vault complex through binding with MVP via its C-terminal MVP-interacting domain (Figure 3-21A) (Kozlov et al., 2006; Yu et al., 2017). To assess the association between MVP and PARP4 levels in our cell lines, I overexpressed an MVP domain-deleted PARP4 variant in a clonal PARP4 knockout SAECK line and showed that immunoprecipitation of PARP4 failed to co-immunoprecipitate MVP (Figure 3-21B). This verifies that PARP4 interacts with MVP through its C-terminal domain. Furthermore, in spite of reduced PARP4 protein seen in MVP-depleted cell lines, PARP4 mRNA levels were comparable with control cells (Figure 3-21C, D). This thus points to the role of post-transcriptional regulation of PARP4 levels by MVP.

To determine whether MVP regulates PARP4 protein stability, I treated control and MVP-depleted SAECK cells with the proteasomal inhibitor MG-132. MG-132-treated cells, expectedly, accumulated ubiquitinated proteins that were not cleared by the proteasome (Figure 3-21E, top panel). In MVP-depleted cell lines, accumulation and stabilization of PARP4 levels were observed over time and eventually reached levels that were comparable with cell lines with intact MVP expression (Figure 3-21E, red boxes), suggesting that the stability of a significant fraction of PARP4 protein indeed depends on MVP. Interestingly, in the untreated condition, PARP4 protein was not completely depleted by loss of MVP (Figure 3-21E), suggesting that a fraction of cellular PARP4 protein was resistant to MVP loss and could exist independently of MVP. Moreover, PARP4's tumor suppressive role likely resides in this MVP-independent fraction, as evidenced by the lack of tumorigenicity in MVP-depleted cells (Figure 3-19).

Indeed, immunofluorescence studies in HeLa cells had previously revealed that PARP4 was not entirely associated with the vault complex but was also found in the nucleus, from which vault complexes were excluded (Kickhoefer et al., 1999a; van Zon et al., 2003a). Notably, PARP4 possesses two classical lysine (K)- or arginine (R)-rich nuclear localization signals (NLS) at amino acid residues 19-25 within the BRCT domain (PQQQKKK) and 1237-1249 between the VWFA and MVP interaction domains (KRKHRKIPFSKRK)

(Figure 3-21A), which are predicted to allow importin-regulated nuclear import (Kalderon et al., 1984; Nguyen Ba et al., 2009). These findings thus suggest possible vault-independent roles for PARP4. However, a different study reported that the nuclear localization of PARP4 could be an artefact (Vyas et al., 2013). Thus, I wanted to determine PARP4's localization in the SAECK cells.

To do so, I performed cellular fractionation and found that while the majority of PARP4 localizes to the cytoplasm, a smaller subset does indeed exist in the nucleus (Figure 3-21F). It would be interesting to subsequently study the requirement of nuclear PARP4 for tumor suppression by deleting PARP4's predicted nuclear localization signals.

Collectively, these different lines of evidence suggest that PARP4's tumor suppressive activity is independent of the vault complex.

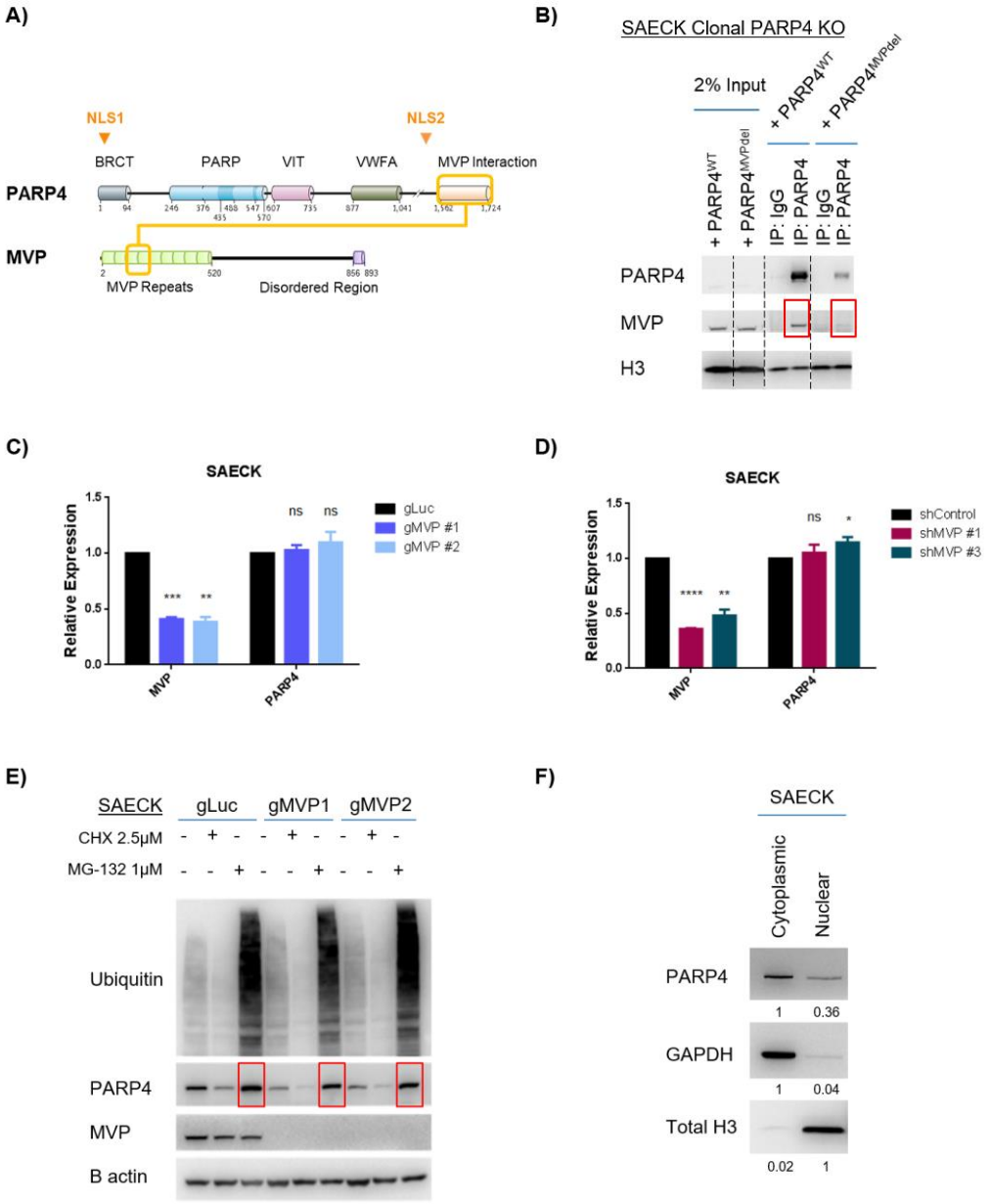


Figure 3-21: MVP-dependent and independent fractions of PARP4

A) Schematic diagram of PARP4 and MVP domain features, with the location of the identified interaction site on both proteins marked out in orange. The location of PARP4's predicted nuclear localization signals (NLS) are indicated by the orange arrowheads. PARP4 domain structure adapted from (Schreiber et al., 2006). **B)** Immunoblot depicting immunoprecipitation of PARP4 from wildtype and MVP domain-deleted PARP4 overexpression in a clonal PARP4 knockout cell line. **C, D)** RT-qPCR analysis of MVP and PARP4 transcript levels in **C)** SAECK gMVP #1 and gMVP #2 cells relative to gLuc, and **D)** SAECK shMVP #1 and shMVP #3 cells relative to shControl. Data represent the mean ± s.e.m., n = 3; *p<0.05, **p<0.01, ***p<0.001,

****p<0.0001, ns: not significant, as determined by ordinary one-way ANOVA followed by the Dunnett test to correct for multiple comparisons. The multiplicity adjusted p-value was reported. **E)** Immunoblot depicting loss of PARP4 protein from MVP-depleted SAECK cells at steady state and upon cycloheximide (CHX) inhibition of protein synthesis. MG-132 inhibition of proteasomal degradation led to accumulation of ubiquitinated proteins and reverted PARP4 levels back to that comparable with the gLuc control (red boxes). Cells were treated with the indicated concentrations of CHX and MG-132 for 24h. **F)** Immunoblot depicting PARP4 protein levels following cell fractionation. Equal amounts of total protein from both fractions were used. GAPDH and total histone H3 were used as specific markers for the cytoplasmic and nuclear fractions respectively. Band intensity was quantified relative to the cytoplasmic or nuclear fraction and indicated under the respective bands.

3.6.3 Graphical summary

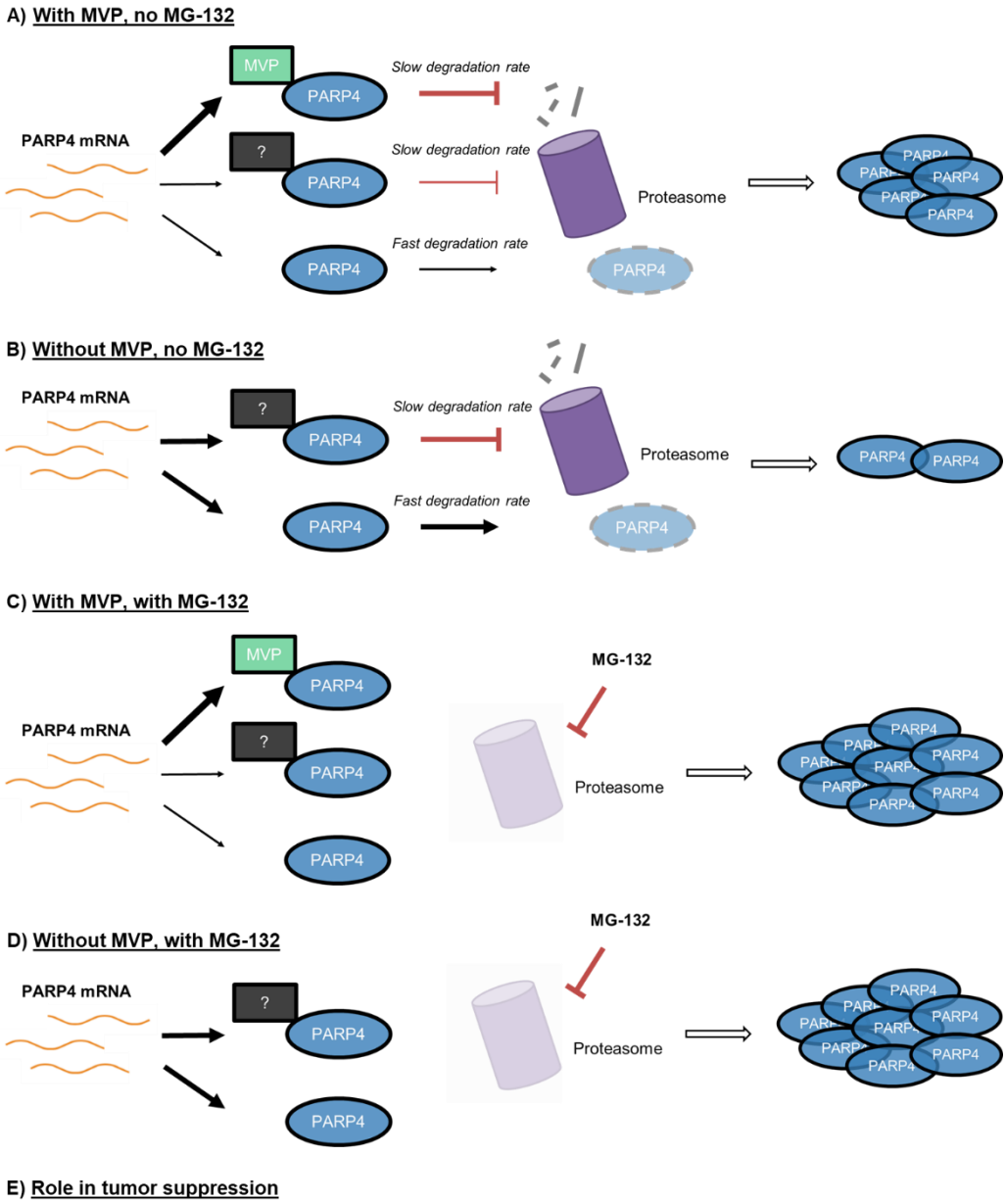


Figure 3-22: Model of the relationship between MVP and PARP4, as well as the effects on tumorigenicity

A) In the presence of MVP, most of PARP4 in the cell is bound to MVP. In its MVP-bound state, PARP4 is more resistant to proteasomal degradation. In contrast, the free form of PARP4 is more rapidly degraded. There is another fraction of PARP4 that is possibly stabilized by binding of other interaction partners. **B)** In the absence of MVP, PARP4 mRNA levels remain unchanged, and translated PARP4 protein is likely redistributed to the free fraction as well as that bound by other interaction partners. The relative distribution between the two fractions, however, is unknown. There is now a greater proportion of translated PARP4 protein in its free form which is rapidly degraded, leading to the observed reduction in PARP4 protein. At the same time, there remains a fraction of PARP4 protein that is more resistant to degradation, possibly protected by other binding partners. **C)** With or **D)** without MVP, all forms of PARP4 protein are stabilized following inhibition of proteasomal degradation by MG-132, thus accounting for the higher levels of PARP4 protein observed. **E)** As MVP loss does not increase tumorigenicity, the MVP-bound form of PARP4 is likely not important for tumor suppression. PARP4 in its free form, however, is rapidly degraded, so it is likely the third fraction of PARP4 that is functionally important for tumor suppression. Whether this fraction resides in the nucleus or cytoplasm remains to be determined.

3.7 hnRNPM is a novel PARP4-binding partner with tumor suppressive activity in LUAD

3.7.1 Identification of hnRNPM as a novel PARP4 binding partner

Having established that PARP4's tumor suppressive role appears to be independent of the vault complex, I then wanted to determine what other proteins PARP4 could be interacting with to mediate its anti-tumorigenic properties. I thus sought to identify PARP4's endogenous binding partners by performing stable isotope labelling by amino acids in cell culture (SILAC) co-immunoprecipitation (co-IP) mass spectrometry.

The SILAC method was chosen because it was more quantitative and specific than regular mass spectrometry (Emmott and Goodfellow, 2014; Wal and Demmers, 2015). Specifically, PARP4 was immunoprecipitated in SAECK shControl cells, with shPARP4 #1 cells serving as a negative control to improve the specificity of the hits. In the SILAC set up (Figure 3-23), IP for the two cell line conditions was each performed using two independent cell cultures, one labelled by amino acids with heavier isotopes (^{15}N- and ^{13}C-labelled lysine and arginine), termed 'heavy', and one with non-labelled amino acids, termed 'light'. The 'heavy' shControl eluate was mixed with the 'light' shPARP4 eluate to yield a 'forward' sample, whereas the 'light' shControl eluate was mixed with the 'heavy' shPARP4 eluate to yield a 'reverse' sample. In each sample, the mass spectrometry analysis would report a SILAC 'heavy' to 'light' (H/L) ratio corresponding to the proportion of 'heavy' peptides to 'light' peptides detected for each protein. Specific PARP4 interaction partners would hence be indicated by a forward SILAC H/L ratio > 1 and reverse SILAC H/L ratio < 1.

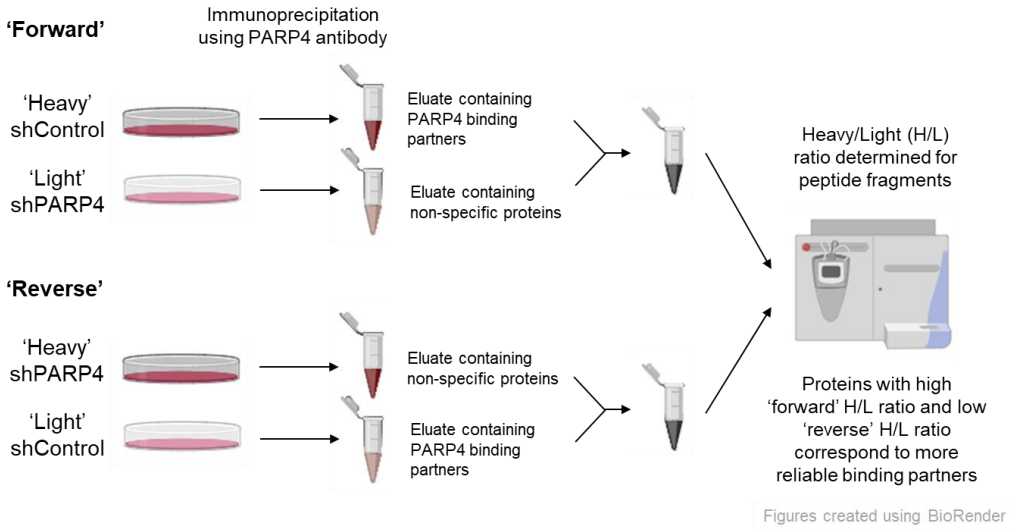


Figure 3-23: Schematic diagram for SILAC co-IP mass spectrometry experiment

After verifying that the immunoprecipitation conditions were indeed able to successfully enrich for PARP4 (Figure 3-24), the 'forward' and 'reverse' SILAC samples were sent for mass spectrometry analysis at the Quantitative Proteomics Core at the Cancer Science Institute (CSI).

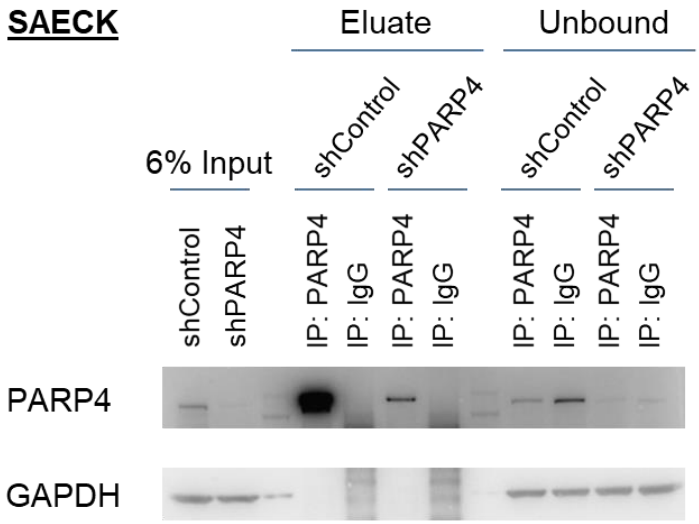


Figure 3-24: Successful enrichment of PARP4 via immunoprecipitation

Immunoblot analysis of PARP4 following immunoprecipitation using PARP4 or IgG control antibody in SAECK shControl and shPARP4 cells. GAPDH was used as a loading control for the input fractions.

The results revealed the bait protein PARP4, as well as MVP, as the most highly enriched proteins with the highest forward ratios alongside the lowest reverse ratios (Figure 3-25). The detection of the well-known interacting partner of PARP4, MVP, indicated that the SILAC co-IP mass spectrometry experiment had indeed worked. Apart from PARP4 and MVP, the remaining proteins had weaker enrichment ratios, and a more lenient cut off (a forward ratio greater than 1.1 and a reverse ratio smaller than 0.9) was chosen to shortlist potential hits (Table 3-5).

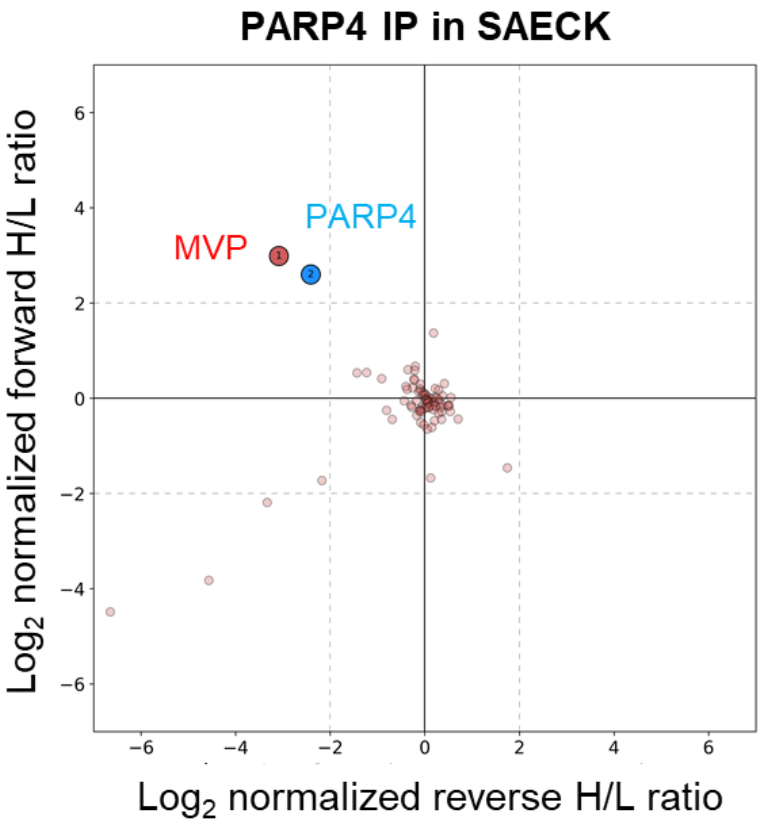


Figure 3-25: SILAC co-IP mass spectrometry analysis of PARP4 in SAECK cells

Plot showing the normalized forward and reverse SILAC H/L ratios of proteins detected in the PARP4 SILAC co-IP mass spectrometry experiment. Hits with forward H/L ratio >1 and reverse H/L ratio < 1 are found in the top left quadrant.

Table 3-5: List of potential hits with normalized forward ratio > 1.1 and normalized reverse ratio < 0.9.

Gene Name	Description	Normalized Forward H/L Ratio	Normalized Reverse H/L Ratio
MVP	Major vault protein	7.95	0.12
PARP4	Poly (ADP-ribose) polymerase 4	6.06	0.19
PHB2	Prohibitin-2	1.60	0.87
PHB	Prohibitin	1.52	0.78
KRT18	Keratin 18	1.50	0.86
PYGB	Glycogen phosphorylase, brain form	1.46	0.43
VIM	Vimentin	1.45	0.37
RPS27A	Ubiquitin-40S ribosomal protein S27a; Ubiquitin; 40S ribosomal protein S27a	1.33	0.53
UBC	Polyubiquitin-C; Ubiquitin		
UBB	Polyubiquitin-B; Ubiquitin		
UBA52	Ubiquitin-60S ribosomal protein L40; Ubiquitin; 60S ribosomal protein L40		
PCNT	Pericentrin	1.33	0.85
DDX3X	ATP-dependent RNA helicase DDX3X	1.30	0.86
DDX3Y	ATP-dependent RNA helicase DDX3Y		
PCM1	Pericentriolar material 1 protein	1.19	0.76
ACBD5	Acyl-CoA-binding domain-containing protein 5	1.17	0.84
HNRNPM	Heterogeneous nuclear ribonucleoprotein M	1.14	0.77

To validate the hits, I performed a targeted immunoprecipitation of PARP4 in SAECK lysates, and observed the co-immunoprecipitation of not only MVP but also heterogeneous nuclear ribonucleoprotein M (hnRNPM) and prohibitin (PHB) (Figure 3-26A). PARP4 could similarly be co-immunoprecipitated upon the reverse experiment of either hnRNPM or PHB immunoprecipitation (Figure 3-26B).

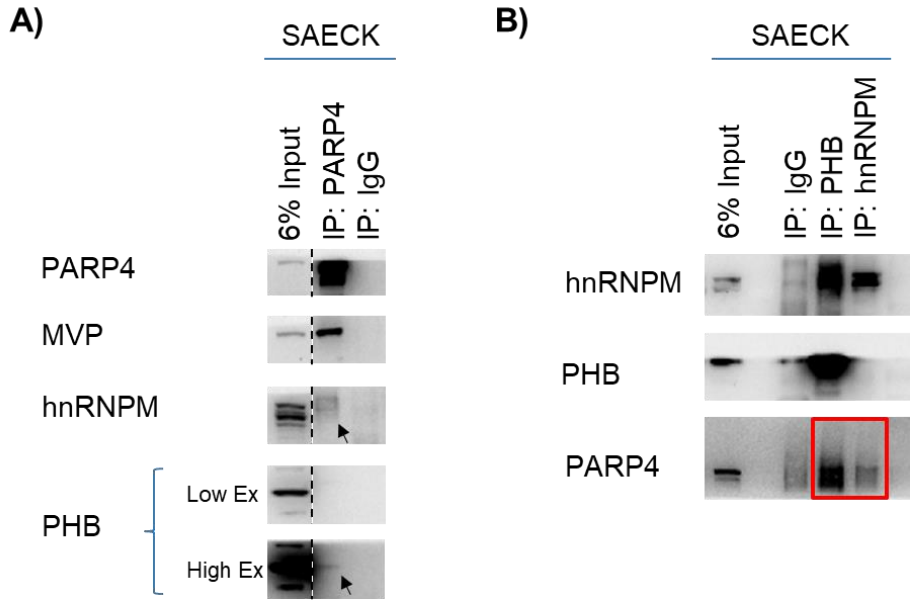


Figure 3-26: Validation of PARP4 interaction partners identified from SILAC co-IP mass spectrometry

A) Immunoblot analysis following immunoprecipitation of PARP4 in SAECK cells. PARP4 and MVP were enriched while a fraction of hnRNPM and PHB (indicated by black arrows) was found in association with PARP4. **B)** Immunoblot analysis following immunoprecipitation of PHB and hnRNPM in SAECK cells. PARP4 was successfully co-immunoprecipitated (indicated by red box).

In addition to these, I was also able to validate some of the other putative PARP4 binding partners from the mass spectrometry data. Under immunoprecipitation conditions that successfully enriched for pericentrin (PCNT), prohibitin-2 (PHB2), pericentriolar material 1 protein (PCM1), keratin 18 (KRT18) and vimentin (VIM) (Figure 3-27, red boxes), co-immunoprecipitation of PARP4 was observed for all the above-mentioned proteins except for PHB2 (Figure 3-27, yellow boxes).

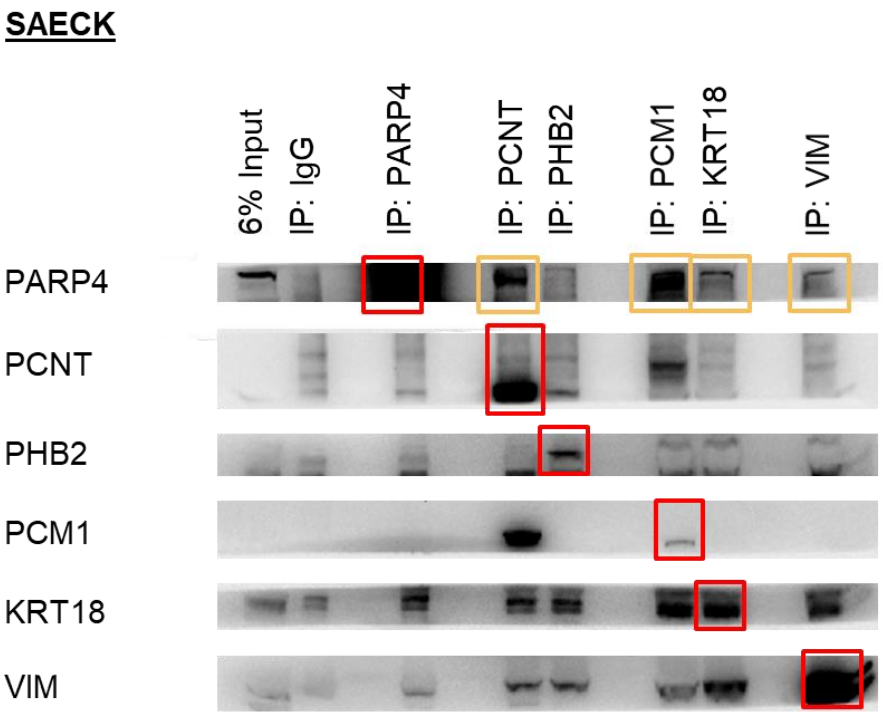


Figure 3-27: Validation of additional PARP4 interaction partners identified from SILAC co-IP mass spectrometry

Immunoblot analysis following immunoprecipitation of PARP4 and prospective interaction partners identified as hits from the SILAC co-IP mass spectrometry experiment. Bands indicating successful immunoprecipitation of the target proteins by their respective antibody relative to IgG control are highlighted by red boxes. Bands indicating co-immunoprecipitation of PARP4 relative to the IgG control are highlighted by yellow boxes.

As the interaction between PARP4 and hnRNPM was most consistent, with hnRNPM successfully co-immunoprecipitating with PARP4 in multiple lung cell lines including A549 and PC-9 (Figure 3-28), I decided to follow up on hnRNPM. Notably, the interaction was preserved in normal lung lines (Figure 3-28, right panel), indicating that this was not cancer-specific.

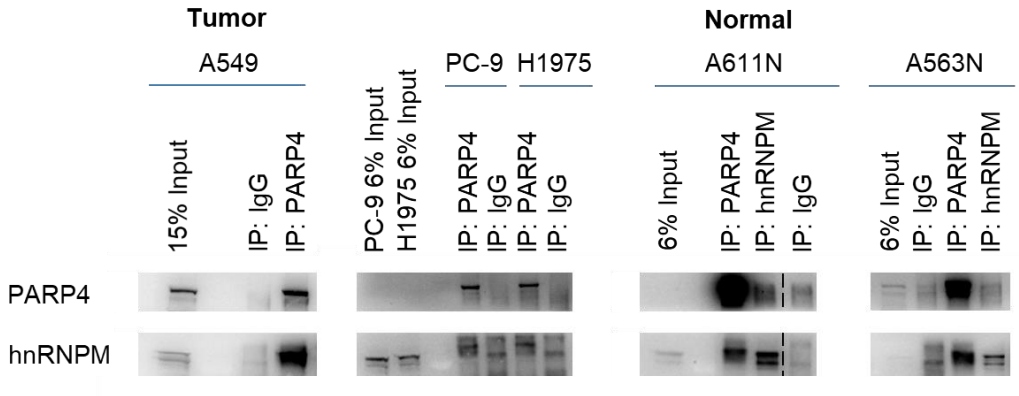


Figure 3-28: Immunoprecipitation of PARP4 and hnRNPM in various lung cell lines

Immunoblot analysis following immunoprecipitation of PARP4 or hnRNPM in the lung cancer cell lines A549, PC-9 and H1975, as well as the normal lung cell lines A611N and A563N.

hnRNPM is a predominantly nuclear protein, although it has been reported to shuttle to and from the cytoplasm (Chen et al., 2019; Geuens et al., 2016; Llères et al., 2010; Marko et al., 2010). From earlier fractionation studies (Figure 3-21F), I observed that while the majority of PARP4 localizes to the cytoplasm, a smaller subset does exist in the nucleus in SAECK cells. To determine whether PARP4's interaction with hnRNPM is nuclear or cytoplasmic, I set out to first isolate cytoplasmic and nuclear fractions from SAECK cells before immunoprecipitating PARP4. Interesting, co-immunoprecipitation of hnRNPM was detected in the nuclear but not the cytoplasmic fraction (Figure 3-29A, red box), suggesting that PARP4 and hnRNPM's interaction occurs predominantly in the nuclear compartment in SAECK cells.

To assess if this was also true in normal lung cells or cancer cell-specific, I performed fractionation on wildtype A653N cells followed by immunoprecipitation of PARP4. Unexpectedly, the nuclear fraction of A653N had more PARP4 than the cytoplasmic fraction (Figure 3-29B). hnRNPM was also successfully co-immunoprecipitated alongside PARP4 in both fractions (Figure 3-29B, red boxes), suggesting that in normal A653N lung cells, the interaction between PARP4 and hnRNPM could occur in both compartments.

These experiments would have to be performed in additional normal and lung cancer cell lines to determine if these findings are more widely relevant. If true, these may suggest additional functionality of PARP4 and hnRNPM's interaction that is present only in normal cells but lost in cancer.

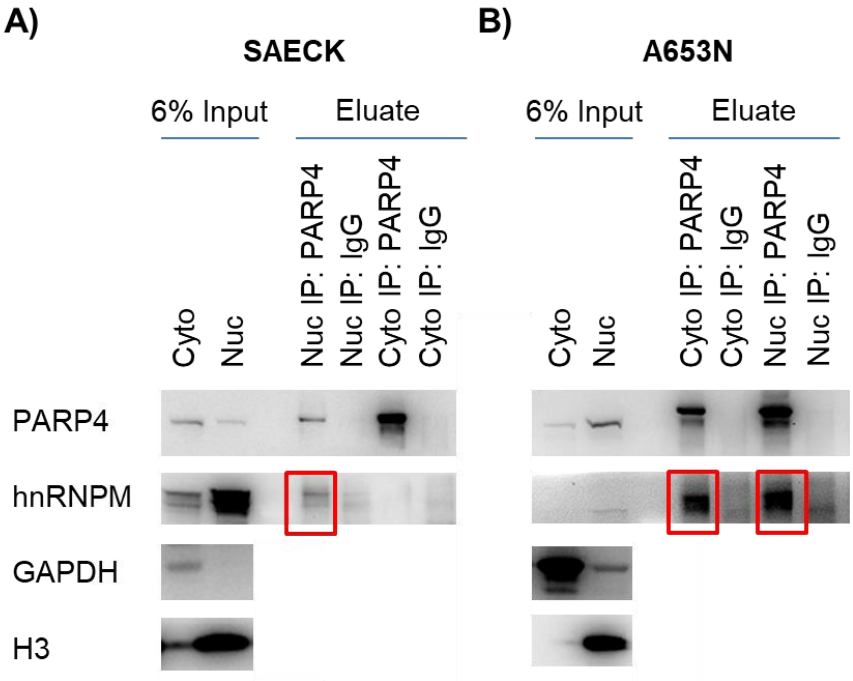


Figure 3-29: PARP4 immunoprecipitation in cytoplasmic and nuclear fractions of SAECK cancer cells and A653N normal cells

A, B) Immunoblot analysis following immunoprecipitation of PARP4 from the cytoplasmic (cyto) and nuclear (nuc) fractions of **A)** SAECK cells and **B)** normal A653N lung cells. GAPDH was used as a cytoplasmic marker while total histone H3 was used as a nuclear marker. Successful co-immunoprecipitation of hnRNPM by PARP4 is highlighted in the red boxes.

Furthermore, given that PARP4 and hnRNPM's interaction in SAECK cells occurs mainly in the nuclear compartment, this could be one reason accounting for the weak SILAC ratios observed for hnRNPM (Table 3-5) even though it was the most reliably validated hit. This was because our initial SILAC co-IP mass spectometry experiment had been performed on SAECK whole cell lysates rather than specifically on the nuclear fraction. In future, co-IP mass

spectrometry experiments performed specifically on the cytoplasmic and nuclear compartments might unveil yet additional PARP4 binding partners.

3.7.2 hnRNPM loss increases tumorigenicity of lung cancer cells

Having established hnRNPM as a novel PARP4 binding partner, I next wanted to understand whether, and how, it might contribute to PARP4's tumor suppressive role.

hnRNPM belongs to the family of heterogeneous ribonucleoproteins (hnRNPs), which comprises RNA-binding proteins integral to various stages of RNA metabolism, such as in regulating splicing and mRNA stability (Geuens et al., 2016; Han et al., 2010).

In the context of cancer, hnRNPM has been reported to drive splicing programs (Figure 3-30) that promote the epithelial to mesenchymal transition (EMT) to support metastasis in breast and gastric cancer (Wang et al., 2021; Xu et al., 2014). Furthermore, hnRNPM-regulated splicing events that were triggered in response to phosphatidylinositol 3-kinase (PI3K) inhibition served to limit the efficacy of chemotherapeutic agents in Ewing sarcoma (Palombo et al., 2020; Passacantilli et al., 2017). hnRNPM was also recently found to prevent aberrant splicing of key genes required for prostate cancer cell growth (Ho et al., 2021). In addition to its splicing roles, hnRNPM was observed to be overexpressed in a panel of colorectal cancer cell lines and was thus proposed as a potential biomarker of colorectal carcinoma (Chen et al., 2014). Under conditions of hypoxic stress leading to suppression of global protein production, hnRNPM was reported to localize to the cytoplasm and bind to internal ribosome entry site (IRES)-containing transcripts associated with functions promoting adaptation to hypoxia, thereby enhancing their translation (Chen et al., 2019). There was one report of a contradictory role in prostate cancer where hnRNPM was found to negatively regulate cancer cell motility and invasiveness by reducing expression of a key EMT transcription factor, Twist1 (Yang et al., 2019). Consistent with

this, the same study found that hnRNPM expression was lost in more aggressive prostate cancer cell lines. All in all, these studies suggest that more often than not, hnRNPM may have tumor-promoting properties in several cancer types. However there have not been any studies on hnRNPM in the context of lung cancer.

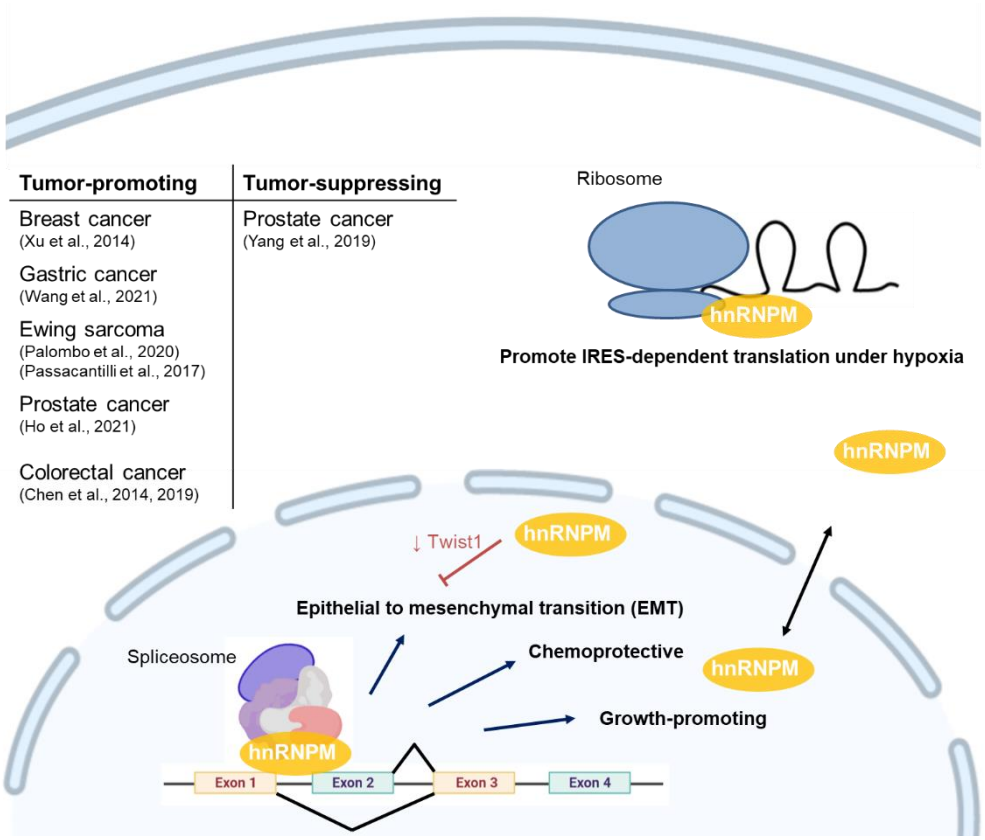


Figure 3-30: Schematic diagram describing the roles of hnRNPM in cancer

hnRNPM has been reported to drive splicing programs that promote epithelial to mesenchymal transition (EMT), reduce the efficacy of chemotherapy and promote cell growth. hnRNPM can shuttle between the nucleus and the cytoplasm, where it selects mRNAs for IRES-dependent translation under conditions of hypoxic stress. However, there was also a report of hnRNPM inhibiting EMT. A summary of studies reporting hnRNPM's role in different cancer types is found in the table. Figure was illustrated using Biorender.

To investigate the potential role of hnRNPM in LUAD, I began by determining whether hnRNPM expression levels were correlated with LUAD patient survival.

Using the same set of LUAD microarray and survival data from the PARP4 and RASA1 comparisons earlier (Győrffy et al., 2013), I found that contrary to the tumor-promoting roles associated with hnRNPM in other cancers, higher hnRNPM expression was correlated with better survival in LUAD patients (Figure 3-31A). This points to a potential tumor suppressive role of hnRNPM in the LUAD context.

Similar to PARP4, hnRNPM copy number loss (53.0%, n=122) was also observed in the majority of the TCGA LUAD cohort (n=230) (Figure 3-31B). In the case of Asian LUAD (n=302), most cases were still hnRNPM copy number diploid (49.3%, n=149), although there were many more instances of copy number loss (41.4%, n=125) than copy number gain (9.3%, n=28) (Figure 3-31C). Unlike PARP4 however, hnRNPM levels were not significantly different between lung tumor and normal groups (Figure 3-31D). Thus, hnRNPM could possibly have a tumor suppressive role in LUAD that would need to be further determined.

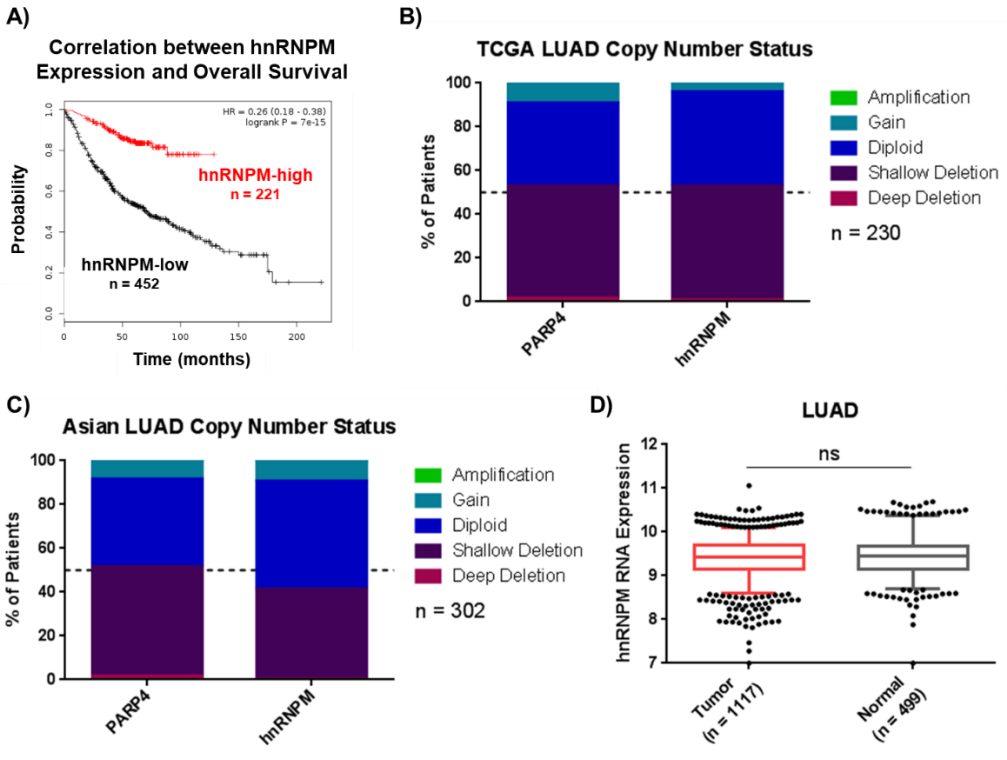


Figure 3-31: Data mining to examine the potential role of hnRNPM in LUAD

A) Kaplan-Meier plot generated using LUAD microarray data (n=720, Affymetrix ID 1555844_s_at for hnRNPM) from the KM Plotter database (Győrffy et al., 2013). The patient group with high hnRNPM gene expression (n=221) is displayed in red while the low expression group (n=452) is marked in black. **B, C)** Bar graphs showing the distribution of PARP4 and hnRNPM copy number status within the **B)** TCGA LUAD cohort (n=230) and the **C)** Asian LUAD cohort (n=302). **D)** Box-and-whisker plot comparing hnRNPM RNA levels between LUAD tumor (n=1117) versus normal tissue (n=499). The 25^{th}, 50^{th} and 75^{th} percentiles are indicated by the boxes, while the whiskers extend to the 5^{th} and 95^{th} percentiles. Beyond, individual points are displayed. ns: not significant, as determined by unpaired two-tailed t-test. Data was retrieved from cBio portal (Cerami et al., 2012; Gao et al., 2013) **(B)**, OncoSG **(C)** and the GENT2 database (Park et al., 2019) **(D)**.

To conclusively determine whether hnRNPM had a tumor suppressive role in lung cancer cells that recapitulated what was observed with PARP4, I depleted hnRNPM in both SAECK and A549 cells using the same shRNAs (Figure 3-32A, D) and characterized their tumorigenicity. Whereas there was once again no significant difference in cell proliferation rates upon hnRNPM loss (Figure 3-32B), hnRNPM knockdown cells formed significantly larger tumors upon subcutaneous implantation into the flanks of immunocompromised NSG mice (Figure 3-32C, E, F). Clearly, loss of hnRNPM was able to recapitulate the phenotype of PARP4 loss *in vivo*. I thus hypothesized that PARP4 and hnRNPM could be acting along the same pathway to mediate their tumor suppressive activities.

Interestingly, loss of hnRNPM did not result in detectable changes to PARP4 protein levels (Figure 3-32A, D), and the converse was also true (Figure 3-32G). I thus concluded that PARP4 and hnRNPM do not regulate each other's protein expression levels.

Notably, post-translational modification of hnRNPM had previously been reported to modify its activity. Specifically, hnRNPM was found to repress expression of a group of innate immune transcripts in mouse macrophages through splicing regulation. Upon pathogenic infection, phosphorylation of hnRNPM at specific serine residues could relieve splicing repression of one of the immune transcripts, IL6 (West et al., 2019). Since I had earlier shown that

the PARP catalytic domain was important for PARP4's tumor suppressive properties, one hypothesis for the relationship between PARP4 and hnRNPM could be that PARP4 ADP-ribosylates hnRNPM to promote its tumor-suppressive activity. In another possibility, PARP4's interaction with hnRNPM could be required and sufficient for altering hnRNPM activity in a manner that affects tumorigenicity.

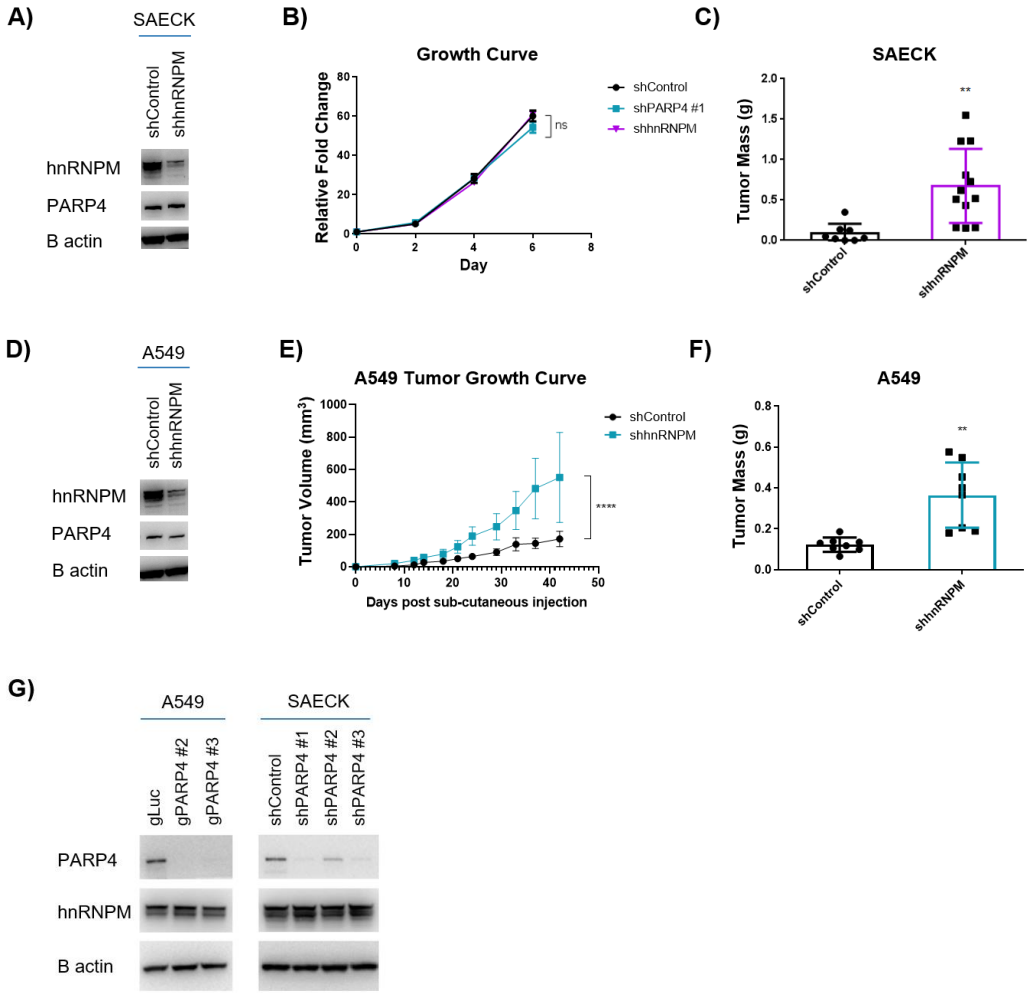


Figure 3-32: hnRNPM knockdown increases SAECK and A549 tumorigenicity

A) Immunoblot analysis of hnRNPM and PARP4 levels in SAECK shControl and shhnRNPM cells. **B)** Growth curves comparing proliferative capacity of SAECK shControl, shPARP4 #1 and shhnRNPM cells as measured by the Cell-Titer Glo assay. Data represent the mean ± s.e.m., n = 3. **C)** Bar graph comparing mass of tumors formed from the subcutaneous injection of 1 million SAECK shControl or shhnRNPM cells per flank of NSG mice after 10 weeks. Data represent the mean ± s.d., n ≥ 8. **D)** Immunoblot

analysis of hnRNPM and PARP4 levels in A549 shControl and shhnRNPM cells. **E)** Growth curve of tumors formed from the subcutaneous injection of 400,000 A549 shControl or shhnRNPM cells per flank of NSG mice. Tumor volume was calculated as 0.5 x length x width^2. **F)** Bar graph comparing mass of tumors harvested from A549 shhnRNPM versus shControl after 7 weeks. Data represent the mean ± s.d., n = 8; **p<0.01, ****p<0.0001, ns: not significant, as determined by unpaired two-tailed t-test. **G)** Immunoblot analysis of PARP4 and hnRNPM levels in control and PARP4-depleted A549 and SAECK cells.

3.7.3 hnRNPM regulates splicing in the lung cancer context

Having demonstrated that PARP4's novel interaction partner, hnRNPM, similarly has a tumor suppressive role in the lung cancer context, I then wanted to examine the underlying mechanisms.

As hnRNPM is known first and foremost as a splicing regulatory factor, with a majority of studies focused on its RNA-binding properties and effects on splicing outcomes (Datar et al., 1993; Ho et al., 2021; Huelga et al., 2012; Venables et al., 2008; Wagner et al., 2021), I sought to examine how hnRNPM's regulation of alternative splicing could contribute to tumor suppression in our lung cancer cell line model.

Splicing programs have been increasingly recognized to be altered in cancer in a number of ways. These include mutations to splicing regulatory *cis* elements of cancer-associated genes (Jung et al., 2015; Supek et al., 2014), as well as mutations or expression level changes to core and auxiliary splicing factors (Anczuków et al., 2012; Graubert et al., 2011; Jensen et al., 2014; Karni et al., 2007; Martin et al., 2013; Yoshida et al., 2011).

Collectively, these splicing alterations affect all aspects of tumor biology, and have been associated with each of the different hallmarks of cancer (Bonnal et al., 2020; Ghigna et al., 2008; Hanahan and Weinberg, 2011; Venables, 2004). For example, alternative splicing of pyruvate kinase (PKM) leads to the production of a PKM2 isoform that promotes aerobic glycolysis in ovarian and gastric cancer (Chao et al., 2017; Shiroki et al., 2017), whereas enhanced expression of the anti-apoptotic Bcl-xL isoform promoted cancer cell survival and malignancy in breast and liver cancers (Olopade et al., 1997; Takehara et al., 2001).

As a splicing factor, hnRNPM has been shown to be aberrantly expressed in several cancers, with hnRNPM-mediated splicing events contributing to tumor-promoting phenotypes such as EMT in breast, gastric and prostate cancer (Chen et al., 2014; Ho et al., 2021; Wang et al., 2021; Xu et al., 2014). Notably,

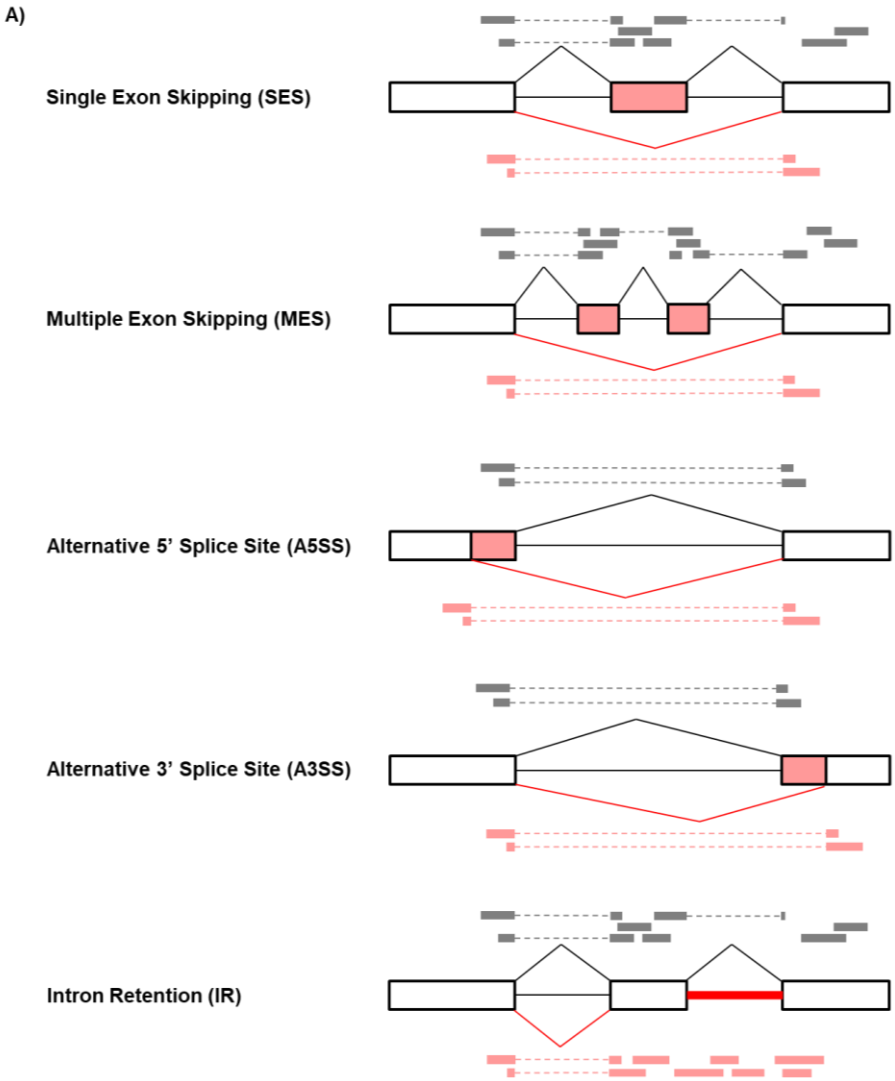
hnRNPM regulates splicing of EMT-associated transcripts in an antagonistic fashion as compared to the epithelial splicing factors ESRP1 and AKAP8, with which hnRNPM shares a subset of regulated splice sites and competes for binding (Harvey et al., 2018; Hu et al., 2020). Additionally, hnRNPM was found to regulate splicing events that could limit the efficacy of chemotherapeutic agents, thereby having a protective effect in Ewing sarcoma (Palombo et al., 2020; Passacantilli et al., 2017).

However, hnRNPM's role in regulating splicing in the lung cancer context has not been studied. Thus, I wanted to conduct a global profiling of splicing changes upon loss of hnRNPM in SAECK cells.

Firstly, direct Nanopore sequencing was performed on cDNA from SAECK shControl and shhnRNPM cell lines. In Nanopore direct cDNA sequencing, each full-length cDNA strand is fed through a protein pore. As each individual nucleotide passes through the pore, an electrical signal characteristic of the base identity is generated and recorded in a read, with each read corresponding to the entire cDNA sequence (Branton et al., 2008; Deamer and Branton, 2002). This long-read sequencing approach was specifically chosen as it allows the entire transcript to be sequenced, thus providing exon connectivity information to aid transcript identification (Kono and Arakawa, 2019). Secondly, the omission of any PCR amplification step through direct cDNA sequencing removes potential amplification bias that might favor and thus overrepresent particular transcripts (Amarasinghe et al., 2020; Ku and Roukos, 2013). Finally, the use of singleplex sequencing enables high sample read depths to be achieved, which in turn would allow rare splice events to be picked up and improve the reliability of the detection of alternative transcript usage and splice events.

Thereafter, the sequencing data was analyzed by our collaborators Dr. Zhang Bin (CSI) and Dr. Yvonne Tay (CSI) using the PSI-Sigma pipeline, which was selected for its capability in detecting and quantitating alternative splicing events using long-read sequencing data (Lin and Krainer, 2019).

Using this pipeline, we were able to study five alternative splicing event types: (1) single exon skipping (SES), (2) multiple exon skipping (MES), (3) intron retention (IR) and use of either (4) an alternative 5' splice site (A5SS) or (5) an alternative 3' splice site (A3SS) at a particular exon (Figure 3-33A). For each splice event, the pipeline computes the Percent Spliced In (PSI) value, which represents the percentage of sequencing reads supporting the inclusion of the splice feature (Figure 3-33B). A more detailed example is described in Figure 3-33B. Finally, the PSI value for the control sample was subtracted from that of the hnRNPM knockdown sample to yield the delta PSI value (ΔPSI), which represents the change in inclusion of a splice feature upon hnRNPM loss.



$$Percent\ Spliced\ In\ (PSI) = \frac{No.\ of\ Supporting\ Reads\ for\ an\ Event}{Total\ No.\ of\ Reads} \times 100\%$$

$$PSI\ for\ SES = \frac{No.\ of\ Inclusion\ Reads\ (grey)}{No.\ of\ Inclusion\ Reads\ (grey) + No.\ of\ Exclusion\ Reads\ (pink)} \times 100\%$$

Figure 3-33: Splicing analysis by the PSI-Sigma pipeline

A) Schematic diagram of each of the five splice event types recognized by the PSI-Sigma pipeline. Exons represented by the boxes are connected by lines representing the introns. The alternatively spliced feature is highlighted in pink or red. Examples of sequencing reads that support the constitutive event are represented in grey, while reads that support the alternative event are represented in pink. A splice event must be represented by a minimum of five supporting sequencing reads before it is successfully detected by the pipeline. **B)** Calculation of the Percent Spliced In (PSI) value. For each splice event, the pipeline computes the PSI value, which represents the percentage of sequencing reads

supporting a particular splice feature out of total reads at the event region. As an example, the PSI value for a variable exon (in pink) for an SES event in **A)** would be determined by the percentage of sequencing reads supporting the inclusion of that exon (grey reads), out of the total number of reads for the splice region that either include (grey reads) or exclude the exon (pink reads).

From our analysis, thousands of splice events across each of the five event categories were detected in control and hnRNPM knockdown SAECK cells (Figure 3-34A). In both cell lines, the greatest number of events belonged to the A3SS category with around 4000 unique events detected, while the least number were MES events with around 1000 unique events detected. Strikingly, there were 17% more unique intron retention events detected in the hnRNPM knockdown cells compared to control (Figure 3-34A, red box), whereas the other categories of splice events had similar unique event counts between the two conditions. Furthermore, intron retention events were also the most dysregulated class of splicing events identified upon hnRNPM loss (Figure 3-34B, red box), with the majority of these being significantly upregulated ($\Delta PSI > 10$). This suggests that in the SAECK lung cancer cells, hnRNPM tends to promote intron removal from many genes.

Apart from intron retention events, single exon skipping events were the next most dysregulated class of events upon hnRNPM knockdown (Figure 3-34B, blue box). The numbers of up- ($\Delta PSI < -10$) and down-regulated ($\Delta PSI > 10$) single exon skipping events were even – 66 and 61 respectively, suggesting that hnRNPM promotes both exon inclusion and exclusion events in SAECK cells. This finding supports what is more recently known that hnRNPM mediates exon inclusion as well as exclusion, as opposed to the classical notion of hnRNPs being solely splicing repressors (Ho et al., 2021; Huelga et al., 2012; Xu et al., 2014).

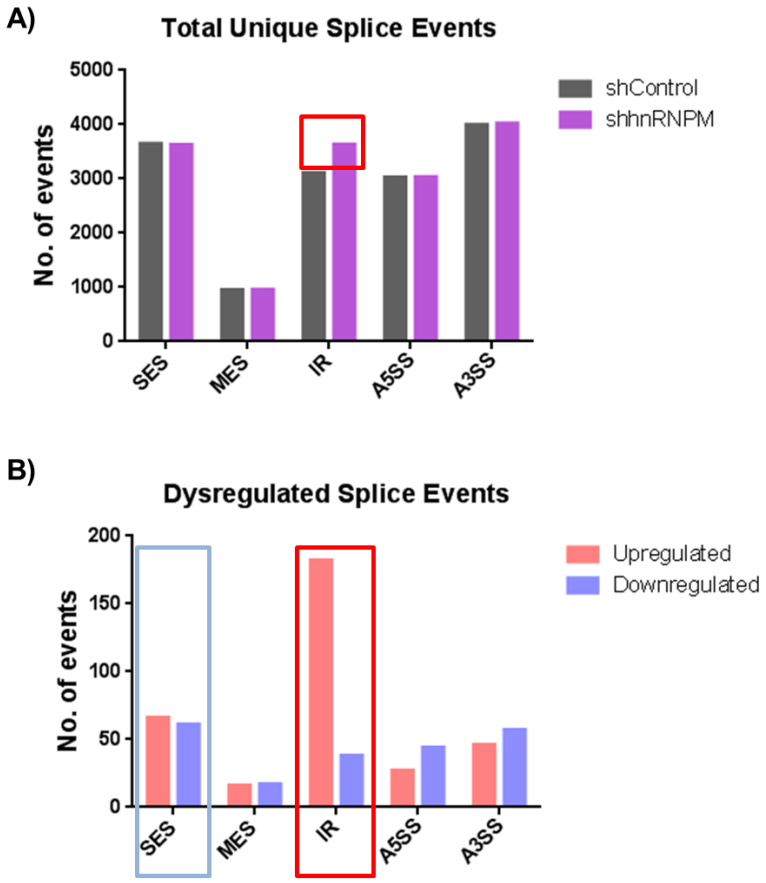


Figure 3-34: Splice events identified by PSI-Sigma analysis of SAECK shControl and shhnRNPM Nanopore sequencing data

A) Number of unique splice events detected within the SAECK shControl and shhnRNPM samples across each of the five event categories. A splice event must be represented by a minimum of five supporting sequencing reads before it is successfully detected by the pipeline. **B)** Number of significantly upregulated or downregulated splice events across the five event categories upon hnRNPM loss. An event was taken to be significantly dysregulated if the absolute ΔPSI value (|ΔPSI|) was greater than 10 and p-value was less than 0.05.

To get a broad indication of the splicing programs mediated by hnRNPM that were affected in the hnRNPM knockdown cells, the entire list of genes with significantly dysregulated splicing events was analyzed using the Enrichr platform (Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021). Among the GO Biological Process 2021 gene sets, I observed enrichment of RNA processing-related genes (Figure 3-35A, darkened), which was expected given that hnRNPM is known to regulate its own splicing and that of other RNA binding and splicing-related proteins (Huelga et al., 2012).

In order to directly determine whether hnRNPM regulates the splicing of known cancer-associated genes, I assessed the degree of overlap between genes with splicing regulated by hnRNPM and cancer genes defined by the Catalogue Of Somatic Mutations In Cancer (COSMIC) database (Futreal et al., 2004). Of the 729 COSMIC-annotated cancer genes, the splicing of 22 genes was controlled by hnRNPM, and this corresponded to 5% of all alternatively spliced genes upon hnRNPM loss (Figure 3-35B). While the functional significance of these events needs to be further explored, this comparison tells us that hnRNPM could contribute to the regulation of a subset of cancer genes through alternative splicing.

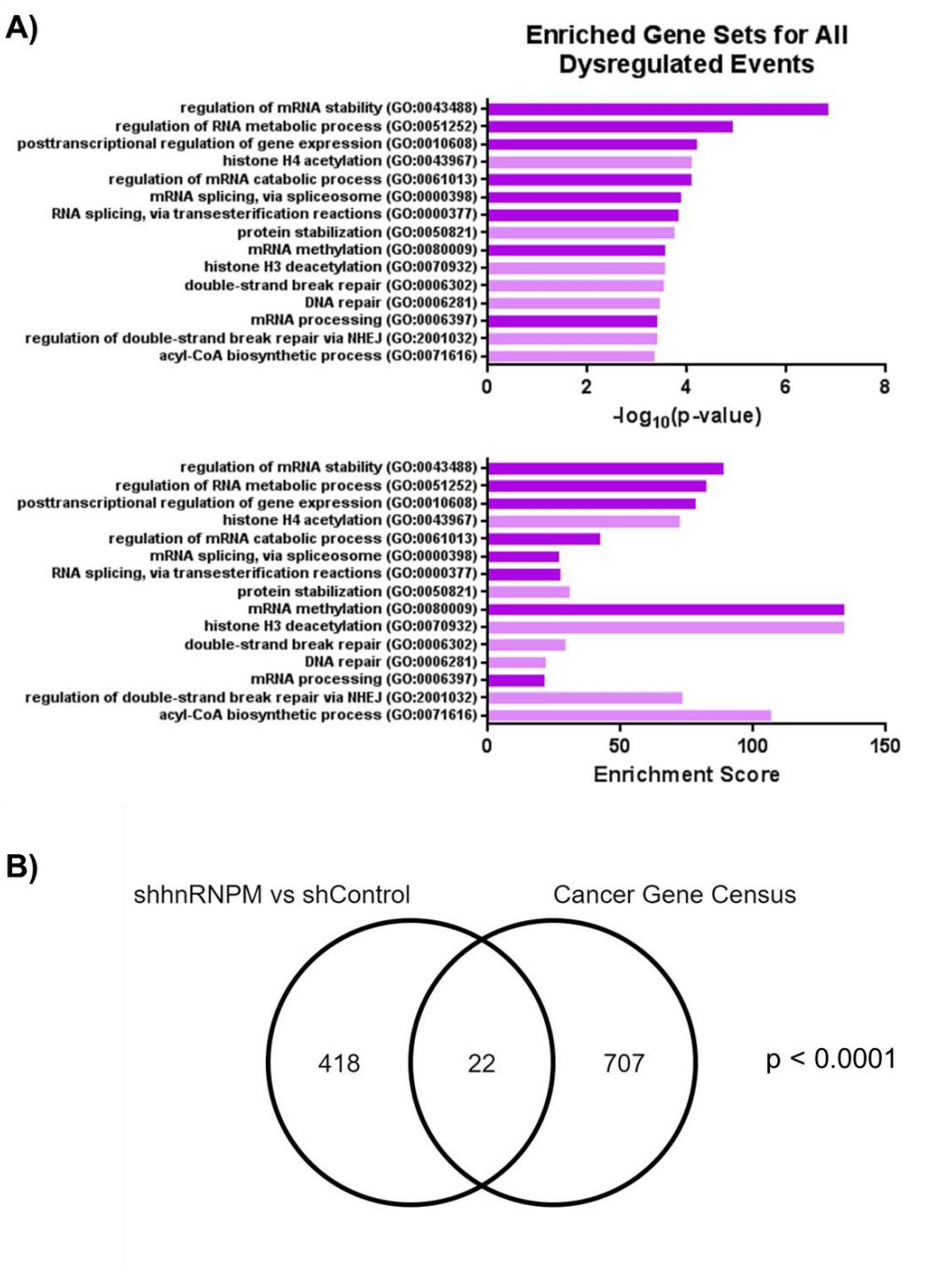


Figure 3-35: Analysis of genes with dysregulated splicing upon hnRNPM knockdown

A) Top 15 enriched GO Biological Process 2021 gene sets for all genes with significantly dysregulated splicing upon hnRNPM knockdown ($|\Delta PSI| > 10$, p-value < 0.05). In the top panel, the gene sets are ranked by their p-values. In the bottom panel, the enrichment scores for the same ranking of gene sets are indicated. Gene sets related to RNA metabolism and splicing are highlighted with a darker shade. Analysis was performed

using the Enrichr platform (Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021). **B)** Overlap between genes with splicing regulated by hnRNPM and cancer-related genes defined by COSMIC (Futreal et al., 2004). p denotes the probability that 22 randomly selected genes detected by the PSI-Sigma pipeline (n=7083) would overlap with genes from the Cancer Gene Census (n=729).

I next selected the top dysregulated IR (Table 3-6, Table 3-7) and SES events (Table 3-8, Table 3-9) identified by the PSI-Sigma pipeline for targeted PCR validation. Specifically, primers were designed to flank the target region to amplify alternative forms of the transcript, thereby yielding PCR products with different sizes that could be separated by gel electrophoresis.

Table 3-6: Top 15 upregulated IR events in SAECK shhnRNPM versus shControl

Gene	Target Region (GRCh38 coordinates)	ΔPSI	p-value
SIAH1	chr16:48361760-48361853	47.37	0.028
IP6K2	chr3:48694241-48694701	29.94	0.018
DPY19L4	chr8:94792358-94792730	29.60	0.041
SYVN1	chr11:65128489-65128562	29.53	0.014
IP6K2	chr3:48694240-48695088	29.51	0.011
IP6K2	chr3:48694825-48695089	28.25	0.005
AP5M1	chr14:57288972-57289875	28.05	0.020
MECP2	chrX:154031287-154031395	27.23	0.009
SRRM2	chr16:2770228-2770351	26.08	0.024
PHKG2	chr16:30756960-30760118	25.52	0.008
SHISA5	chr3:48468983-48469022	25.39	0.010
TMEM107	chr17:8174285-8174525	22.16	0.006
ERRFI1	chr1:8012957-8013241	21.48	0.042
CERS2	chr1:150965744-150965831	21.38	0.014
ERRFI1	chr1:8012956-8013357	21.34	0.044

Table 3-7: Top 15 downregulated IR events in SAECK shhnRNPM versus shControl

Gene	Target Region (GRCh38 coordinates)	ΔPSI	p-value
SLC19A1	chr21:45513559-45514895	-33.04	0.002
POT1	chr7:124822552-124823970	-18.22	0.033
ASB1	chr2:238450260-238451784	-17.43	0.009
NGRN	chr15:90265877-90266287	-16.27	0.048
DCTN5	chr16:23672198-23675193	-14.57	0.037
TOR2A	chr9:127732279-127732563	-13.73	0.019
AP003352.1	chr8:98041860-98042653	-13.30	0.001
DYNC1H1	chr14:102032468-102033064	-13.20	0.007
FGF2	chr4:122893226-122897623	-12.44	0.017
IFT27	chr22:36775795-36776030	-12.14	0.030
MIF4GD	chr17:75267631-75268082	-11.43	0.042
ASB1	chr2:238446656-238451784	-11.33	0.029
AP003352.1	chr8:98041859-98042652	-10.94	0.013
ACO2	chr22:41527421-41527900	-10.51	0.030
NEMF	chr14:49782595-49782829	-10.45	0.035

Table 3-8: Top 15 upregulated SES events in SAECK shhnRNPM versus shControl

Gene	Target Region (GRCh38 coordinates)	ΔPSI	p-value
KIAA1191	chr5:176359481-176359567	-37.92	0.006
FIP1L1	chr4:53414615-53414722	-23.32	0.019
ALDOC	chr17:28575965-28576020	-21.07	0.033
GNB1L	chr22:19821228-19821375	-17.81	0.042
SNHG1	chr11:62854888-62854938	-16.84	0.049
CHID1	chr11:893427-893519	-16.56	0.042
YBEY	chr21:46291334-46291462	-15.87	0.004
WASHC2C	chr10:45786612-45786674	-15.57	0.026
CDK7	chr5:69252418-69252451	-15.41	0.027
CA12	chr15:63328098-63328130	-15.05	0.050
AP2B1	chr17:35670857-35670898	-14.10	0.022
RPAIN	chr17:5425971-5426299	-13.98	0.044
C1orf52	chr1:85258935-85259061	-13.84	0.027
SEC24C	chr10:73746630-73746716	-13.56	0.002
CHCHD7	chr8:56212762-56212889	-13.29	0.044

Table 3-9: Top 15 downregulated SES events in SAECK shhnRNPM versus shControl

Gene	Target Region (GRCh38 coordinates)	ΔPSI	p-value
SNHG17	chr20:38422092-38422241	34.27	0.025
POGZ	chr1:151440928-151441086	29.47	0.036
NAP1L4	chr11:2989129-2989271	28.14	0.026
ZMYND8	chr20:47212642-47212725	27.77	0.029
MRPL33	chr2:27774424-27774530	27.49	0.018
APLP2	chr11:130123612-130123779	26.51	0.021
PAK1	chr11:77397027-77397093	25.05	0.001
KIF23	chr15:69425282-69425323	23.61	0.003
HNRNPR	chr1:23340852-23341020	22.60	0.008
NCAPG2	chr7:158650832-158650972	21.36	0.001
RABGEF1	chr7:66768831-66769025	20.60	0.044
TMEM11	chr17:21210936-21211227	20.22	0.049
ANKRD11	chr16:89431256-89431326	16.61	0.007
INTS14	chr15:65607159-65607442	15.94	0.039
HP1BP3	chr1:20780345-20780540	15.50	0.010

Using the targeted PCR assays, I was able to successfully validate the expected changes for a number of dysregulated IR and SES events in the hnRNPM knockdown cells. In the case of IR events, PCR primers (black arrows) were designed to amplify the region flanking the retained intron in red (Figure 3-36A, left). The band intensity of the shorter PCR product corresponding to intron removal was then compared to that of the longer intron-containing product to determine the relative degree of intron retention.

In the case of upregulated IR events, hnRNPM loss led to a significant reduction in the relative lower band intensity of PRPF4B (Figure 3-36A) – a kinase which also has roles in splicing (Corkery et al., 2015), as well as TMEM107 (Figure 3-36B), which is involved in Sonic hedgehog signaling and ciliogenesis (Lambacher et al., 2016).

As for downregulated IR events, loss of hnRNPM resulted in significantly increased lower band intensities for SLC19A1 (Figure 3-36C) – a transporter of folate and the immunotransmitter signaling molecule 2’3’-cyclic GMP-AMP

(cGAMP) (Luteijn et al., 2019; Ritchie et al., 2019), ASB1 (Figure 3-36D), which is implicated in proteasomal degradation and inflammation (Hou et al., 2021; Kohroki et al., 2005), as well as MIF4GD (Figure 3-36E), which has roles in cell cycle regulation by binding and stabilizing the cyclin-dependent kinase (CDK) inhibitor p27 (Wan et al., 2015).

In the case of SES events, hnRNPM knockdown resulted in increased exon skipping in both FIP1L1 (Figure 3-37A), which is one of the 22 overlapping COSMIC genes involved in mRNA polyadenylation (Davis et al., 2018), as well as SEC24C (Figure 3-37B), which is a coat protein on vesicles from the endoplasmic reticulum and involved in the sorting and transport of cargo (Adolf et al., 2016; Siddiqi et al., 2010; Wang et al., 2018a). This can be seen from the increase in relative band intensities of the lower band corresponding to the shorter, exon-skipped PCR product in the hnRNPM knockdown samples (Figure 3-37).

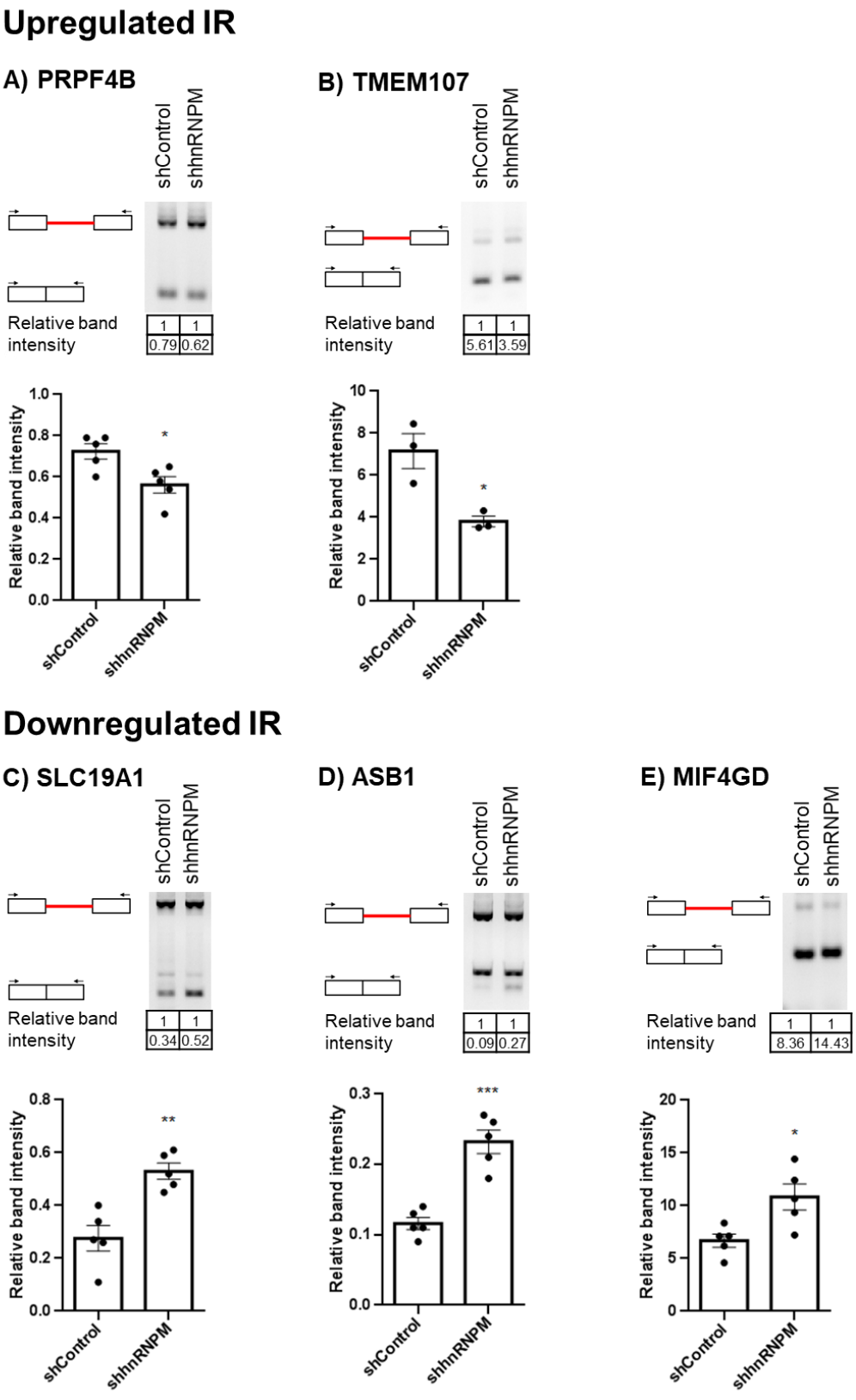


Figure 3-36: PCR validation of dysregulated IR events

Representative gel electrophoresis images of targeted PCR validation of upregulated **(A, B)** and downregulated **(C, D, E)** IR events in SAECK shControl and shhnRNPM samples. Band intensity was quantified, with that of the lower band normalized to that of the upper band, and indicated below the respective lanes. To the left of the respective bands, a schematic diagram corresponding to the intron-retained (upper band) or intron-removed (lower band) product can be found. The red line represents the retained intron while the black arrows represent the PCR primers. The bar graph at the bottom summarizes the results from experimental replicates. Data represent the mean ± s.e.m., n ≥ 3; *p<0.05, **p<0.01, ***p<0.001, as determined by unpaired two-tailed t-test.

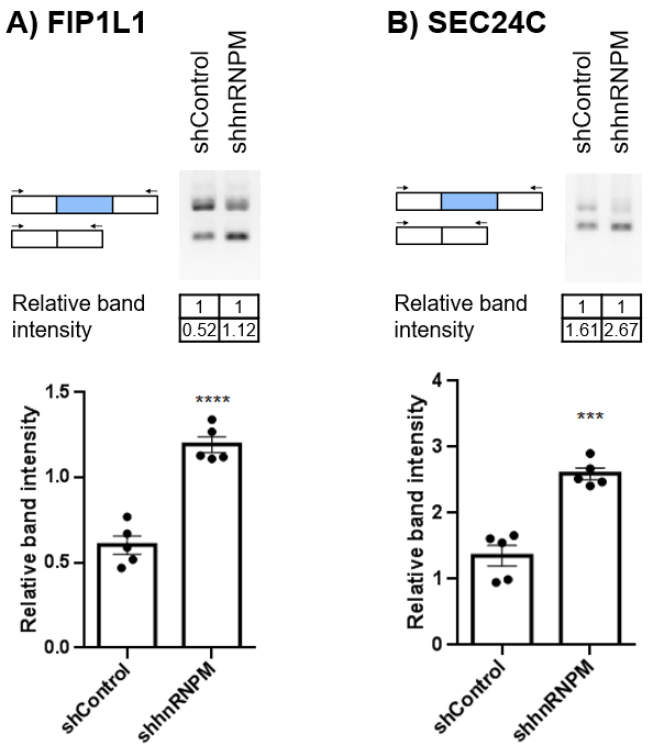


Figure 3-37: PCR validation of upregulated SES events

A, B) Representative gel electrophoresis images of targeted PCR validation of upregulated SES events in SAECK shControl and shhnRNPM samples. Band intensity was quantified, with that of the lower band normalized to that of the upper band, and indicated below the respective lanes. To the left of the respective bands, a schematic diagram corresponding to the exon-included (upper band) or exon-skipped (lower band) product can be found. The blue bar represents the alternatively skipped exon while the black arrows represent the PCR primers. The bar graph at the bottom summarizes the results from experimental replicates. Data represent the mean ± s.e.m., n = 5; *p<0.05, ***p<0.001, ****p<0.0001, as determined by unpaired two-tailed t-test.

While the significance of these validated splicing events has not been previously reported, an analysis of the resultant transcript structures using annotations from the Ensembl genome browser provided some initial clues on possible functional outcomes (Howe et al., 2021). Notably, several of the validated changes in intron retention were predicted to alter the splicing outcome between a functional transcript isoform versus one that has no protein product (Figure 3-38).

For example, the increased intron retention in PRPF4B would likely correspond to a higher expression of the main coding transcript as opposed to two alternative transcripts that were either reported to undergo nonsense-mediated decay (NMD) or did not yield a protein product (Figure 3-38A). In contrast, the increased intron retention in TMEM107 was expected to result in increased levels of two transcripts that encoded a truncated protein or no protein product at all (Figure 3-38B). In the same vein, the downregulated intron retention in both SLC19A1 and ASB1 could correspond to a higher proportion of transcripts that lost part of their 3' untranslated region (3' UTR) and did not encode any protein product (Figure 3-38C, D). Finally, the decreased intron retention in MIF4GD meant that there was a higher expression of the main coding transcript relative to two transcript isoforms that did not form proteins (Figure 3-38E).

In the case of SES events, hnRNPM loss promotes FIP1L1 exon skipping to result in a shortened isoform (Figure 3-38F) that loses a disordered region responsible for its interaction with poly(A) polymerase alpha (PAPOLA), which mediates polyadenylation of mRNA (Laishram, 2014). As for SEC24C, the increased exon skipping was expected to result in loss of part of its 5'UTR sequence (Figure 3-38G).

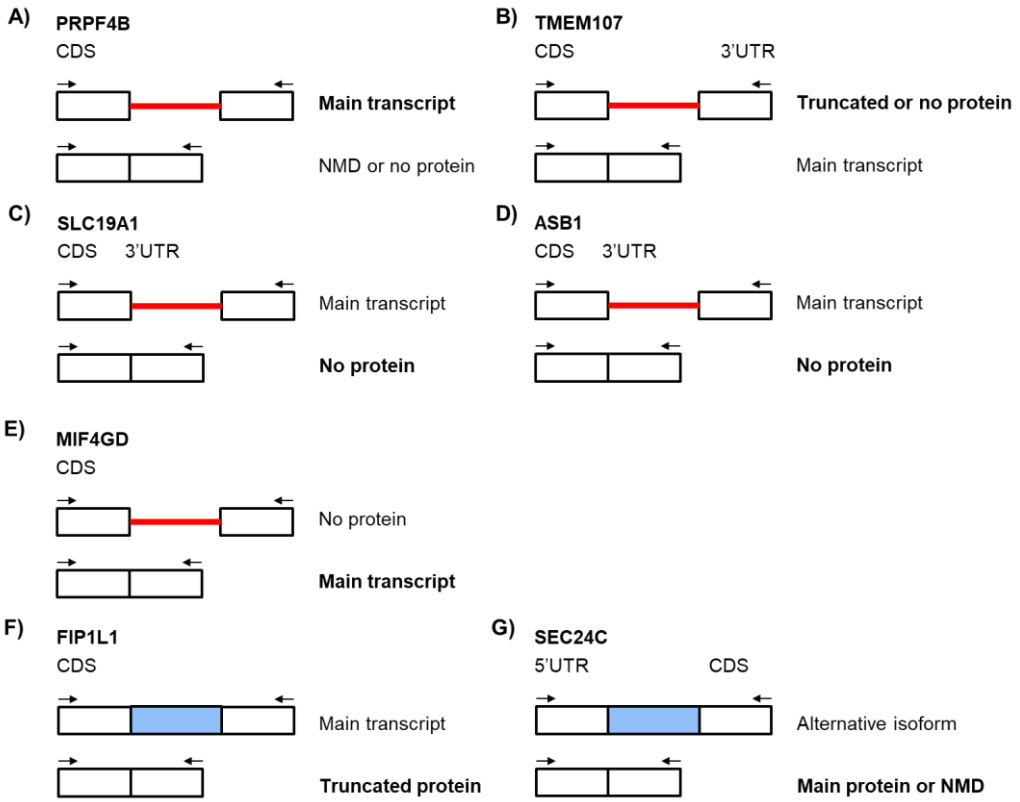


Figure 3-38: Summary of outcomes of dysregulated splicing upon hnRNPM knockdown

The expected outcome upon hnRNPM loss is highlighted in bold.

Sequence and structural elements within the 5' and 3' UTRs of transcripts have been found to control the localization, stability and translation of mRNA (Chatterjee and Pal, 2009; Leppek et al., 2018; Mignone et al., 2002), and it is conceivable that the outcome of some of these splicing events could be dysregulated protein expression. I would subsequently need to examine whether loss of hnRNPM functionally results in changes to the protein products of these genes, and whether these events contribute to LUAD tumorigenesis.

3.7.4 Dysregulation of splicing observed in Asian LUAD cohort cases with PARP4 copy number loss

As I hypothesized that PARP4 could alter hnRNPM function through its interaction with and/or ADP-ribosylation of hnRNPM, I wanted to determine whether PARP4 loss similarly resulted in dysregulated splicing. To gain a preliminary understanding of how PARP4 might be implicated in splicing regulation, I looked to available RNA-seq data from the Asian LUAD cohort to compare splicing profiles between PARP4 copy number loss and diploid patients. This could reveal a number of insights: firstly, this analysis could provide an indication of how PARP4 contributes to the splicing landscape in LUAD. Secondly, this could point out clinically relevant splice events. Thirdly, I wanted to determine if there were overlaps with the hnRNPM-regulated splice events identified from the Nanopore sequencing analysis, which could point to co-regulation by PARP4.

To ensure that PARP4 expression levels were sufficiently different between the two comparison groups to be able to pick up meaningful splicing differences, I identified a third of PARP4 copy diploid cases with the highest PARP4 expression levels (n=25) as well as the third of PARP4 copy number loss cases with the lowest PARP4 expression levels (n=27). Indeed, as compared to using all available patient data, the difference in PARP4 levels between the smaller comparison groups became more pronounced (Figure 3-39A).

As previous Illumina sequencing of these patient samples generated short-read sequencing data, rMATS (Shen et al., 2014), which is a pipeline suited for splicing analyses based on short-read data, was used for the data analysis. The rMATS analysis was performed by Cheryl Phua (Genome Institute of Singapore) and specifically focused on IR and SES events as these were most dysregulated in the hnRNPM analysis above.

This analysis revealed 1030 significantly upregulated and 49 significantly downregulated IR events, as well as 383 significantly upregulated and 857 significantly downregulated SES events (Figure 3-39B). Strikingly, the much greater number of upregulated versus downregulated IR events mirrors what was earlier observed upon hnRNPM knockdown in SAECK cells (Figure 3-34B). This could potentially suggest a role for PARP4 in regulating intron splicing.

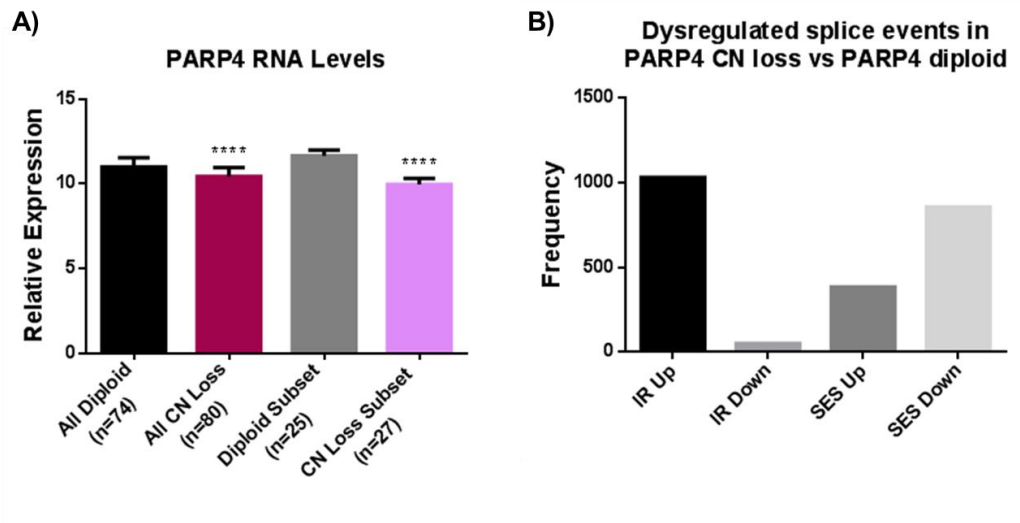


Figure 3-39: Comparison of splicing between PARP4 copy number loss versus PARP4 diploid patients

A) PARP4 RNA expression in PARP4 diploid versus PARP4 copy number loss cases, before and after further stratification by PARP4 expression levels. Data represent the mean ± s.d., ****p<0.0001, as determined by unpaired two-tailed t-test. **B)** Number of significantly upregulated and downregulated IR and SES events detected by rMATS analysis of stratified PARP4 copy number loss versus diploid patients. An event with $|\Delta PSI| > 5$ and p-value < 0.05 was considered to be significantly dysregulated.

An Enrichr analysis further revealed that these dysregulated IR and SES events between the PARP4 copy number loss and diploid patients were collectively enriched in genes involved in RNA metabolism, as GO terms for regulation of splicing, RNA processing and RNA export were among the top 15 significantly and highly enriched gene sets (Figure 3-40). This supports a wider role for PARP4 in the regulation of RNA metabolism.

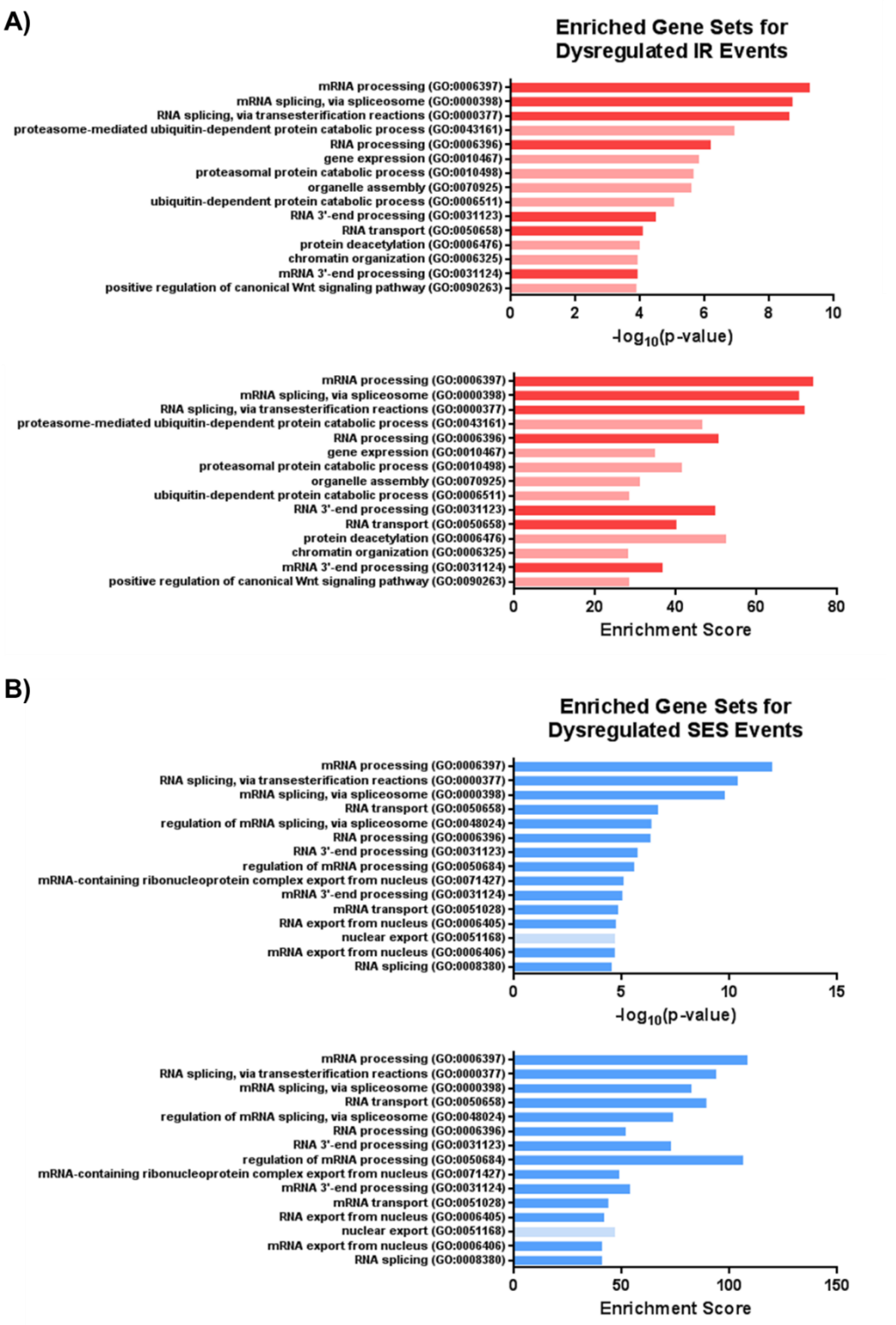


Figure 3-40: Enrichr analysis of genes with dysregulated IR or SES events in Asian LUAD cohort patients

A, B) Top 15 enriched GO Biological Process 2021 gene sets for genes with significantly dysregulated **A)** IR or **B)** SES events in the PARP4 copy number loss versus PARP4 diploid splicing analysis splicing ($|\Delta PSI| > 5$, p-value < 0.05). In the top panel, the gene sets are ranked by their p-values. In the bottom panel, the enrichment scores for the same ranking of gene sets are indicated. Gene sets related to RNA metabolism and splicing are highlighted with a darker shade. Analysis was performed using the Enrichr platform (Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021).

Furthermore, when these genes with dysregulated IR and SES events were compared with those from the hnRNPM analysis above using the same filtering threshold settings, there was a subset of genes commonly dysregulated in both analyses. 12% of genes with hnRNPM-regulated IR events (Figure 3-41A) and 17% of genes with hnRNPM-regulated SES events (Figure 3-41B) were also differentially spliced between the PARP4 copy number loss and diploid groups. This suggests that there could be potentially overlapping regulation of alternative splicing between hnRNPM and PARP4 that will need to be further investigated.

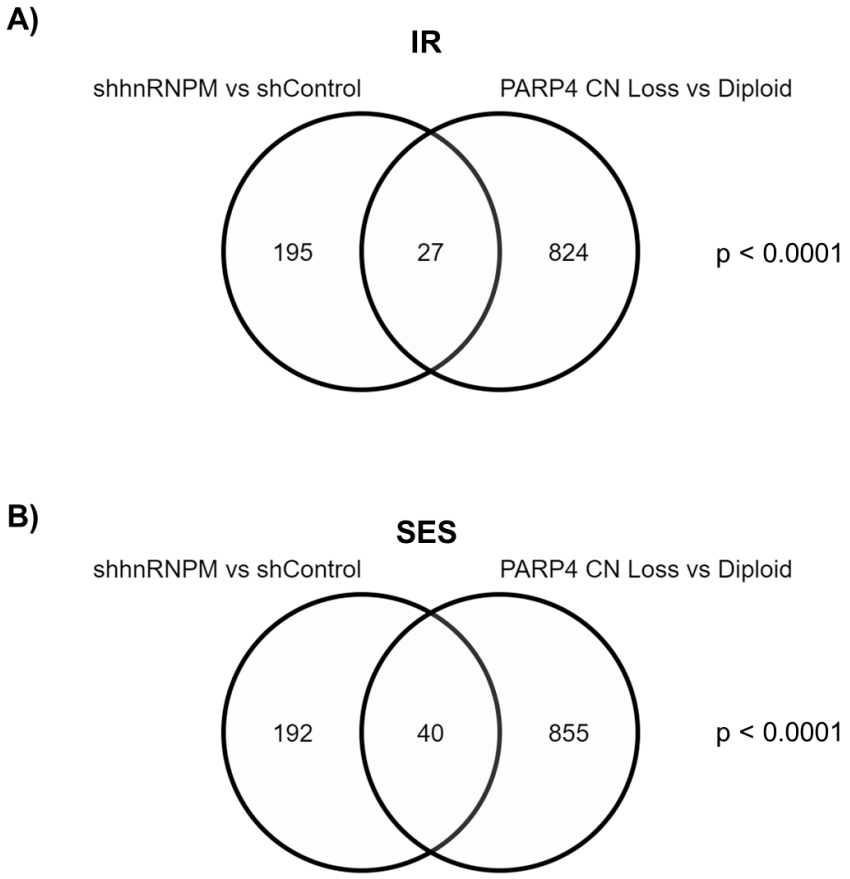


Figure 3-41: Overlap in genes with dysregulated splicing between the hnRNPM knockdown and PARP4 copy number analyses

A) Overlap in genes with significantly dysregulated IR ($|\Delta PSI| > 5$, p-value < 0.05) identified from the SAECK shhnRNPM versus shControl analysis and the PARP4 copy number loss versus diploid analysis. p denotes the probability that 27 randomly selected genes detected by the PSI-Sigma pipeline (n=7083) would overlap with genes from the PARP4 CN loss vs diploid analysis (n=851). **B)** Overlap in genes with significantly dysregulated SES ($|\Delta PSI| > 5$, p-value < 0.05) identified from the SAECK shhnRNPM versus shControl analysis and the PARP4 copy number loss versus diploid analysis. p denotes the probability that 40 randomly selected genes detected by the PSI-Sigma pipeline (n=7083) would overlap with genes from the PARP4 CN loss vs diploid analysis (n=895).

To conclude, our splicing analyses using Nanopore sequencing and the PSI-Sigma pipeline has provided a glimpse of the splicing landscape regulated by hnRNPM in transformed lung cells and allowed better profiling of intron retention events that are still not well studied. PARP4 may also have a wider role in splicing, as evidenced by the changes in the landscape of IR and SES between PARP4-diploid and PARP4 copy number-deleted patients.

Strikingly, there were some parallels between the SAECK hnRNPM knockdown and patient cohort PARP4 copy number loss analyses. These were namely in the form of upregulated intron retention events, as well as shared genes with dysregulated IR and SES splicing. It would be informative to further examine these overlapping events and their functional contribution to tumorigenicity.

3.7.5 Graphical summary

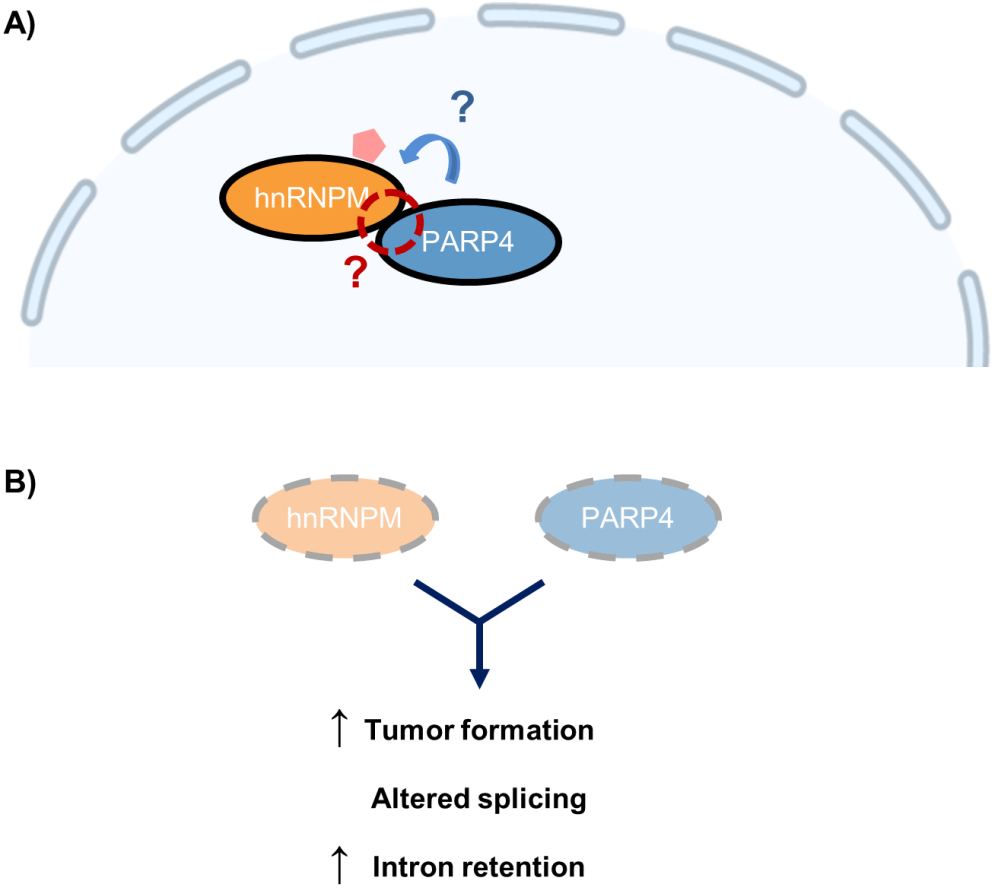


Figure 3-42: Schematic diagram illustrating the relationship between hnRNPM and PARP4 and their proposed functions

A) hnRNPM was found to be a novel PARP4 interaction partner. In SAECK cells, the interaction was specifically localized to the nucleus. The interaction between hnRNPM and PARP4, as well as potential ADP-ribosylation of hnRNPM by PARP4, are proposed to modulate tumorigenicity. These would need to be further studied and defined. **B)** Loss of hnRNPM or PARP4 results in the shared phenotypes of increased tumorigenicity and altered splicing, with notably upregulated intron retention events.

CHAPTER 4: Discussion

4.1 Identification of PARP4 as a novel tumor suppressor in LUAD

Somatic mutations in PARP4 were reported in 6% of the Asian LUAD cohort, and PARP4 was predicted as a novel candidate driver (Chen et al., 2020b). Interestingly, PARP4's role as a tumor suppressor in LUAD had not been picked up from earlier LUAD cohort studies. Whereas PARP4 copy number loss was equally prevalent in both the Asian as well as the Caucasian TCGA LUAD cohorts (Figure 3-4), PARP4 mutations were found to be associated with a Chinese ethnic bias (Figure 3-2).

Furthermore, prior to this cohort, previous Asian LUAD sequencing cohorts were small in size, with the largest among them having only 131 patients, as opposed to 305 patients in our cohort (Table 1-2). Smaller cohorts meant that rare driver events could have been missed due to low statistical power (Jiang et al., 2019; Tokheim et al., 2016). These collectively highlight the importance of having a representative cohort in terms of ethnicity and size, for the prediction and identification of novel and relevant drivers.

PARP4 is one of the lesser understood members of the PARP family. In this study, novel tumor suppressive functions in LUAD were attributed to PARP4. Notably, using our model cell line systems, loss of PARP4 and its associated tumor suppressor function were shown to be relevant only in SAECK but not SAECTS, suggesting the requirement of oncogenic stimuli. Furthermore, following from the analysis of the Asian cohort data, PARP4 copy number loss was found to be associated with EGFR or KRAS mutations, and I further demonstrated that loss of PARP4 increases the tumorigenicity of both EGFR- or KRAS-driven lung cancer cell lines.

Prior to this study, PARP4-deficient mice were found to be more likely than wildtype mice to develop colon and lung tumors when challenged with chemical carcinogens, pointing to a potential tumor suppressive role (Raval-Fernandes et al., 2005). PARP4 had previously been proposed as a candidate

cancer susceptibility gene, with germline variants (T1170I (n=5), G496V (n=1)) identified in six among 14 patients who developed independent thyroid and breast cancers, had a familial history of cancer, and were null for germline mutations in PTEN, which is a known thyroid and breast cancer susceptibility gene (Ikeda et al., 2016). To support this, PARP4 depletion was subsequently performed and found to increase proliferation in the HCC1143 breast cancer cell line. In light of these findings, an independent study sought to evaluate PARP4's potential as a candidate cancer susceptibility gene in BRCA1/2-wildtype breast cancer patients (n=198) and identified eight different germline PARP4 missense variants in 5.5% of the patients (n=11) (Prawira et al., 2019). However, in their validation studies, PARP4 CRISPR knockout clones derived from MDA-MB-231 breast cancer cells were reported to have a slight but significant reduction in proliferation and colony forming ability, with one of three clones unexpectedly having increased motility. It was thus concluded that PARP4 may not be functionally relevant in breast cancer cells. In yet another study, siRNA knockdown of PARP4 in HeLa cells neither decreased cell viability nor yielded any obvious morphological defects (Vyas et al., 2013). These discordant findings suggest that the cell line and cancer context may be important determinants of PARP4's tumor suppressive activity.

Notably, the experiments conducted across these studies were mostly limited to proliferation assays. In our study however, PARP4 loss did not affect proliferation under normal cell culture conditions, yet significantly promoted *in vitro* and *in vivo* tumorigenicity. This suggests that the 3D context provided by the soft agar assay and xenograft transplantation experiments may be important for PARP4's tumor suppressive activity. It would thus be important to re-examine the relevance of PARP4 in other cancer cell lines using these assays.

4.2 PARP4 mutations and domains relevant to tumorigenicity

As for the significance of the recurrent I1039T mutation in PARP4 that was found in our cohort, PolyPhen-2 analysis predicted a deleterious effect for the variant (Figure 3-16). I have also shown through comparative studies between $PARP4^{WT}$ and $PARP4^{I1039T}$ overexpression in a clonal knockout background that $PARP4^{I1039T}$-expressing cells had lower protein levels in spite of comparable transcript levels, and formed larger tumors. Furthermore, patients bearing the I1039T mutation did not have significantly different PARP4 expression levels compared to patients without PARP4 mutations. I thus concluded that the I1039T mutation could potentially reduce the stability of PARP4 protein, effectively resulting in a loss-of-function phenotype. Further analyses on the effect of the I1039T mutation on PARP4 protein stability by inhibiting protein synthesis or degradation using cycloheximide or MG-132 respectively could be performed to confirm this. This should also be confirmed in additional cell lines.

Deleterious amino acid substitutions have been proposed to either replace key functional residues, such as those involved in catalytic activity or post-translational modification, or affect protein scaffolding by destabilizing interaction sites (Bromberg and Rost, 2009; Teng et al., 2010; Wang and Moult, 2001). In the case of the I1039T mutation, isoleucine, a nonpolar amino acid, is converted to threonine, a polar amino acid bearing a hydroxyl group that is amenable to post-translation modifications. Threonine is a potential phosphorylation site and whether phosphorylated or unphosphorylated, the presence of either a negative charge or polar group where there was originally a hydrophobic residue could conceivably disrupt protein interactions important in stabilizing PARP4. In fact, mutations causing gain or loss of phosphorylation was previously found to be significantly enriched in cancer somatic mutation datasets as opposed to control datasets (Radivojac et al., 2008). To fully understand the impact of I1039T on protein stability and function, more experiments will be needed.

To determine whether PARP4's enzymatic activity of ADP-ribosylation was important in its tumor suppressive functions, I ablated the PARP catalytic domain and showed that loss of this domain resulted in increased tumorigenicity. However, the PARP domain in PARP1 was also shown to have additional protein interaction activities and it would thus be important to distinguish the two functionalities (Kumari et al., 1998; Pinnola et al., 2007; Wacker et al., 2007). As the key active site residues in the PARP catalytic domain had been identified as a triad of histidine, tyrosine and glutamate (H-Y-E) (Gupte et al., 2017), I plan to generate a catalytically dead version of PARP4 by converting the negatively charged glutamate 547 residue of the catalytic triad to a positively charged lysine residue (E547K), which had previously been shown to abrogate PARP1 catalytic activity (Marsischky et al., 1995). By conducting rescue experiments where we overexpress either wildtype, PARP domain-deleted or catalytically dead PARP4, we could tease out if the catalytic activity or scaffolding role of the PARP domain contributes to PARP4's tumor suppressive role.

4.3 Additional functionality for PARP4 beyond that of the vault complex

PARP4 was initially discovered in the vault complex (Kickhoefer et al., 1999a), with what little functional studies of its activity focused on the context of the vault complex. It thus came as a surprise that PARP4's tumor suppressive activity appears to be independent of MVP. Since MVP is the main structural component of the vault complex, this then meant that PARP4's tumor suppressive activity is also independent of the vault complex. This was shown from experiments whereby depletion of MVP did not result in significant increase in tumorigenicity, and that MVP levels had no effect on LUAD patient survival.

Intriguingly, the interaction between PARP4 and MVP, which had previously been reported in other model systems and was suggested to help stabilize PARP4 protein (Kozlov et al., 2006; Yu et al., 2017), was further confirmed in the SAECK cells through deletion of PARP4's MVP-interaction domain. Furthermore, varying MVP levels had no effect on PARP4 transcript

levels but affected PARP4 protein stability, as observed from the cycloheximide and MG-132 studies.

Having observed that there remains a fraction of PARP4 that was resistant to proteasomal degradation in the absence of MVP, one possibility for MVP-independent PARP4 stability is its association with other proteins. In fact, SILAC co-IP mass spectometry later uncovered novel PARP4 binding partners, of which I was able to validate several in the SAECK cells, with hnRNPM being the most consistent of them and validated across multiple cell lines. While hnRNPM depletion in SAECK and A549 cells had no appreciable effect on PARP4 protein levels, this could be because most of PARP4 appears to be already bound to and stabilized by MVP (Figure 3-21, Figure 3-25). Hence, it would be interesting to study whether hnRNPM stabilizes PARP4 in the context of MVP loss.

Furthermore, I found that while PARP4 localizes mainly to the cytoplasm in SAECK cells, a small fraction could also be found in the nucleus. This was in line with some of the discordant studies observing PARP4 nuclear localization (Kickhoefer et al., 1999a; Liu et al., 2004; Vyas et al., 2013; van Zon et al., 2003b). Moreover, analysis of PARP4 protein sequence revealed the presence of two predicted nuclear localization signals – one lies near the N terminus within the BRCT domain while the other is located between the VWFA and MVP interaction domains. Importantly, while the interaction between PARP4 and hnRNPM could occur in both the cytoplasmic and nuclear fractions in normal A653N lung cells, the interaction between PARP4 and hnRNPM was detected only in the nuclear compartment of SAECK cells. This suggests an important role for nuclear PARP4 in its tumor suppressive properties.

The possibility of an MVP-independent nuclear fraction of PARP4 that is responsible for its tumor suppressor activity is an appealing hypothesis. For one, this would be consistent with hnRNPM being a predominantly nuclear protein, and could account for the proposed roles of PARP4 in regulating splicing, which occurs in the nucleus. To test whether PARP4's tumor suppressive activity indeed resides in the nuclear fraction, it would be important to perturb PARP4's

cellular localization to study its effects on tumorigenicity. Ablating PARP4's predicted nuclear localization sequences should expectedly reduce nuclear PARP4 localization, unless PARP4's mechanism of nuclear import relies on other means (Freitas and Cunha, 2009). If PARP4 can be successfully ablated from the nucleus using this method, then its interaction with hnRNPM as well as its effect on tumorigenicity can both be tested. Furthermore, it would be interesting to separately evaluate MVP's effects on stabilizing cytoplasmic and nuclear PARP4 through cycloheximide and MG-132 studies. According to our hypothesis, MVP would likely stabilize only cytoplasmic but not nuclear PARP4. It would be more difficult to perform the converse experiment of ablating cytoplasmic PARP4. Stronger nuclear localization sequences could perhaps be incorporated into PARP4's protein sequence to force its nuclear localization.

Additionally, as cytoplasmic and nuclear PARP4 may have unique interaction partners, it could be informative to perform additional co-IP mass spectrometry studies on the individual fractions. Given that PARP4 regulates a subset of hnRNPM-regulated events but additionally alters the splicing landscape in ways different from hnRNPM, its nuclear localization could further suggest that PARP4 could be binding to other splicing regulatory factors in the nucleus.

4.4 hnRNPM is a novel PARP4 interaction partner that has tumor suppressive activity in LUAD

In this study, hnRNPM was found to be a novel PARP4 binding partner through SILAC co-IP mass spectrometry. In fact, the interaction between PARP4 and hnRNPM could be reliably recapitulated in several lung normal and cancer cell lines. Loss of hnRNPM through shRNA knockdown revealed that hnRNPM too could function as a tumor suppressor in LUAD. Just like PARP4, low hnRNPM expression was correlated with poor overall survival (Győrffy et al., 2013). The experiments collectively showed that loss of either PARP4 or hnRNPM led to increased tumorigenicity, suggesting that both proteins have tumor suppressive roles.

Interestingly, this is one of the rare studies suggesting a tumor suppressive role for hnRNPM, as it was reported to have tumor-promoting effects in most other cancers (Chen et al., 2014; Ho et al., 2021; Palombo et al., 2020; Passacantilli et al., 2017; Wang et al., 2021; Xu et al., 2014). Hence, this points to hnRNPM as a cancer-related gene which has tumor suppressive or tumor-promoting effects depending on its cellular context.

As I have shown that loss of PARP4 and hnRNPM results in several shared phenotypes, notably increased tumorigenicity, altered splicing and increased dysregulation of IR events, I hypothesized that hnRNPM and PARP4 could be acting in the same pathway. I further suggest that PARP4's interaction with, and potential modification of hnRNPM could perhaps be important in mediating the tumor suppressor phenotype. However, I have not directly shown this to be true. To do so convincingly, I would first need to identify the exact domains required for PARP4 and hnRNPM's interaction so that I can subsequently perturb them to determine the outcome on PARP4 and hnRNPM's tumor suppressive properties.

Furthermore, post-translational modification of hnRNPM in the form of serine phosphorylation was shown to have direct effects on hnRNPM's splicing regulation in mouse macrophages (West et al., 2019). Specifically, the modified serines were located in one of hnRNPM's RNA recognition motifs and were suggested to affect its splice site recognition. It is thus a plausible hypothesis that PARP4 could ADP-ribosylate hnRNPM to similarly alter its activity. Intriguingly, a search on an online database (Ayyappan et al., 2021; Vivelo et al., 2017) of known ADP-ribosylation modifications showed that hnRNPM could be ADP-ribosylated by PARP1, although the exact site of modification and its suggested function were not reported. Furthermore, ADP-ribosylation of hnRNPs had previously been shown to modulate splicing (Ji and Tulin, 2009, 2013). Thus, PARP4's binding to and modification of hnRNPM could possibly affect hnRNPM's splicing activity. This would need to be further studied.

4.5 Role of hnRNPM in dysregulated splicing

hnRNPM is a well-known splicing factor. There have been several notable studies profiling hnRNPM's RNA specificity, binding properties and regulation of alternative splicing, performed with the aim of dissecting hnRNPM's roles and functionalities in different contexts.

In one study, siRNA targeting of hnRNPM alongside a few other hnRNPs was performed to determine their effect on a selection of 56 alternative splicing events within apoptotic genes across three cell lines – HeLa (cervical cancer), PC-3 (prostate cancer) and BJT (foreskin derivative) (Venables et al., 2008). Overlaps between cell lines for the same hnRNP were minimal or poor, suggesting that the effects of individual hnRNP loss on splicing could be cell line-dependent. This was followed by a more global analysis where the effect of hnRNPM depletion from HEK293 cells was examined through splice junction array studies, as well as cross-linking and immunoprecipitation followed by high throughput sequencing (CLIP-seq) to study hnRNPM's binding sites on RNA (Huelga et al., 2012). Through such CLIP-seq studies as well as targeted *in vitro* affinity experiments using synthetic RNA fragments, hnRNPM was found to selectively bind GU-rich motifs, particularly within intronic regions of target RNAs (Datar et al., 1993; Dreyfuss et al., 1988; Hovhannisyan and Carstens, 2007; Huelga et al., 2012; Swanson and Dreyfuss, 1988).

Interestingly, we observed that loss of hnRNPM from SAECK cells resulted primarily in the increase of intron retention events, suggesting that hnRNPM could have a role in controlling the proper removal of introns from transcripts.

Prior to this study, there had only been two other reports tying hnRNPM to intron retention events. In the first study, hnRNPM and SFPQ, the latter of which is also a splicing factor, were found to bind to the same retained intron within SFPQ itself. The retained intron leads to nuclear loss of SFPQ and underlies amyotrophic lateral sclerosis (ALS) (Luisier et al., 2018). hnRNPM was also found to be a component of nuclear stress bodies, which are membrane-

less aggregates comprising 141 proteins – of which many are RNA-binding proteins, that nucleate around the HSATIII lncRNA upon thermal stress and go on to promote the retention of 533 distinct introns during recovery from the thermal exposure (Ninomiya et al., 2020). However, the study did not conclude whether hnRNPM itself was directly involved and required in promoting these intron retention events.

In fact, IR events are common in cancers. An analysis comparing RNA processing between normal and tumor tissue across multiple cancer types revealed that almost all cancers had alterations in intron retention (Dvinge and Bradley, 2015). In a separate study, intron retention was also observed to be a common mechanism underlying the inactivation of tumor suppressors (Jung et al., 2015).

Intron retention events could result in a few different outcomes. More often than not, IR results in the introduction of a premature termination codon and leads to transcript degradation through nonsense-mediated decay (Belgrader et al., 1994; Braunschweig et al., 2014). In fact, a study comparing protein and mRNA levels across nine tissue types found that genes with IR events were significantly associated with lower protein levels (Middleton et al., 2017). Some transcripts bearing premature termination codons have been observed to escape nonsense-mediated decay, but these transcripts tend not to be translated (Middleton et al., 2017). Finally, the retained intron could introduce a new functional element within the mRNA to produce a different isoform (Bell et al., 2008; Buckley et al., 2011).

In the case of hnRNPM knockdown in SAECK cells, several of the validated changes in intron retention were indeed predicted to alter the splicing outcome between a functional transcript isoform versus one that has no protein product.

4.6 Possible PARP4 and hnRNPM interaction in regulating splicing

While the comparison between PARP4 copy number loss and diploid patients could have confounding factors such as differences in known driver backgrounds, a preliminary analysis revealed the presence of significant changes in the regulation of alternative splicing between the two groups, as marked by significantly dysregulated IR and SES events. As analysis on the cohort data was specifically performed on IR and SES events, I am unable to comment on the importance of IR relative to other splice event categories, unlike how we have demonstrated that IR was most perturbed among the five event categories in the hnRNPM knockdown cells. Nevertheless, it was striking that the PARP4 CN loss patients tended to have more upregulated intron retention events as opposed to downregulated ones, mirroring our observation in the hnRNPM knockdown cells.

Furthermore, from the comparison between PARP4 CN loss and diploid patients, dysregulated IR or SES events were enriched in genes involved in splicing and other RNA processing pathways. Given that splicing factors have been reported to regulate their own splicing as well as that of other RNA-binding proteins (Geuens et al., 2016; Huelga et al., 2012), one hypothesis from this finding was that PARP4 copy number loss could be associated with changes in the expression of various splicing factors and RNA-binding proteins. Analysis of the transcriptomic differences between the two patient groups would thus be important. If this were true, it could possibly also account for PARP4's likely indirect mechanism of regulating alternative splicing, since PARP4 does not have known RNA-binding domains and ability.

Indeed, while my experiments show that hnRNPM and PARP4 affect alternative splicing, whether these are direct effects with hnRNPM directly binding and causing the splicing change, or indirect effects where PARP4 or hnRNPM modify the activity of other splicing factors, remains to be understood. To determine this, CLIP-seq studies could be performed, whereby cross-linking is performed to capture the interactions between protein and RNA, followed by immunoprecipitation of hnRNPM or PARP4 and sequencing of the bound RNA

fragments. This would also establish whether PARP4 has any direct RNA-binding abilities.

Having observed a subset of genes with dysregulated IR or SES that are overlapping between hnRNPM and PARP4, it would be important to follow up with targeted PCR validation and additional experiments to understand their direct relevance to tumorigenicity. In addition to PCR assays comparing hnRNPM and PARP4 knockdown cells to control, re-expression of hnRNPM and PARP4 could demonstrate the specificity of PARP4 and hnRNPM's role in regulating these splice events. Lastly, and most crucially, following the identification of the binding site between PARP4 and hnRNPM, it would be important to determine if loss of the interaction alters these events in the same manner as loss of the entire protein. These would then enable me to conclude whether these events are jointly or separately regulated by PARP4 and hnRNPM.

In any case, we can still conclude from our findings that hnRNPM and PARP4, at least independently, both have tumor suppressive functions and can mediate splicing changes.

CHAPTER 5: Future Directions

While this work has uncovered PARP4 as a novel tumor suppressor in LUAD and attributed new functionalities to a gene that is not well-studied, there remain several important and interesting questions to be answered: (1) what is the exact nature of PARP4's relationship with hnRNPM; (2) is PARP4's ADP-ribosylation activity involved and if so, how; (3) are the dysregulated splicing events functionally important in tumorigenicity; and finally, (4) what is the clinical utility of this new knowledge?

5.1 To further understand the role of I1039T on PARP4 protein stability

Having observed that the recurrent I1039T mutation appears to affect PARP4 expression at the post-transcriptional level, comparing the effects of cycloheximide and MG-132 treatment between SAECK PARP4 clonal knockout cell lines expressing $PARP4^{WT}$ or $PARP4^{I1039T}$ would demonstrate whether the mutation reduces protein stability as hypothesized.

5.2 To understand the nature of PARP4 and hnRNPM's interaction

5.2.1 To identify the binding site between PARP4 and hnRNPM

To narrow down the site on PARP4 responsible for binding to hnRNPM, a collection of PARP4 clonal knockout cell lines overexpressing PARP4 variants with deletion of each of PARP4's annotated domains has been generated (Figure 5-1). In addition to these mutants, long intervening sequences between the individual domains could affect binding and should also be deleted and studied.

To identify the site on hnRNPM responsible for binding to PARP4, hnRNPM clonal knockout cell lines expressing different structural regions of hnRNPM could also be generated.

Collectively, these variants can be used in co-immunoprecipitation studies. Specifically, immunoprecipitation of hnRNPM should be performed in the mutant PARP4 lines to assess which of the PARP4 variants cannot be successfully co-immunoprecipitated, and is thus responsible for hnRNPM-

binding. In a similar manner, immunoprecipitation of PARP4 could be performed in the mutant hnRNPM lines. The interaction site on PARP4 and hnRNPM can thus be respectively identified.

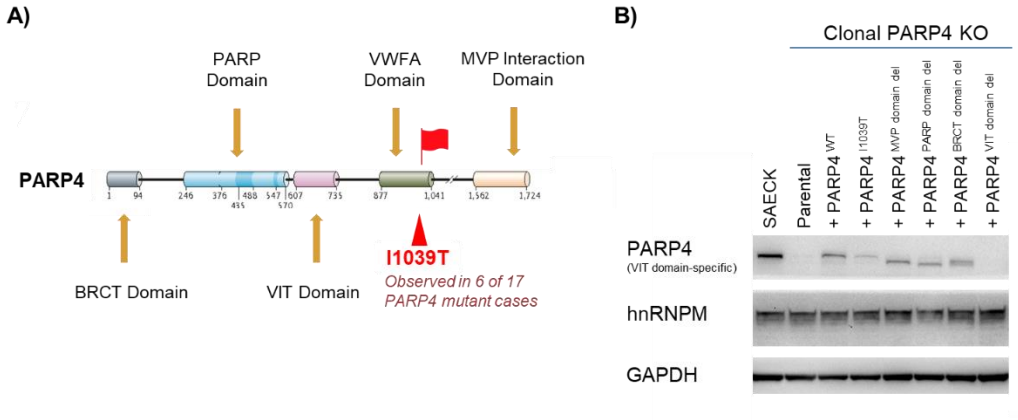


Figure 5-1: Generation of PARP4 domain mutants

A) Schematic diagram of PARP4 domain features. PARP4 domain structure adapted from (Schreiber et al., 2006). **B)** Immunoblot depicting PARP4 and hnRNPM protein levels in the various PARP4 mutant cell lines. GAPDH was used as a loading control. The domain-deleted PARP4 variants expectedly had a downward shift in their molecular weights.

5.2.2 To determine if PARP4's interaction with hnRNPM is required for tumor suppression

Following the identification of PARP4 and hnRNPM's interaction sites, the requirement of this interaction for their tumor suppressive activities can be evaluated. Specifically, the respective mutant cell lines lacking the interaction site can be subjected to tumorigenicity assays. If PARP4 and hnRNPM's interaction is indeed responsible for the tumor suppressive activities, then these mutants should have increased tumorigenicity compared to their respective wildtype control.

5.2.3 To understand the properties of the MVP-independent PARP4 fraction

Having shown that there is a fraction of PARP4 that is resistant to proteasomal degradation in spite of loss of MVP (Figure 3-21), it could be interesting to determine whether this MVP-independent fraction is: (1) localized to the nucleus, and/or (2) stabilized by hnRNPM.

To study (1), the cycloheximide and MG-132 experiments can be performed as before on control and MVP-depleted cells, with the addition of a subsequent fractionation step to compare PARP4 protein levels in the cytoplasmic and nuclear compartments between different conditions. If the MVP-independent fraction of PARP4 is indeed localized to the nucleus, MVP depletion should only affect the levels of cytoplasmic but not nuclear PARP4.

To study (2), hnRNPM knockdown lines could be generated from parental MVP-depleted cell lines. Transcript and protein levels of PARP4 could then be compared between the control and hnRNPM knockdown lines. If PARP4 protein but not transcript levels appear to be further reduced upon loss of hnRNPM, then both sets of cell lines can subsequently be treated with cycloheximide and MG-132 to examine their effect on PARP4 protein levels. If hnRNPM indeed stabilizes PARP4, I would expect to see even lesser PARP4 protein in the hnRNPM knockdown compared to control in the absence of treatment, in spite of similar PARP4 protein levels in the MG-132-treated condition.

5.3 To identify the role of nuclear PARP4 in tumorigenicity

In parallel to the experiments proposed in the preceding section, the role of PARP4's nuclear localization in tumorigenicity can be directly determined by ablating nuclear PARP4. Specifically, PARP4's predicted nuclear localization sequences (amino acid residues 19-25 within the BRCT domain, and 1237-1249 between the VWFA and MVP interaction domains) can be deleted. After confirming that PARP4 is indeed lost from the nucleus but not the cytoplasm in

these variants, their effects on tumorigenicity, interaction with hnRNPM, stability independent of MVP and splicing can all be examined.

5.4 To identify the role of PARP4's catalytic activity and ADP-ribosylation targets

Having earlier shown that PARP4's PARP domain is important in tumorigenicity, and based on the understanding that the PARP domain could have other roles in mediating protein interactions apart from its catalytic activity, it would be important to determine precisely which of these functionalities is required for tumor suppression. Hence, I intend to generate a catalytically inactive E547K variant that disrupts the catalytic triad, while retaining the PARP domain structure.

Comparing the tumorigenicity of this variant relative to the PARP domain-deleted and wildtype PARP4 would give an indication of the degree by which ADP-ribosylation activity contributes to PARP4's tumor suppression.

If PARP4's catalytic activity is shown to be important, it would be interesting to identify PARP4's ADP-ribosylation targets and determine whether PARP4 ADP-ribosylates hnRNPM.

I have made preliminary attempts to uncover clues about PARP4's ADP-ribosylation activity. By immunoprecipitating PARP4 in SAECK cells and probing the eluates for poly-ADP-ribose (PAR) or mono-ADP-ribose (MAR) modifications (Figure 5-2), I observed the presence of unique bands that were absent in the IgG negative control, suggesting that PARP4 could potentially be responsible for modifications on its binding partners.

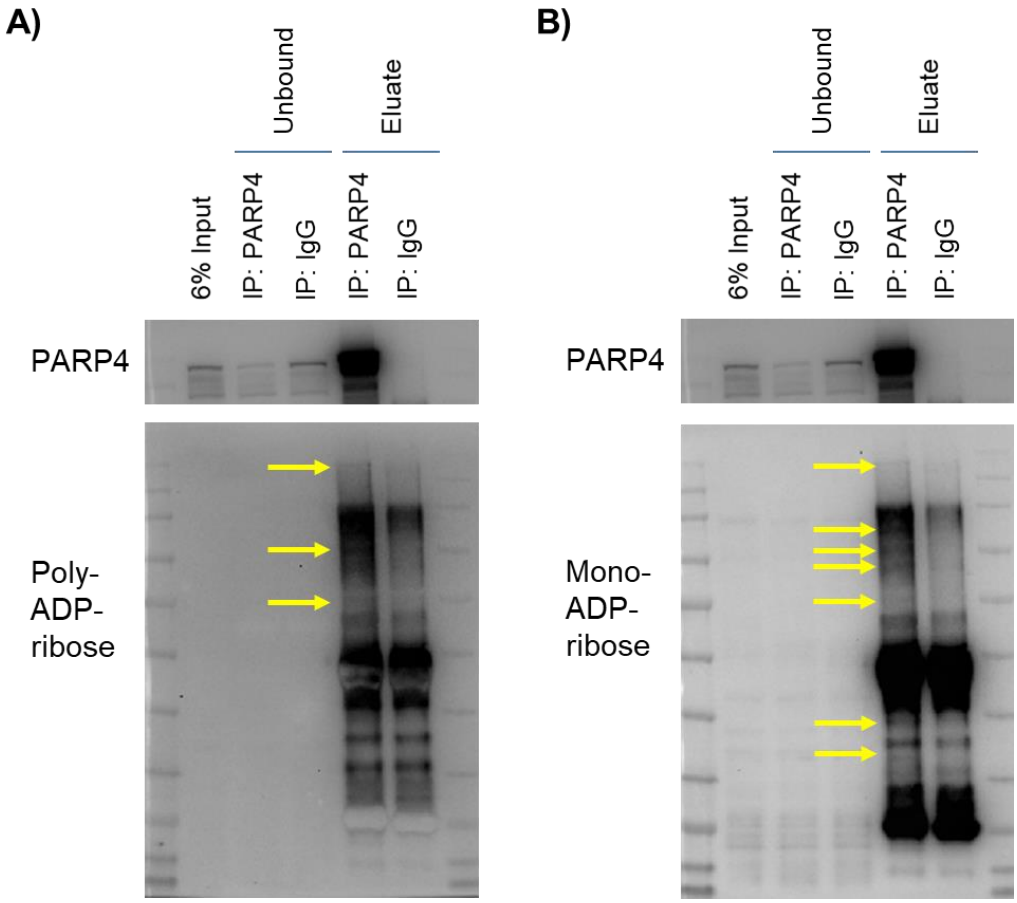


Figure 5-2: ADP-ribosylation of PARP4 binding partners

Immunoblot comparing **A)** poly-ADP-ribose and **B)** mono-ADP-ribose bands between IgG control and PARP4 eluates from immunoprecipitation performed on SAECK lysates. Bands specific to PARP4 eluates are indicated with yellow arrows. The same blot was first probed for poly-ADP-ribose, and then stripped and blocked each time before probing with mono-ADP-ribose and followed lastly by PARP4. The same PARP4 panel is included above both blots for easier viewing and comparison of the lanes.

In an attempt to determine this more directly, I had also modified an *in vitro* ADP-ribosylation assay from (Vyas et al., 2014) that uses radioactively labelled NAD+ to one based on biotinylated NAD+. In this method, immunoprecipitation of PARP4 was followed by incubation with biotinylated NAD+, which would subsequently allow incorporated ADP-ribose groups to be detected by immunoblotting and chemiluminescence methods when streptavidin-conjugated horse radish peroxidase was used. However, in contrast to the high

degree of PARP4 enrichment attained, auto-ADP-ribosylation of PARP4 was only weakly detected (Figure 5-3), suggesting that the modification of the assay from radiolabeling to biotinylation, while making it safer for handling, greatly reduced its sensitivity and utility.

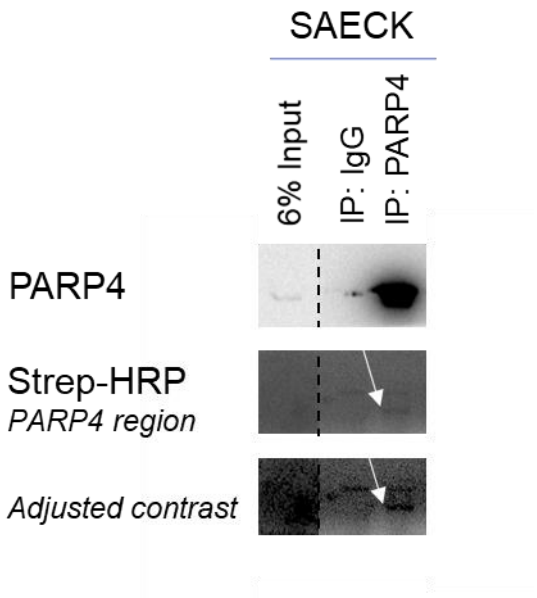


Figure 5-3: Modified *in vitro* ADP-ribosylation assay

Immunoblot depicting results of the in vitro ADP-ribosylation assay conducted on SAECK lysate. PARP4 protein levels and signal from streptavidin-HRP (strep-HRP) at the molecular weight corresponding to PARP4 are shown. A faint band corresponding to auto-modification of PARP4 which is absent in the IgG control is indicated by the white arrow. Below, the contrast was adjusted for better viewing.

As there has not been any global profiling study defining PARP4's ADP-ribosylation targets, unlike what has been done for several of the PARP family members (Ayyappan et al., 2021; Gibson et al., 2016; Jungmichel et al., 2013; Vivelo et al., 2017; Zhang et al., 2013), it could be informative to systematically identify PARP4's ADP-ribosylation targets via mass spectometry.

To specifically attribute ADP-ribosylation to a given PARP family member, several experimental setups involve the mutation of the active site in the PARP family member of interest to accommodate non-natural, chemically modified NAD analogs, thereby allowing downstream targets to be distinguished

by their unique modified ADP-ribosylation labels (Carter-O'Connell et al., 2014, 2016; Gibson et al., 2016; Leutert et al., 2016; Mangerich and Altmeyer, 2016; Rodriguez et al., 2021).

Here, I propose an alternative experimental design that subverts the need for further modifying PARP4 and using the specially synthesized NAD analogs. In this method, ADP-ribosylated proteins can be immunoprecipitated using a poly- or mono-ADP-ribosylation-binding antibody, followed by mass spectrometry to compare the hits between a cell line with endogenous levels of PARP4 versus a matched PARP4-depleted cell line.

5.5 To understand the role of hnRNPM- and PARP4-regulated splicing events on tumorigenicity

Having identified and validated specific IR and SES events that are controlled by hnRNPM, it would be important to demonstrate whether they could similarly be controlled by PARP4, as well as their specificity by rescue experiments with hnRNPM and PARP4 overexpression. Using the various mutants mentioned in the preceding sections, the role of PARP4 interaction and ADP-ribosylation of hnRNPM on splicing outcomes could also be examined.

To examine the role of these splice events in LUAD tumorigenicity, splice-switching oligonucleotides (SSOs) can be employed to reverse the splice events and study their effects on tumorigenicity. Specifically, SSOs are chemically modified antisense oligonucleotides that bind to RNA sequence elements to block access to splicing factors, thereby affecting splicing outcomes (Bauman et al., 2009; Havens and Hastings, 2016). Furthermore, SSOs have demonstrated therapeutic potential and are FDA-approved for the treatment of spinal muscular atrophy (SMA) and Duchenne muscular dystrophy (DMD) (Marabti and Abdel-Wahab, 2021).

Thus, successful SSOs that affect LUAD tumorigenicity may have therapeutic potential that can be further tapped on.

5.6 To determine the clinical utility of PARP4

Having established PARP4's role as a tumor suppressor whereby loss of PARP4 increases tumorigenicity and is associated with poorer prognosis in LUAD patients, it would be useful to identify drug compounds that could preferentially target PARP4-low cancer cells. This could then enhance the repertoire of therapeutic strategies available to LUAD patients. Furthermore, as normal lung cells were shown to have high PARP4 expression, compounds identified using such a screening strategy could be more specific in targeting lung cancer cells.

In a pilot experiment, I performed drug screening on SAECK shControl and shPARP4#1 cells using two drug libraries that were available to us (Figure 5-4). The first drug library comprised 317 anti-cancer compounds from SelleckChem while the second drug library was composed of 303 metabolic drugs from Med Chem Express. Using the first library, I hoped to identify existing anti-cancer therapeutics that may benefit from patient stratification using PARP4 levels. With the second library, I wanted to find novel metabolic drugs that could be re-purposed in anti-cancer therapy.

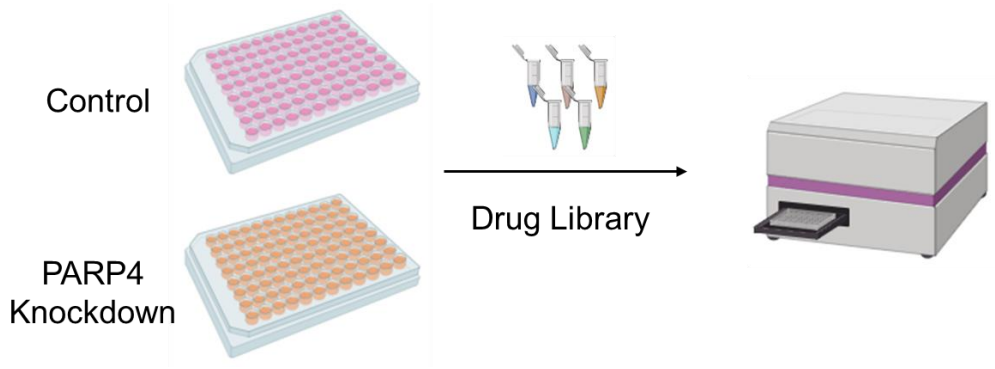


Figure 5-4: Schematic diagram of drug screening process

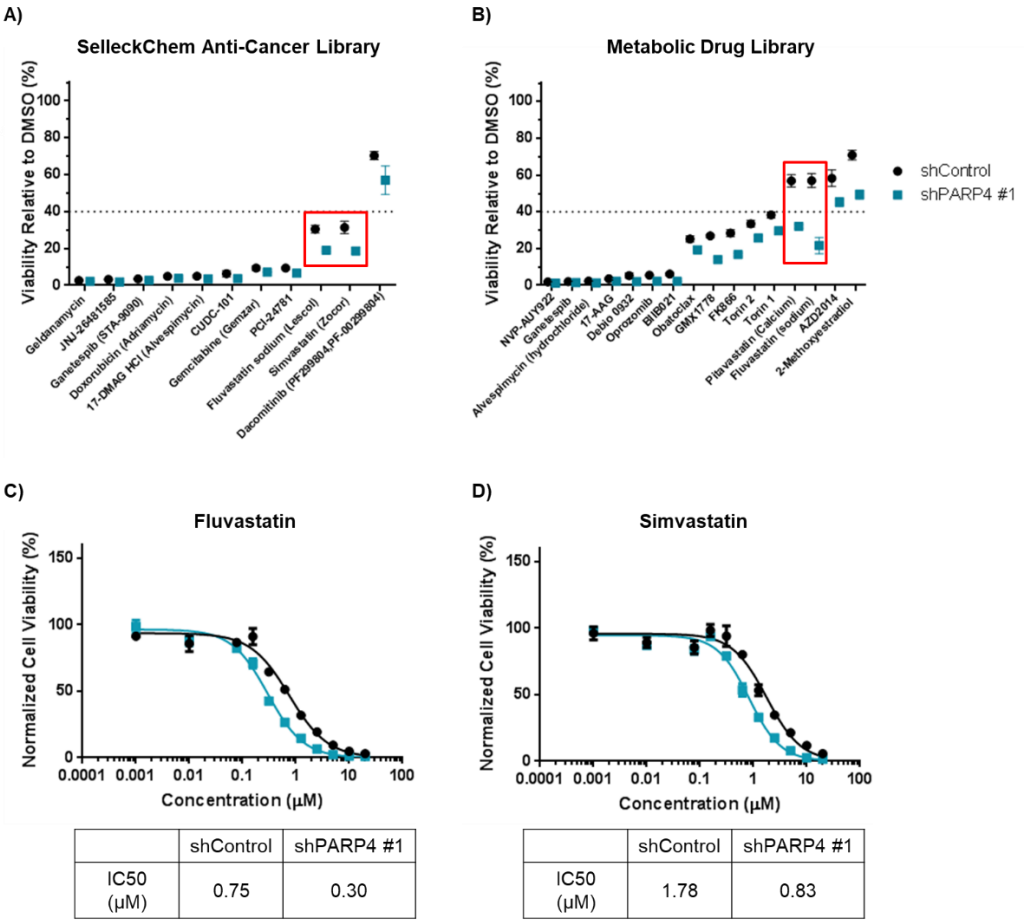
From the initial screening results, compounds that were effective against the PARP4 knockdown cells also appeared to reduce the viability of control cells (Figure 5-5A, B). This suggests that compounds from these libraries could potently kill the SAECK cells even if they were not selective for PARP4 expression levels. Some of the drug hits which preferentially killed the PARP4-depleted cells resulted in cell viabilities over 40% (dacomitinib from the SelleckChem library (Figure 5-5A) and AZD2014 from the metabolic library (Figure 5-5B), for example) and would thus be of limited use as single agents, although their effectiveness in potential combination therapy remains to be studied.

Across both drug libraries, the statins – fluvastatin, pitavastatin and simvastatin – stood out, and were between 1.5 to 3 times more effective at killing the PARP4 knockdown cells (Figure 5-5A, B).

Indeed, through follow-up secondary screening experiments, I observed close to 2-fold reduction in IC50 of fluvastatin and simvastatin in the PARP4 knockdown cells, although the decrease in absolute IC50 magnitude was admittedly small at 0.45µM for fluvastatin and 0.95µM for simvastatin (Figure 5-5C, D).

Nevertheless, the initial drug screening followed by secondary validation serve as proof of concept that it is possible to identify drugs that preferentially target PARP4-low cancer cells. Further screening with other compound libraries could thus potentially uncover other more effective hits.

Having shown that splicing is perturbed in PARP4-low cancer cells, drugs modulating splicing should be included in future screens.



	shControl	shPARP4 #1
IC50 (µM)	0.75	0.30

	shControl	shPARP4 #1
IC50 (µM)	1.78	0.83

Figure 5-5: Results of drug screening and secondary validation

A, B) Results of drug screening conducted using the **A)** SelleckChem anti-cancer library and **B)** metabolic drug library. Drugs were added with the kind help of Mr. Matan Thangavelu Thangavelu. SAECK shControl and shPARP4 #1 cells were incubated with the respective drugs at a final concentration of 1µM for 72h. Following which cell viability was determined by the Cell Titer-Glo assay and normalized to the DMSO controls. Drugs yielding a minimally 12.5% greater relative viability in control versus PARP4-depleted cells were included in the figures. **C, D)** Secondary validation was performed on SAECK shControl and shPARP4 #1 cells using **C)** fluvastatin and **D)** simvastatin. The respective IC50 values were determined and indicated in the table below each graph. Data represent the mean ± s.e.m., n = 3.

5.7 To determine the role of PARP4 in other cancers

Having established PARP4's tumor suppressive role in LUAD, it would be interesting to examine its relevance in other cancers. Earlier efforts to examine PARP4's role as a cancer susceptibility gene in breast cancers used proliferation assays to measure effects of PARP4 loss in two breast cancer cell lines, and resulted in discordant observations (Ikeda et al., 2016; Prawira et al., 2019). Thus, to examine PARP4's potential tumor-suppressive role in other cancer types, the depletion of PARP4 from representative cell lines should be followed with *in vitro* and *in vivo* tumorigenicity assays. Should PARP4 be relevant in other cancer types, our findings and therapeutic opportunities for PARP4 could be transferrable to these other cancers.

CHAPTER 6: Conclusion

In summary, we have uncovered PARP4 as a novel tumor suppressor in LUAD based on WES sequencing data from the largest Asian LUAD cohort available to date. In addition to bioinformatics-based prediction methods, we performed *in vitro* and *in vivo* functional validation to show that PARP4 indeed plays a tumor suppressive role in various lung cancer cell line models. Furthermore, a recurrent I1039T mutation was reported in our cohort and we subsequently observed that it reduces PARP4 protein levels.

In addition to its known association with the vault complex, we have identified novel roles of PARP4 in regulating tumorigenicity and controlling splicing, thereby successfully ascribing novel functions to a relatively under-studied protein.

Finally, hnRNPM was identified and validated as a novel PARP4 interaction partner and we showed that loss of hnRNPM results in several phenotypes that are shared with PARP4 loss. Importantly, loss of hnRNPM also contributes to LUAD tumorigenicity.

This work points to the value of cohort sequencing studies in identifying new cancer-associated genes that could have therapeutical applications. Furthermore, it highlights the importance of having representative cohorts in terms of ethnicity and size. As PARP4 is a relatively little-known gene in the context of cancer, there is still much room to explore PARP4's roles in modulating tumorigenicity.

CHAPTER 7: References

Adolf, F., Rhiel, M., Reckmann, I., and Wieland, F.T. (2016). Sec24C/D-isoform–specific sorting of the preassembled ER–Golgi Q-SNARE complex. Mol. Biol. Cell *27*, 2697–2707. https://doi.org/10.1091/mbc.E16-04-0229.

Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. *Chapter 7*, Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249. https://doi.org/10.1038/nmeth0410-248.

Alimirzaie, S., Mohamadkhani, A., Masoudi, S., Sellars, E., Boffetta, P., Malekzadeh, R., Akbari, M.R., and Pourshams, A. (2018). Mutations in Known and Novel cancer Susceptibility Genes in Young Patients with Pancreatic Cancer. Arch. Iran. Med. *21*, 228–233. .

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., and Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. Genome Biol. *21*, 30. https://doi.org/10.1186/s13059-020-1935-5.

Amé, J.-C., Spenlehauer, C., and de Murcia, G. (2004). The PARP superfamily. BioEssays News Rev. Mol. Cell. Dev. Biol. *26*, 882–893. https://doi.org/10.1002/bies.20085.

Anczuków, O., Rosenberg, A.Z., Akerman, M., Das, S., Zhan, L., Karni, R., Muthuswamy, S.K., and Krainer, A.R. (2012). The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. Nat. Struct. Mol. Biol. *19*, 220–228. https://doi.org/10.1038/nsmb.2207.

Ayyappan, V., Wat, R., Barber, C., Vivelo, C.A., Gauch, K., Visanpattanasin, P., Cook, G., Sazeides, C., and Leung, A.K.L. (2021). ADPriboDB 2.0: an updated database of ADP-ribosylated proteins. Nucleic Acids Res. *49*, D261–D265. https://doi.org/10.1093/nar/gkaa941.

Azarm, K., and Smith, S. (2020). Nuclear PARPs and genome integrity. Genes Dev. *34*, 285–301. https://doi.org/10.1101/gad.334730.119.

Bai, P. (2015). Biology of Poly(ADP-Ribose) Polymerases: The Factotums of Cell Maintenance. Mol. Cell *58*, 947–958. https://doi.org/10.1016/j.molcel.2015.01.034.

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al.

(2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371-385.e18. https://doi.org/10.1016/j.cell.2018.02.060.

Bartholomew, C., Eastlake, L., Dunn, P., and Yiannakis, D. (2017). EGFR targeted therapy in lung cancer; an evolving story. Respir. Med. Case Rep. *20*, 137–140. https://doi.org/10.1016/j.rmcr.2017.01.016.

Bauman, J., Jearawiriyapaisarn, N., and Kole, R. (2009). Therapeutic Potential of Splice-Switching Oligonucleotides. Oligonucleotides *19*, 1–13. https://doi.org/10.1089/oli.2008.0161.

Beattie, T.L., Zhou, W., Robinson, M.O., and Harrington, L. (1998). Reconstitution of human telomerase activity in vitro. Curr. Biol. CB *8*, 177–180. https://doi.org/10.1016/s0960-9822(98)70067-3.

Belgrader, P., Cheng, J., Zhou, X., Stephenson, L.S., and Maquat, L.E. (1994). Mammalian nonsense codons can be cis effectors of nuclear mRNA half-life. Mol. Cell. Biol. *14*, 8219–8228. https://doi.org/10.1128/mcb.14.12.8219-8228.1994.

Bell, T.J., Miyashiro, K.Y., Sul, J.-Y., McCullough, R., Buckley, P.T., Jochems, J., Meaney, D.F., Haydon, P., Cantor, C., Parsons, T.D., et al. (2008). Cytoplasmic BK(Ca) channel intron-containing mRNAs contribute to the intrinsic excitability of hippocampal neurons. Proc. Natl. Acad. Sci. U. S. A. *105*, 1901–1906. https://doi.org/10.1073/pnas.0711796105.

Berger, W., Spiegl-Kreinecker, S., Buchroithner, J., Elbling, L., Pirker, C., Fischer, J., and Micksche, M. (2001). Overexpression of the human major vault protein in astrocytic brain tumor cells. Int. J. Cancer *94*, 377–382. https://doi.org/10.1002/ijc.1486.

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J.C., Huang, J.H., Alexander, S., et al. (2007). Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc. Natl. Acad. Sci. U. S. A. *104*, 20007–20012. https://doi.org/10.1073/pnas.0710052104.

Blanco, R., Iwakawa, R., Tang, M., Kohno, T., Angulo, B., Pio, R., Montuenga, L.M., Minna, J.D., Yokota, J., and Sanchez-Cespedes, M. (2009). A gene-alteration profile of human lung cancer cell lines. Hum. Mutat. *30*, 1199–1206. https://doi.org/10.1002/humu.21028.

Bong, I.P.N., Ng, C.C., Fakiruddin, S.K., Lim, M.N., and Zakaria, Z. (2016). Small interfering RNA-mediated silencing of nicotinamide phosphoribosyltransferase (NAMPT) and lysosomal trafficking regulator (LYST) induce growth inhibition and apoptosis in human multiple myeloma cells: A preliminary study. Bosn. J. Basic Med. Sci. *16*, 268–275. https://doi.org/10.17305/bjbms.2016.1568.

Bonnal, S.C., López-Oreja, I., and Valcárcel, J. (2020). Roles and mechanisms of alternative splicing in cancer — implications for care. Nat. Rev. Clin. Oncol. *17*, 457–474. https://doi.org/10.1038/s41571-020-0350-x.

Bork, P., Hofmann, K., Bucher, P., Neuwald, A.F., Altschul, S.F., and Koonin, E.V. (1997). A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *11*, 68–76. .

Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K.W., Vogelstein, B., and Nowak, M.A. (2010). Accumulation of driver and passenger mutations during tumor progression. Proc. Natl. Acad. Sci. *107*, 18545–18550. https://doi.org/10.1073/pnas.1010978107.

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., et al. (2008). The potential and challenges of nanopore sequencing. Nat. Biotechnol. *26*, 1146–1153. https://doi.org/10.1038/nbt.1495.

Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res. *24*, 1774–1786. https://doi.org/10.1101/gr.177790.114.

Bromberg, Y., and Rost, B. (2009). Correlating protein function and stability through the analysis of single amino acid substitutions. BMC Bioinformatics *10*, S8. https://doi.org/10.1186/1471-2105-10-S8-S8.

Buckley, P.T., Lee, M.T., Sul, J.-Y., Miyashiro, K.Y., Bell, T.J., Fisher, S.A., Kim, J., and Eberwine, J. (2011). Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons. Neuron *69*, 877–884. https://doi.org/10.1016/j.neuron.2011.02.028.

Callebaut, I., and Mornon, J.P. (1997). From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. FEBS Lett. *400*, 25–30. https://doi.org/10.1016/s0014-5793(96)01312-9.

Campbell, J.D., Alexandrov, A., Kim, J., Wala, J., Berger, A.H., Pedamallu, C.S., Shukla, S.A., Guo, G., Brooks, A.N., Murray, B.A., et al. (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat. Genet. *48*, 607–616. https://doi.org/10.1038/ng.3564.

Cangemi, R., Mensah, A., Albertini, V., Jain, A., Mello-Grand, M., Chiorino, G., Catapano, C.V., and Carbone, G.M. (2008). Reduced expression and tumor suppressor function of the ETS transcription factor ESE-3 in prostate cancer. Oncogene *27*, 2877–2885. https://doi.org/10.1038/sj.onc.1210953.

Cardnell, R.J.G., Behrens, C., Diao, L., Fan, Y., Tang, X., Tong, P., Minna, J.D., Mills, G.B., Heymach, J.V., Wistuba, I.I., et al. (2015). An Integrated Molecular Analysis of Lung Adenocarcinomas Identifies Potential Therapeutic Targets among TTF1-Negative Tumors, Including DNA Repair Proteins and Nrf2. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *21*, 3480–3491. https://doi.org/10.1158/1078-0432.CCR-14-3286.

Carrot-Zhang, J., Yao, X., Devarakonda, S., Deshpande, A., Damrauer, J.S., Silva, T.C., Wong, C.K., Choi, H.Y., Felau, I., Robertson, A.G., et al. (2021). Whole-genome characterization of lung adenocarcinomas lacking alterations in the RTK/RAS/RAF pathway. Cell Rep. *34*, 108707. https://doi.org/10.1016/j.celrep.2021.108707.

Carter-O'Connell, I., Jin, H., Morgan, R.K., David, L.L., and Cohen, M.S. (2014). Engineering the substrate specificity of ADP-ribosyltransferases for identifying direct protein targets. J. Am. Chem. Soc. *136*, 5201–5204. https://doi.org/10.1021/ja412897a.

Carter-O'Connell, I., Jin, H., Morgan, R.K., Zaja, R., David, L.L., Ahel, I., and Cohen, M.S. (2016). Identifying Family-Member-Specific Targets of Mono-ARTDs by Using a Chemical Genetics Approach. Cell Rep. *14*, 621–631. https://doi.org/10.1016/j.celrep.2015.12.045.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discov. *2*, 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095.

Challa, S., Khulpateea, B.R., Nandu, T., Camacho, C.V., Ryu, K.W., Chen, H., Peng, Y., Lea, J.S., and Kraus, W.L. (2021). Ribosome ADP-ribosylation inhibits translation and maintains proteostasis in cancers. Cell *184*, 4531-4546.e26. https://doi.org/10.1016/j.cell.2021.07.005.

Chao, T.-K., Huang, T.-S., Liao, Y.-P., Huang, R.-L., Su, P.-H., Shen, H.-Y., Lai, H.-C., and Wang, Y.-C. (2017). Pyruvate kinase M2 is a poor prognostic marker of and a therapeutic target in ovarian cancer. PloS One *12*, e0182166. https://doi.org/10.1371/journal.pone.0182166.

Chatterjee, S., and Pal, J.K. (2009). Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. Biol. Cell *101*, 251–262. https://doi.org/10.1042/BC20080104.

Chen, C.-Y., Chen, J., He, L., and Stiles, B.L. (2018). PTEN: Tumor Suppressor and Metabolic Regulator. Front. Endocrinol. *9*, 338. https://doi.org/10.3389/fendo.2018.00338.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128. https://doi.org/10.1186/1471-2105-14-128.

Chen, H., Li, J., Wang, Y., Ng, P.K.-S., Tsang, Y.H., Shaw, K.R., Mills, G.B., and Liang, H. (2020a). Comprehensive assessment of computational algorithms in predicting cancer driver mutations. Genome Biol. *21*, 43. https://doi.org/10.1186/s13059-020-01954-z.

Chen, J., Yang, H., Teo, A.S.M., Amer, L.B., Sherbaf, F.G., Tan, C.Q., Alvarez, J.J.S., Lu, B., Lim, J.Q., Takano, A., et al. (2020b). Genomic landscape of lung adenocarcinoma in East Asians. Nat. Genet. *52*, 177–186. https://doi.org/10.1038/s41588-019-0569-6.

Chen, S., Zhang, J., Duan, L., Zhang, Y., Li, C., Liu, D., Ouyang, C., Lu, F., and Liu, X. (2014). Identification of HnRNP M as a novel biomarker for colorectal carcinoma by quantitative proteomics. Am. J. Physiol.-Gastrointest. Liver Physiol. *306*, G394–G403. https://doi.org/10.1152/ajpgi.00328.2013.

Chen, T., Wang, Z., Zhou, W., Chong, Z., Meric-Bernstam, F., Mills, G.B., and Chen, K. (2016). Hotspot mutations delineating diverse mutational signatures and biological utilities across cancer types. BMC Genomics *17*, 394. https://doi.org/10.1186/s12864-016-2727-x.

Chen, T.-M., Lai, M.-C., Li, Y.-H., Chan, Y.-L., Wu, C.-H., Wang, Y.-M., Chien, C.-W., Huang, S.-Y., Sun, H.S., and Tsai, S.-J. (2019). hnRNPM induces translation switch under hypoxia to promote colon cancer development. EBioMedicine *41*, 299–309. https://doi.org/10.1016/j.ebiom.2019.02.059.

Chen, Y.-J., Roumeliotis, T.I., Chang, Y.-H., Chen, C.-T., Han, C.-L., Lin, M.-H., Chen, H.-W., Chang, G.-C., Chang, Y.-L., Wu, C.-T., et al. (2020c). Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression. Cell *182*, 226-244.e17. https://doi.org/10.1016/j.cell.2020.06.012.

Chen, Y.-L., Huang, W.-C., Yao, H.-L., Chen, P.-M., Lin, P.-Y., Feng, F.-Y., and Chu, P.-Y. (2017). Down-regulation of RASA1 Is Associated with Poor Prognosis in Human Hepatocellular Carcinoma. Anticancer Res. *37*, 781–785. https://doi.org/10.21873/anticanres.11377.

Cirello, V., Colombo, C., Pogliaghi, G., Proverbio, M.C., Rossi, S., Mussani, E., Tosi, D., Bulfamante, G., Bonoldi, E., Gherardi, G., et al. (2019). Genetic variants of PARP4 gene and PARP4P2 pseudogene in patients with multiple primary tumors including thyroid cancer. Mutat. Res. *816–818*, 111672. https://doi.org/10.1016/j.mrfmmm.2019.111672.

Cohen, M.S., and Chang, P. (2018). Insights into the biogenesis, function, and regulation of ADP-ribosylation. Nat. Chem. Biol. *14*, 236–243. https://doi.org/10.1038/nchembio.2568.

Collier, O., Stoven, V., and Vert, J.-P. (2019). LOTUS: A single- and multitask machine learning algorithm for the prediction of cancer driver genes. PLOS Comput. Biol. *15*, e1007381. https://doi.org/10.1371/journal.pcbi.1007381.

Corkery, D.P., Holly, A.C., Lahsaee, S., and Dellaire, G. (2015). Connecting the speckles: Splicing kinases and their role in tumorigenesis and treatment response. Nucleus *6*, 279–288. https://doi.org/10.1080/19491034.2015.1062194.

Daniels, C.M., Ong, S.-E., and Leung, A.K.L. (2015). The Promise of Proteomics for the Study of ADP-Ribosylation. Mol. Cell *58*, 911–924. https://doi.org/10.1016/j.molcel.2015.06.012.

Datar, K.V., Dreyfuss, G., and Swanson, M.S. (1993). The human hnRNP M proteins: identification of a methionine/arginine-rich repeat motif in ribonucleoproteins. Nucleic Acids Res. *21*, 439–446. https://doi.org/10.1093/nar/21.3.439.

Daugherty, M.D., Young, J.M., Kerns, J.A., and Malik, H.S. (2014). Rapid Evolution of PARP Genes Suggests a Broad Role for ADP-Ribosylation in Host-Virus Conflicts. PLOS Genet. *10*, e1004403. https://doi.org/10.1371/journal.pgen.1004403.

Davis, A.G., Shima, T., Wang, R., Zheng, D., Tian, B., and Zhang, D.-E. (2018). FIP1L1 Regulates Alternative Polyadenylation of Leukemia-Associated Genes in Acute Myeloid Leukemia. Blood *132*, 3882. https://doi.org/10.1182/blood-2018-99-119582.

Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell *155*, 948–962. https://doi.org/10.1016/j.cell.2013.10.011.

Deamer, D.W., and Branton, D. (2002). Characterization of nucleic acids by nanopore analysis. Acc. Chem. Res. *35*, 817–825. https://doi.org/10.1021/ar000138m.

Dees, N.D., Zhang, Q., Kandoth, C., Wendl, M.C., Schierding, W., Koboldt, D.C., Mooney, T.B., Callaway, M.B., Dooling, D., Mardis, E.R., et al. (2012). MuSiC: Identifying mutational significance in cancer genomes. Genome Res. *22*, 1589–1598. https://doi.org/10.1101/gr.134635.111.

Dela Cruz, C.S., Tanoue, L.T., and Matthay, R.A. (2011). Lung Cancer: Epidemiology, Etiology, and Prevention. Clin. Chest Med. *32*. https://doi.org/10.1016/j.ccm.2011.09.001.

Ding, L., Getz, G., Wheeler, D.A., Mardis, E.R., McLellan, M.D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D.M., Morgan, M.B., et al. (2008). Somatic mutations affect key pathways in lung adenocarcinoma. Nature *455*, 1069–1075. https://doi.org/10.1038/nature07423.

Dolma, S., Lessnick, S.L., Hahn, W.C., and Stockwell, B.R. (2003). Identification of genotype-selective antitumor agents using synthetic lethal chemical screening in engineered human tumor cells. Cancer Cell *3*, 285–296. https://doi.org/10.1016/S1535-6108(03)00050-3.

Dreyfuss, G., Swanson, M.S., and Piñol-Roma, S. (1988). Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. Trends Biochem. Sci. *13*, 86–91. https://doi.org/10.1016/0968-0004(88)90046-1.

Dvinge, H., and Bradley, R.K. (2015). Widespread intron retention diversifies most cancer transcriptomes. Genome Med. *7*, 45. https://doi.org/10.1186/s13073-015-0168-9.

Emmott, E., and Goodfellow, I. (2014). Identification of Protein Interaction Partners in Mammalian Cells Using SILAC-immunoprecipitation Quantitative Proteomics. J. Vis. Exp. JoVE 51656. https://doi.org/10.3791/51656.

Feng, X., and Koh, D.W. (2013). Chapter Five - Roles of Poly(ADP-Ribose) Glycohydrolase in DNA Damage and Apoptosis. In International Review of Cell and Molecular Biology, K.W. Jeon, ed. (Academic Press), pp. 227–281.

Forde, P.M., and Ettinger, D.S. (2013). Targeted therapy for non-small-cell lung cancer: past, present and future. Expert Rev. Anticancer Ther. *13*, 745–758. https://doi.org/10.1586/era.13.47.

Freitas, N., and Cunha, C. (2009). Mechanisms and Signals for the Nuclear Import of Proteins. Curr. Genomics *10*, 550–557. https://doi.org/10.2174/138920209789503941.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat. Rev. Cancer *4*, 177–183. https://doi.org/10.1038/nrc1299.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1. https://doi.org/10.1126/scisignal.2004088.

Gazdar, A. (2009). Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. Oncogene *28*, S24–S31. https://doi.org/10.1038/onc.2009.198.

Geuens, T., Bouhy, D., and Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. Hum. Genet. *135*, 851–867. https://doi.org/10.1007/s00439-016-1683-5.

Ghigna, C., Valacca, C., and Biamonti, G. (2008). Alternative splicing and tumor progression. Curr. Genomics *9*, 556–570. https://doi.org/10.2174/138920208786847971.

Gibson, B.A., Zhang, Y., Jiang, H., Hussey, K.M., Shrimp, J.H., Lin, H., Schwede, F., Yu, Y., and Kraus, W.L. (2016). Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. Science *353*, 45–50. https://doi.org/10.1126/science.aaf7865.

Gonzalez-Perez, A., and Lopez-Bigas, N. (2012). Functional impact bias reveals cancer drivers. Nucleic Acids Res. *40*, e169. https://doi.org/10.1093/nar/gks743.

Gordon, L.G., White, N.M., Elliott, T.M., Nones, K., Beckhouse, A.G., Rodriguez-Acevedo, A.J., Webb, P.M., Lee, X.J., Graves, N., and Schofield, D.J. (2020). Estimating the costs of genomic sequencing in cancer control. BMC Health Serv. Res. *20*, 492. https://doi.org/10.1186/s12913-020-05318-y.

Govindan, R., Ding, L., Griffith, M., Subramanian, J., Dees, N.D., Kanchi, K.L., Maher, C.A., Fulton, R., Fulton, L., Wallis, J., et al. (2012). Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell *150*, 1121–1134. https://doi.org/10.1016/j.cell.2012.08.024.

Graham, N.A., Minasyan, A., Lomova, A., Cass, A., Balanis, N.G., Friedman, M., Chan, S., Zhao, S., Delgado, A., Go, J., et al. (2017). Recurrent patterns of DNA copy number alterations in tumors reflect metabolic selection pressures. Mol. Syst. Biol. *13*, 914. https://doi.org/10.15252/msb.20167159.

Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E., et al. (2011). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat. Genet. *44*, 53–57. https://doi.org/10.1038/ng.1031.

Gupte, R., Liu, Z., and Kraus, W.L. (2017). PARPs and ADP-ribosylation: recent advances linking molecular functions to biological outcomes. Genes Dev. *31*, 101–126. https://doi.org/10.1101/gad.291518.116.

Győrffy, B., Surowiak, P., Budczies, J., and Lánczky, A. (2013). Online survival analysis software to assess the prognostic value of biomarkers using

transcriptomic data in non-small-cell lung cancer. PloS One *8*, e82241. https://doi.org/10.1371/journal.pone.0082241.

Ha, S.Y., Choi, S.-J., Cho, J.H., Choi, H.J., Lee, J., Jung, K., Irwin, D., Liu, X., Lira, M.E., Mao, M., et al. (2015). Lung cancer in never-smoker Asian females is driven by oncogenic mutations, most often involving EGFR. Oncotarget *6*, 5465–5474. https://doi.org/10.18632/oncotarget.2925.

Haber, D.A., and Settleman, J. (2007). Drivers and passengers. Nature *446*, 145–146. https://doi.org/10.1038/446145a.

Hahn, W.C., and Weinberg, R.A. (2002). Modelling the molecular circuitry of cancer. Nat. Rev. Cancer *2*, 331–341. https://doi.org/10.1038/nrc795.

Hahn, W.C., Counter, C.M., Lundberg, A.S., Beijersbergen, R.L., Brooks, M.W., and Weinberg, R.A. (1999). Creation of human tumour cells with defined genetic elements. Nature *400*, 464–468. https://doi.org/10.1038/22780.

Hahn, W.C., Dessain, S.K., Brooks, M.W., King, J.E., Elenbaas, B., Sabatini, D.M., DeCaprio, J.A., and Weinberg, R.A. (2002). Enumeration of the Simian Virus 40 Early Region Elements Necessary for Human Cell Transformation. Mol. Cell. Biol. *22*, 2111–2123. https://doi.org/10.1128/MCB.22.7.2111-2123.2002.

Haigis, K.M. (2017). KRAS Alleles: The Devil Is In The Detail. Trends Cancer *3*, 686–697. https://doi.org/10.1016/j.trecan.2017.08.006.

Hainaut, P., and Pfeifer, G.P. (2016). Somatic TP53 Mutations in the Era of Genome Sequencing. Cold Spring Harb. Perspect. Med. *6*, a026179. https://doi.org/10.1101/cshperspect.a026179.

Hamill, D.R., and Suprenant, K.A. (1997). Characterization of the sea urchin major vault protein: a possible role for vault ribonucleoprotein particles in nucleocytoplasmic transport. Dev. Biol. *190*, 117–128. https://doi.org/10.1006/dbio.1997.8676.

Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. Biochem. J. *430*, 379–392. https://doi.org/10.1042/BJ20100396.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

Harrington, L., McPhail, T., Mar, V., Zhou, W., Oulton, R., Bass, M.B., Arruda, I., and Robinson, M.O. (1997). A mammalian telomerase-associated protein. Science *275*, 973–977. https://doi.org/10.1126/science.275.5302.973.

Harvey, S.E., Xu, Y., Lin, X., Gao, X.D., Qiu, Y., Ahn, J., Xiao, X., and Cheng, C. (2018). Coregulation of alternative splicing by hnRNPM and ESRP1 during EMT. RNA N. Y. N *24*, 1326–1338. https://doi.org/10.1261/rna.066712.118.

Havens, M.A., and Hastings, M.L. (2016). Splice-switching antisense oligonucleotides as therapeutic drugs. Nucleic Acids Res. *44*, 6549–6563. https://doi.org/10.1093/nar/gkw533.

Helfrich, B.A., Raben, D., Varella-Garcia, M., Gustafson, D., Chan, D.C., Bemis, L., Coldren, C., Barón, A., Zeng, C., Franklin, W.A., et al. (2006). Antitumor activity of the epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor gefitinib (ZD1839, Iressa) in non-small cell lung cancer cell lines correlates with gene copy number and EGFR mutations but not EGFR protein levels. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *12*, 7117–7125. https://doi.org/10.1158/1078-0432.CCR-06-0760.

Herbst, R.S., Heymach, J.V., and Lippman, S.M. (2008). Lung Cancer. N. Engl. J. Med. *359*, 1367–1380. https://doi.org/10.1056/NEJMra0802714.

Herrmann, C., Golkaramnay, E., Inman, E., Rome, L., and Volknandt, W. (1999). Recombinant Major Vault Protein Is Targeted to Neuritic Tips of PC12 Cells. J. Cell Biol. *144*, 1163–1172. .

Hirsch, F.R., Scagliotti, G.V., Mulshine, J.L., Kwon, R., Curran, W.J., Wu, Y.-L., and Paz-Ares, L. (2017). Lung cancer: current therapies and new targeted treatments. Lancet Lond. Engl. *389*, 299–311. https://doi.org/10.1016/S0140-6736(16)30958-8.

Ho, J.S., Di Tullio, F., Schwarz, M., Low, D., Incarnato, D., Gay, F., Tabaglio, T., Zhang, J., Wollmann, H., Chen, L., et al. (2021). HNRNPM controls circRNA biogenesis and splicing fidelity to sustain cancer cell fitness. ELife *10*, e59654. https://doi.org/10.7554/eLife.59654.

Horos, R., Büscher, M., Kleinendorst, R., Alleaume, A.-M., Tarafder, A.K., Schwarzl, T., Dziuba, D., Tischer, C., Zielonka, E.M., Adak, A., et al. (2019). The Small Non-coding Vault RNA1-1 Acts as a Riboregulator of Autophagy. Cell *176*, 1054-1067.e12. https://doi.org/10.1016/j.cell.2019.01.030.

Hou, J.P., and Ma, J. (2014). DawnRank: discovering personalized driver genes in cancer. Genome Med. *6*, 56. https://doi.org/10.1186/s13073-014-0056-8.

Hou, P., Jia, P., Yang, K., Li, Z., Tian, T., Lin, Y., Zeng, W., Xing, F., Chen, Y., Li, C., et al. (2021). An unconventional role of an ASB family protein in NF-κB activation and inflammatory response during microbial infection and colitis. Proc. Natl. Acad. Sci. U. S. A. *118*, e2015416118. https://doi.org/10.1073/pnas.2015416118.

Hovhannisyan, R.H., and Carstens, R.P. (2007). Heterogeneous ribonucleoprotein m is a splicing regulatory protein that can enhance or silence splicing of alternatively spliced exons. J. Biol. Chem. *282*, 36265–36274. https://doi.org/10.1074/jbc.M704188200.

Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhai, J., et al. (2021). Ensembl 2021. Nucleic Acids Res. *49*, D884–D891. https://doi.org/10.1093/nar/gkaa942.

Hu, X., Harvey, S.E., Zheng, R., Lyu, J., Grzeskowiak, C.L., Powell, E., Piwnica-Worms, H., Scott, K.L., and Cheng, C. (2020). The RNA-binding protein AKAP8 suppresses tumor metastasis by antagonizing EMT-associated alternative splicing. Nat. Commun. *11*, 486. https://doi.org/10.1038/s41467-020-14304-1.

Hu, Y., Stephen, A.G., Cao, J., Tanzer, L.R., Slapak, C.A., Harrison, S.D., Devanarayan, V., Dantzig, A.H., Starling, J.J., Rome, L.H., et al. (2002). A very early induction of major vault protein accompanied by increased drug resistance in U-937 cells. Int. J. Cancer *97*, 149–156. https://doi.org/10.1002/ijc.1590.

Huang, L., and Fu, L. (2015). Mechanisms of resistance to EGFR tyrosine kinase inhibitors. Acta Pharm. Sin. B *5*, 390–401. https://doi.org/10.1016/j.apsb.2015.07.001.

Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., et al. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep. *1*, 167–178. https://doi.org/10.1016/j.celrep.2012.02.001.

Huffman, K.E., and Corey, D.R. (2005). Major vault protein does not play a role in chemoresistance or drug localization in a non-small cell lung cancer cell line. Biochemistry *44*, 2253–2261. https://doi.org/10.1021/bi047948g.

Hussain, S., Sajini, A.A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., Paramor, M., Gleeson, J.G., Odom, D.T., Ule, J., et al. (2013). NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell Rep. *4*, 255–261. https://doi.org/10.1016/j.celrep.2013.06.029.

Ikeda, Y., Kiyotani, K., Yew, P.Y., Kato, T., Tamura, K., Yap, K.L., Nielsen, S.M., Mester, J.L., Eng, C., Nakamura, Y., et al. (2016). Germline PARP4 mutations in patients with primary thyroid and breast cancers. Endocr. Relat. Cancer *23*, 171–179. https://doi.org/10.1530/ERC-15-0359.

Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., et al. (2012).

Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell *150*, 1107–1120. https://doi.org/10.1016/j.cell.2012.08.029.

Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. CA. Cancer J. Clin. *61*, 69–90. https://doi.org/10.3322/caac.20107.

Jensen, M.A., Wilkinson, J.E., and Krainer, A.R. (2014). Splicing factor SRSF6 promotes hyperplasia of sensitized skin. Nat. Struct. Mol. Biol. *21*, 189–197. https://doi.org/10.1038/nsmb.2756.

Ji, Y., and Tulin, A.V. (2009). Poly(ADP-ribosyl)ation of heterogeneous nuclear ribonucleoproteins modulates splicing. Nucleic Acids Res. *37*, 3501–3513. https://doi.org/10.1093/nar/gkp218.

Ji, Y., and Tulin, A.V. (2013). Post-Transcriptional Regulation by Poly(ADP-ribosyl)ation of the RNA-Binding Proteins. Int. J. Mol. Sci. *14*, 16168–16183. https://doi.org/10.3390/ijms140816168.

Jiang, L., Zheng, J., Kwan, J.S.H., Dai, S., Li, C., Li, M.J., Yu, B., To, K.F., Sham, P.C., Zhu, Y., et al. (2019). WITER: a powerful method for estimation of cancer-driver genes using a weighted iterative regression modelling background mutation counts. Nucleic Acids Res. *47*, e96. https://doi.org/10.1093/nar/gkz566.

Jordan, E.J., Kim, H.R., Arcila, M.E., Barron, D., Chakravarty, D., Gao, J., Chang, M.T., Ni, A., Kundra, R., Jonsson, P., et al. (2017). Prospective comprehensive molecular characterization of lung adenocarcinomas for efficient patient matching to approved and emerging therapies. Cancer Discov. *7*, 596–609. https://doi.org/10.1158/2159-8290.CD-16-1337.

Jung, H., Lee, D., Lee, J., Park, D., Kim, Y.J., Park, W.-Y., Hong, D., Park, P.J., and Lee, E. (2015). Intron retention is a widespread mechanism of tumor-suppressor inactivation. Nat. Genet. *47*, 1242–1248. https://doi.org/10.1038/ng.3414.

Jungmichel, S., Rosenthal, F., Altmeyer, M., Lukas, J., Hottiger, M.O., and Nielsen, M.L. (2013). Proteome-wide Identification of Poly(ADP-Ribosyl)ation Targets in Different Genotoxic Stress Responses. Mol. Cell *52*, 272–285. https://doi.org/10.1016/j.molcel.2013.08.026.

Kadara, H., Choi, M., Zhang, J., Parra, E.R., Rodriguez-Canales, J., Gaffney, S.G., Zhao, Z., Behrens, C., Fujimoto, J., Chow, C., et al. (2017). Whole-exome sequencing and immune profiling of early-stage lung adenocarcinoma with fully annotated clinical follow-up. Ann. Oncol. *28*, 75–82. https://doi.org/10.1093/annonc/mdw436.

Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984). A short amino acid sequence able to specify nuclear location. Cell *39*, 499–509. https://doi.org/10.1016/0092-8674(84)90457-4.

Kamburov, A., Lawrence, M.S., Polak, P., Leshchiner, I., Lage, K., Golub, T.R., Lander, E.S., and Getz, G. (2015). Comprehensive assessment of cancer missense mutation clustering in protein structures. Proc. Natl. Acad. Sci. *112*, E5486–E5495. https://doi.org/10.1073/pnas.1516373112.

Kanai, M., Tong, W.-M., Sugihara, E., Wang, Z.-Q., Fukasawa, K., and Miwa, M. (2003). Involvement of Poly(ADP-Ribose) Polymerase 1 and Poly(ADP-Ribosyl)ation in Regulation of Centrosome Function. Mol. Cell. Biol. https://doi.org/10.1128/MCB.23.7.2451-2462.2003.

Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat. Struct. Mol. Biol. *14*, 185–193. https://doi.org/10.1038/nsmb1209.

Katsyv, I., Wang, M., Song, W.M., Zhou, X., Zhao, Y., Park, S., Zhu, J., Zhang, B., and Irie, H.Y. (2016). EPRS is a critical regulator of cell proliferation and estrogen signaling in ER+ breast cancer. Oncotarget *7*, 69592–69605. https://doi.org/10.18632/oncotarget.11870.

Kawaguchi, T., Matsumura, A., Fukai, S., Tamura, A., Saito, R., Zell, J.A., Maruyama, Y., Ziogas, A., Kawahara, M., and Ignatius Ou, S.-H. (2010). Japanese ethnicity compared with Caucasian ethnicity and never-smoking status are independent favorable prognostic factors for overall survival in non-small cell lung cancer: a collaborative epidemiologic study of the National Hospital Organization Study Group for Lung Cancer (NHSGLC) in Japan and a Southern California Regional Cancer Registry databases. J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer *5*, 1001–1010. https://doi.org/10.1097/JTO.0b013e3181e2f607.

Kawahara, A., Yamamoto, C., Nakashima, K., Azuma, K., Hattori, S., Kashihara, M., Aizawa, H., Basaki, Y., Kuwano, M., Kage, M., et al. (2010). Molecular diagnosis of activating EGFR mutations in non-small cell lung cancer using mutation-specific antibodies for immunohistochemical analysis. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *16*, 3163–3170. https://doi.org/10.1158/1078-0432.CCR-09-3239.

Kedersha, N.L., and Rome, L.H. (1986). Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. J. Cell Biol. *103*, 699–709. https://doi.org/10.1083/jcb.103.3.699.

Kedersha, N.L., Miquel, M.C., Bittner, D., and Rome, L.H. (1990). Vaults. II. Ribonucleoprotein structures are highly conserved among higher and lower eukaryotes. J. Cell Biol. *110*, 895–901. https://doi.org/10.1083/jcb.110.4.895.

Kedersha, N.L., Heuser, J.E., Chugani, D.C., and Rome, L.H. (1991). Vaults. III. Vault ribonucleoprotein particles open into flower-like structures with octagonal symmetry. J. Cell Biol. *112*, 225–235. https://doi.org/10.1083/jcb.112.2.225.

Kershah, S.M., Desouki, M.M., Koterba, K.L., and Rowan, B.G. (2004). Expression of estrogen receptor coregulators in normal and malignant human endometrium. Gynecol. Oncol. *92*, 304–313. .

Kickhoefer, V.A., Rajavel, K.S., Scheffer, G.L., Dalton, W.S., Scheper, R.J., and Rome, L.H. (1998). Vaults are up-regulated in multidrug-resistant cancer cell lines. J. Biol. Chem. *273*, 8971–8974. https://doi.org/10.1074/jbc.273.15.8971.

Kickhoefer, V.A., Siva, A.C., Kedersha, N.L., Inman, E.M., Ruland, C., Streuli, M., and Rome, L.H. (1999a). The 193-Kd Vault Protein, Vparp, Is a Novel Poly(Adp-Ribose) Polymerase. J. Cell Biol. *146*, 917–928. .

Kickhoefer, V.A., Stephen, A.G., Harrington, L., Robinson, M.O., and Rome, L.H. (1999b). Vaults and Telomerase Share a Common Subunit, TEP1 *. J. Biol. Chem. *274*, 32712–32717. https://doi.org/10.1074/jbc.274.46.32712.

Kickhoefer, V.A., Liu, Y., Kong, L.B., Snow, B.E., Stewart, P.L., Harrington, L., and Rome, L.H. (2001). The Telomerase/vault-associated protein TEP1 is required for vault RNA stability and its association with the vault particle. J. Cell Biol. *152*, 157–164. https://doi.org/10.1083/jcb.152.1.157.

Kohno, T., Nakaoku, T., Tsuta, K., Tsuchihara, K., Matsumoto, S., Yoh, K., and Goto, K. (2015). Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer. Transl. Lung Cancer Res. *4*, 156–164. https://doi.org/10.3978/j.issn.2218-6751.2014.11.11.

Kohroki, J., Nishiyama, T., Nakamura, T., and Masuho, Y. (2005). ASB proteins interact with Cullin5 and Rbx2 to form E3 ubiquitin ligase complexes. FEBS Lett. *579*, 6796–6802. https://doi.org/10.1016/j.febslet.2005.11.016.

Kolli, S., Zito, C.I., Mossink, M.H., Wiemer, E.A.C., and Bennett, A.M. (2004). The major vault protein is a novel substrate for the tyrosine phosphatase SHP-2 and scaffold protein in epidermal growth factor signaling. J. Biol. Chem. *279*, 29374–29385. https://doi.org/10.1074/jbc.M313955200.

Komarov, P.G., Shtil, A.A., Holian, O., Tee, L., Buckingham, L., Mechetner, E.B., Roninson, I.B., and Coon, J.S. (1998). Activation of the LRP (lung resistance-related protein) gene by short-term exposure of human leukemia cells to phorbol ester and cytarabine. Oncol. Res. *10*, 185–192. .

Kono, N., and Arakawa, K. (2019). Nanopore sequencing: Review of potential applications in functional genomics. Dev. Growth Differ. *61*, 316–326. https://doi.org/10.1111/dgd.12608.

Kozlov, G., Vavelyuk, O., Minailiuc, O., Banville, D., Gehring, K., and Ekiel, I. (2006). Solution Structure of a Two-repeat Fragment of Major Vault Protein. J. Mol. Biol. *356*, 444–452. https://doi.org/10.1016/j.jmb.2005.11.064.

Ku, C.-S., and Roukos, D.H. (2013). From next-generation sequencing to nanopore sequencing technology: paving the way to personalized genomic medicine. Expert Rev. Med. Devices *10*, 1–6. https://doi.org/10.1586/erd.12.63.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90-97. https://doi.org/10.1093/nar/gkw377.

Kumar, M., Keller, B., Makalou, N., and Sutton, R.E. (2001). Systematic Determination of the Packaging Limit of Lentiviral Vectors. Hum. Gene Ther. *12*, 1893–1905. https://doi.org/10.1089/104303401753153947.

Kumari, S.R., Mendoza-Alvarez, H., and Alvarez-Gonzalez, R. (1998). Functional interactions of p53 with poly(ADP-ribose) polymerase (PARP) during apoptosis following DNA damage: covalent poly(ADP-ribosyl)ation of p53 by exogenous PARP and noncovalent binding of p53 to the M(r) 85,000 proteolytic fragment. Cancer Res. *58*, 5075–5078. .

Ladanyi, M., and Pao, W. (2008). Lung adenocarcinoma: guiding EGFR-targeted therapy and beyond. Mod. Pathol. *21*, S16–S22. https://doi.org/10.1038/modpathol.3801018.

Laishram, R.S. (2014). Poly(A) polymerase (PAP) diversity in gene expression – Star-PAP vs canonical PAP. FEBS Lett. *588*, 2185–2197. https://doi.org/10.1016/j.febslet.2014.05.029.

Lambacher, N.J., Bruel, A.-L., van Dam, T.J.P., Szymańska, K., Slaats, G.G., Kuhns, S., McManus, G.J., Kennedy, J.E., Gaff, K., Wu, K.M., et al. (2016). TMEM107 recruits ciliopathy proteins to subdomains of the ciliary transition zone and causes Joubert syndrome. Nat. Cell Biol. *18*, 122–131. https://doi.org/10.1038/ncb3273.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218. https://doi.org/10.1038/nature12213.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014).

Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501. https://doi.org/10.1038/nature12912.

Lee, J.J.-K., Park, S., Park, H., Kim, S., Lee, J., Lee, J., Youk, J., Yi, K., An, Y., Park, I.K., et al. (2019). Tracing Oncogene Rearrangements in the Mutational History of Lung Adenocarcinoma. Cell *177*, 1842-1857.e21. https://doi.org/10.1016/j.cell.2019.05.013.

Leidecker, O., Bonfiglio, J.J., Colby, T., Zhang, Q., Atanassov, I., Zaja, R., Palazzo, L., Stockum, A., Ahel, I., and Matic, I. (2016). Serine is a new target residue for endogenous ADP-ribosylation on histones. Nat. Chem. Biol. *12*, 998–1000. https://doi.org/10.1038/nchembio.2180.

Leppek, K., Das, R., and Barna, M. (2018). Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat. Rev. Mol. Cell Biol. *19*, 158–174. https://doi.org/10.1038/nrm.2017.103.

Leslie Pedrioli, D.M., Leutert, M., Bilan, V., Nowak, K., Gunasekera, K., Ferrari, E., Imhof, R., Malmström, L., and Hottiger, M.O. (2018). Comprehensive ADP-ribosylome analysis identifies tyrosine as an ADP-ribose acceptor site. EMBO Rep. *19*, e45310. https://doi.org/10.15252/embr.201745310.

Leung, A.K.L. (2014). Poly(ADP-ribose): an organizer of cellular architecture. J. Cell Biol. *205*, 613–619. https://doi.org/10.1083/jcb.201402114.

Leutert, M., Pedrioli, D.M.L., and Hottiger, M.O. (2016). Identification of PARP-Specific ADP-Ribosylation Targets Reveals a Regulatory Function for ADP-Ribosylation in Transcription Elongation. Mol. Cell *63*, 181–183. https://doi.org/10.1016/j.molcel.2016.07.006.

Li, C., Fang, R., Sun, Y., Han, X., Li, F., Gao, B., Iafrate, A.J., Liu, X.-Y., Pao, W., Chen, H., et al. (2011). Spectrum of oncogenic driver mutations in lung adenocarcinomas from East Asian never smokers. PloS One *6*, e28204. https://doi.org/10.1371/journal.pone.0028204.

Li, S., Choi, Y.-L., Gong, Z., Liu, X., Lira, M., Kan, Z., Oh, E., Wang, J., Ting, J.C., Ye, X., et al. (2016). Comprehensive Characterization of Oncogenic Drivers in Asian Lung Adenocarcinoma. J. Thorac. Oncol. *11*, 2129–2140. https://doi.org/10.1016/j.jtho.2016.08.142.

Lin, K.-T., and Krainer, A.R. (2019). PSI-Sigma: a comprehensive splicing-detection method for short-read and long-read RNA-seq analysis. Bioinforma. Oxf. Engl. *35*, 5048–5054. https://doi.org/10.1093/bioinformatics/btz438.

Liu, Q., Yu, S., Zhao, W., Qin, S., Chu, Q., and Wu, K. (2018). EGFR-TKIs resistance via EGFR-independent signaling pathways. Mol. Cancer *17*, 53. https://doi.org/10.1186/s12943-018-0793-1.

Liu, Y., Snow, B.E., Kickhoefer, V.A., Erdmann, N., Zhou, W., Wakeham, A., Gomez, M., Rome, L.H., and Harrington, L. (2004). Vault poly(ADP-ribose) polymerase is associated with mammalian telomerase and is dispensable for telomerase function and vault structure in vivo. Mol. Cell. Biol. *24*, 5314–5323. https://doi.org/10.1128/MCB.24.12.5314-5323.2004.

Liu, Y., Liu, T., Sun, Q., Niu, M., Jiang, Y., and Pang, D. (2015). Downregulation of Ras GTPase-activating protein 1 is associated with poor survival of breast invasive ductal carcinoma patients. Oncol. Rep. *33*, 119–124. https://doi.org/10.3892/or.2014.3604.

Liu, Z., Zhang, W., Phillips, J.B., Arora, R., McClellan, S., Li, J., Kim, J.-H., Sobol, R.W., and Tan, M. (2019). Immunoregulatory protein B7-H3 regulates cancer stem cell enrichment and drug resistance through MVP-mediated MEK activation. Oncogene *38*, 88–102. https://doi.org/10.1038/s41388-018-0407-9.

Llères, D., Denegri, M., Biggiogera, M., Ajuh, P., and Lamond, A.I. (2010). Direct interaction between hnRNP-M and CDC5L/PLRG1 proteins affects alternative splice site choice. EMBO Rep. *11*, 445–451. https://doi.org/10.1038/embor.2010.64.

Long, N.P., Lee, W.J., Huy, N.T., Lee, S.J., Park, J.H., and Kwon, S.W. (2016). Novel Biomarker Candidates for Colorectal Cancer Metastasis: A Meta-analysis of In Vitro Studies. Cancer Inform. *15*, 11–17. https://doi.org/10.4137/CIN.S40301.

Luisier, R., Tyzack, G.E., Hall, C.E., Mitchell, J.S., Devine, H., Taha, D.M., Malik, B., Meyer, I., Greensmith, L., Newcombe, J., et al. (2018). Intron retention and nuclear loss of SFPQ are molecular hallmarks of ALS. Nat. Commun. *9*, 2010. https://doi.org/10.1038/s41467-018-04373-8.

Luo, P., Ding, Y., Lei, X., and Wu, F.-X. (2019). deepDriver: Predicting Cancer Driver Genes Based on Somatic Mutations Using Deep Convolutional Neural Networks. Front. Genet. *10*, 13. https://doi.org/10.3389/fgene.2019.00013.

Luo, W., Tian, P., Wang, Y., Xu, H., Chen, L., Tang, C., Shu, Y., Zhang, S., Wang, Z., Zhang, J., et al. (2018). Characteristics of genomic alterations of lung adenocarcinoma in young never-smokers. Int. J. Cancer *143*, 1696–1705. https://doi.org/10.1002/ijc.31542.

Luteijn, R.D., Zaver, S.A., Gowen, B.G., Wyman, S.K., Garelis, N.E., Onia, L., McWhirter, S.M., Katibah, G.E., Corn, J.E., Woodward, J.J., et al. (2019). SLC19A1 transports immunoreactive cyclic dinucleotides. Nature *573*, 434–438. https://doi.org/10.1038/s41586-019-1553-0.

Lynch, T.J., Bell, D.W., Sordella, R., Gurubhagavatula, S., Okimoto, R.A., Brannigan, B.W., Harris, P.L., Haserlat, S.M., Supko, J.G., Haluska, F.G., et al. (2004). Activating mutations in the epidermal growth factor receptor underlying

responsiveness of non-small-cell lung cancer to gefitinib. N. Engl. J. Med. *350*, 2129–2139. https://doi.org/10.1056/NEJMoa040938.

Mangerich, A., and Altmeyer, M. (2016). Identifying ADP-ribosylation targets by chemical genetics. Transl. Cancer Res. *5*. https://doi.org/10.21037/10422.

Marabti, E.E., and Abdel-Wahab, O. (2021). Therapeutic Modulation of RNA Splicing in Malignant and Non-Malignant Disease. Trends Mol. Med. *27*, 643–659. https://doi.org/10.1016/j.molmed.2021.04.005.

Marko, M., Leichter, M., Patrinou-Georgoula, M., and Guialis, A. (2010). hnRNP M interacts with PSF and p54(nrb) and co-localizes within defined nuclear structures. Exp. Cell Res. *316*, 390–400. https://doi.org/10.1016/j.yexcr.2009.10.021.

Marsischky, G.T., Wilson, B.A., and Collier, R.J. (1995). Role of glutamic acid 988 of human poly-ADP-ribose polymerase in polymer formation. Evidence for active site similarities to the ADP-ribosylating toxins. J. Biol. Chem. *270*, 3247–3254. https://doi.org/10.1074/jbc.270.7.3247.

Martello, R., Leutert, M., Jungmichel, S., Bilan, V., Larsen, S.C., Young, C., Hottiger, M.O., and Nielsen, M.L. (2016). Proteome-wide identification of the endogenous ADP-ribosylome of mammalian cells and tissue. Nat. Commun. *7*, 12917. https://doi.org/10.1038/ncomms12917.

Martin, M., Maßhöfer, L., Temming, P., Rahmann, S., Metz, C., Bornfeld, N., van de Nes, J., Klein-Hitpass, L., Hinnebusch, A.G., Horsthemke, B., et al. (2013). Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. Nat. Genet. *45*, 933–936. https://doi.org/10.1038/ng.2674.

Meijer, G.A., Schroeijers, A.B., Flens, M.J., Meuwissen, S.G., van der Valk, P., Baak, J.P., and Scheper, R.J. (1999). Increased expression of multidrug resistance related proteins Pgp, MRP1, and LRP/MVP occurs early in colorectal carcinogenesis. J. Clin. Pathol. *52*, 450–454. https://doi.org/10.1136/jcp.52.6.450.

Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41. https://doi.org/10.1186/gb-2011-12-4-r41.

Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J.J.-L., Bomane, A., Cosson, B., Eyras, E., Rasko, J.E.J., et al. (2017). IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biol. *18*, 51. https://doi.org/10.1186/s13059-017-1184-4.

Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002). Untranslated regions of mRNAs. Genome Biol. *3*, reviews0004.1. https://doi.org/10.1186/gb-2002-3-3-reviews0004.

Miller, M.L., Reznik, E., Gauthier, N.P., Aksoy, B.A., Korkut, A., Gao, J., Ciriello, G., Schultz, N., and Sander, C. (2015). Pan-Cancer Analysis of Mutation Hotspots in Protein Domains. Cell Syst. *1*, 197–209. https://doi.org/10.1016/j.cels.2015.08.014.

Mossink, M.H., Zon, A. van, Fränzel-Luiten, E., Schoester, M., Kickhoefer, V.A., Scheffer, G.L., Scheper, R.J., Sonneveld, P., and Wiemer, E.A.C. (2002). Disruption of the Murine Major Vault Protein (MVP/LRP) Gene Does Not Induce Hypersensitivity to Cytostatics. Cancer Res. *62*, 7298–7304. .

Muir, B., and Nunney, L. (2015). The expression of tumour suppressors and proto-oncogenes in tissues susceptible to their hereditary cancers. Br. J. Cancer *113*, 345–353. https://doi.org/10.1038/bjc.2015.205.

Munnur, D., Bartlett, E., Mikolčević, P., Kirby, I.T., Rack, J.G.M., Mikoč, A., Cohen, M.S., and Ahel, I. (2019). Reversible ADP-ribosylation of RNA. Nucleic Acids Res. *47*, 5658–5669. https://doi.org/10.1093/nar/gkz305.

Nam, S.W., Park, K.C., Choi, H.S., Lee, B., and Kim, S.-W. (2014). Identification of Zinc Finger, MYM-type 2 (ZMYM2) as a regulator of sorafenib resistance in hepatocellular carcinoma cell lines. J. Gastroenterol. Hepatol. *29*, 633–639. .

Nguyen Ba, A.N., Pogoutse, A., Provart, N., and Moses, A.M. (2009). NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. BMC Bioinformatics *10*, 202. https://doi.org/10.1186/1471-2105-10-202.

Ninomiya, K., Adachi, S., Natsume, T., Iwakiri, J., Terai, G., Asai, K., and Hirose, T. (2020). LncRNA-dependent nuclear stress bodies promote intron retention through SR protein phosphorylation. EMBO J. *39*, e102729. https://doi.org/10.15252/embj.2019102729.

Ohashi, R., Schraml, P., Batavia, A., Angori, S., Simmler, P., Rupp, N., Ajioka, Y., Oliva, E., and Moch, H. (2019). Allele Loss and Reduced Expression of CYCLOPS Genes is a Characteristic Feature of Chromophobe Renal Cell Carcinoma. Transl. Oncol. *12*, 1131–1137. https://doi.org/10.1016/j.tranon.2019.05.005.

Olivier, M., Hollstein, M., and Hainaut, P. (2010). TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. Cold Spring Harb. Perspect. Biol. *2*, a001008. https://doi.org/10.1101/cshperspect.a001008.

Olopade, O.I., Adeyanju, M.O., Safa, A.R., Hagos, F., Mick, R., Thompson, C.B., and Recant, W.M. (1997). Overexpression of BCL-x protein in primary breast cancer is associated with high tumor grade and nodal metastases. Cancer J. Sci. Am. *3*, 230–237. .

Osathanon, T., Nowwarote, N., and Pavasant, P. (2016). Expression and influence of Notch signaling in oral squamous cell carcinoma. J. Oral Sci. *58*, 283–294. https://doi.org/10.2334/josnusd.15-0535.

Palombo, R., Verdile, V., and Paronetto, M.P. (2020). Poison-Exon Inclusion in DHX9 Reduces Its Expression and Sensitizes Ewing Sarcoma Cells to Chemotherapeutic Treatment. Cells *9*, 328. https://doi.org/10.3390/cells9020328.

Pao, W., and Chmielecki, J. (2010). Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. Nat. Rev. Cancer *10*, 760–774. https://doi.org/10.1038/nrc2947.

Pao, W., and Girard, N. (2011). New driver mutations in non-small-cell lung cancer. Lancet Oncol. *12*, 175–180. https://doi.org/10.1016/S1470-2045(10)70087-5.

Pao, W., Miller, V., Zakowski, M., Doherty, J., Politi, K., Sarkaria, I., Singh, B., Heelan, R., Rusch, V., Fulton, L., et al. (2004). EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc. Natl. Acad. Sci. U. S. A. *101*, 13306–13311. https://doi.org/10.1073/pnas.0405220101.

Paolella, B.R., Gibson, W.J., Urbanski, L.M., Alberta, J.A., Zack, T.I., Bandopadhayay, P., Nichols, C.A., Agarwalla, P.K., Brown, M.S., Lamothe, R., et al. (2017). Copy-number and gene dependency analysis reveals partial copy loss of wild-type SF3B1 as a novel cancer vulnerability. ELife *6*, e23268. https://doi.org/10.7554/eLife.23268.

Park, S.-J., Yoon, B.-H., Kim, S.-K., and Kim, S.-Y. (2019). GENT2: an updated gene expression database for normal and tumor tissues. BMC Med. Genomics *12*, 101. https://doi.org/10.1186/s12920-019-0514-7.

Passacantilli, I., Frisone, P., De Paola, E., Fidaleo, M., and Paronetto, M.P. (2017). hnRNPM guides an alternative splicing program in response to inhibition of the PI3K/AKT/mTOR pathway in Ewing sarcoma cells. Nucleic Acids Res. *45*, 12270–12284. https://doi.org/10.1093/nar/gkx831.

Persson, H., Kvist, A., Vallon-Christersson, J., Medstrand, P., Borg, A., and Rovira, C. (2009). The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. Nat. Cell Biol. *11*, 1268–1271. https://doi.org/10.1038/ncb1972.

Pinnola, A., Naumova, N., Shah, M., and Tulin, A.V. (2007). Nucleosomal Core Histones Mediate Dynamic Regulation of Poly(ADP-ribose) Polymerase 1 Protein Binding to Chromatin and Induction of Its Enzymatic Activity *. J. Biol. Chem. *282*, 32511–32519. https://doi.org/10.1074/jbc.M705989200.

Pon, J.R., and Marra, M.A. (2015). Driver and Passenger Mutations in Cancer. Annu. Rev. Pathol. Mech. Dis. *10*, 25–50. https://doi.org/10.1146/annurev-pathol-012414-040312.

Prawira, A., Munusamy, P., Yuan, J., Chan, C.H.T., Koh, G.L., Shuen, T.W.H., Hu, J., Yap, Y.S., Tan, M.H., Ang, P., et al. (2019). Assessment of PARP4 as a candidate breast cancer susceptibility gene. Breast Cancer Res. Treat. *177*, 145–153. https://doi.org/10.1007/s10549-019-05286-w.

Prior, I.A., Lewis, P.D., and Mattos, C. (2012). A comprehensive survey of Ras mutations in cancer. Cancer Res. *72*, 2457–2467. https://doi.org/10.1158/0008-5472.CAN-11-2612.

Pütz, S.M., Vogiatzi, F., Stiewe, T., and Sickmann, A. (2010). Malignant transformation in a defined genetic background: proteome changes displayed by 2D-PAGE. Mol. Cancer *9*, 254. https://doi.org/10.1186/1476-4598-9-254.

Rack, J.G.M., Palazzo, L., and Ahel, I. (2020). (ADP-ribosyl)hydrolases: structure, function, and biology. Genes Dev. *34*, 263–284. https://doi.org/10.1101/gad.334631.119.

Radivojac, P., Baenziger, P.H., Kann, M.G., Mort, M.E., Hahn, M.W., and Mooney, S.D. (2008). Gain and loss of phosphorylation sites in human cancer. Bioinformatics *24*, i241–i247. https://doi.org/10.1093/bioinformatics/btn267.

Ramazzotti, D., Lal, A., Wang, B., Batzoglou, S., and Sidow, A. (2018). Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. Nat. Commun. *9*, 4453. https://doi.org/10.1038/s41467-018-06921-8.

Raval-Fernandes, S., Kickhoefer, V.A., Kitchen, C., and Rome, L.H. (2005). Increased susceptibility of vault poly(ADP-ribose) polymerase-deficient mice to carcinogen-induced tumorigenesis. Cancer Res. *65*, 8846–8852. https://doi.org/10.1158/0008-5472.CAN-05-0770.

Ray Chaudhuri, A., and Nussenzweig, A. (2017). The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. Nat. Rev. Mol. Cell Biol. *18*, 610–621. https://doi.org/10.1038/nrm.2017.53.

Ridge, C.A., McErlean, A.M., and Ginsberg, M.S. (2013). Epidemiology of Lung Cancer. Semin. Interv. Radiol. *30*, 93–98. https://doi.org/10.1055/s-0033-1342949.

Ritchie, C., Cordova, A.F., Hess, G.T., Bassik, M.C., and Li, L. (2019). SLC19A1 Is an Importer of the Immunotransmitter cGAMP. Mol. Cell *75*, 372-381.e5. https://doi.org/10.1016/j.molcel.2019.05.006.

Rodriguez, K.M., Buch-Larsen, S.C., Kirby, I.T., Siordia, I.R., Hutin, D., Rasmussen, M., Grant, D.M., David, L.L., Matthews, J., Nielsen, M.L., et al. (2021). Chemical genetics and proteome-wide site mapping reveal cysteine MARylation by PARP-7 on immune-relevant protein targets. ELife *10*, e60480. https://doi.org/10.7554/eLife.60480.

Sainz de Aja, J., Dost, A.F.M., and Kim, C.F. (2021). Alveolar progenitor cells and the origin of lung cancer. J. Intern. Med. *289*, 629–635. https://doi.org/10.1111/joim.13201.

Sajini, A.A., Choudhury, N.R., Wagner, R.E., Bornelöv, S., Selmi, T., Spanos, C., Dietmann, S., Rappsilber, J., Michlewski, G., and Frye, M. (2019). Loss of 5-methylcytosine alters the biogenesis of vault-derived small RNAs to coordinate epidermal differentiation. Nat. Commun. *10*, 2550. https://doi.org/10.1038/s41467-019-10020-7.

Scarpa, E.S., Fabrizio, G., and Di Girolamo, M. (2013). A role of intracellular mono-ADP-ribosylation in cancer biology. FEBS J. *280*, 3551–3562. https://doi.org/10.1111/febs.12290.

Scheffer, G.L., Wijngaard, P.L., Flens, M.J., Izquierdo, M.A., Slovak, M.L., Pinedo, H.M., Meijer, C.J., Clevers, H.C., and Scheper, R.J. (1995). The drug resistance-related protein LRP is the human major vault protein. Nat. Med. *1*, 578–582. https://doi.org/10.1038/nm0695-578.

Scheper, R.J., Broxterman, H.J., Scheffer, G.L., Kaaijk, P., Dalton, W.S., van Heijningen, T.H., van Kalken, C.K., Slovak, M.L., de Vries, E.G., and van der Valk, P. (1993). Overexpression of a M(r) 110,000 vesicular protein in non-P-glycoprotein-mediated multidrug resistance. Cancer Res. *53*, 1475–1479. .

Schreiber, V., Dantzer, F., Ame, J.-C., and de Murcia, G. (2006). Poly(ADP-ribose): novel functions for an old molecule. Nat. Rev. Mol. Cell Biol. *7*, 517–528. https://doi.org/10.1038/nrm1963.

Schroeijers, A.B., Siva, A.C., Scheffer, G.L., de Jong, M.C., Bolick, S.C., Dukers, D.F., Slootstra, J.W., Meloen, R.H., Wiemer, E., Kickhoefer, V.A., et al. (2000). The Mr 193,000 vault protein is up-regulated in multidrug-resistant cancer cell lines. Cancer Res. *60*, 1104–1110. .

Sequist, L.V., Joshi, V.A., Jänne, P.A., Muzikansky, A., Fidias, P., Meyerson, M., Haber, D.A., Kucherlapati, R., Johnson, B.E., and Lynch, T.J. (2007). Response to treatment and survival of patients with non-small cell lung cancer undergoing somatic EGFR mutation testing. The Oncologist *12*, 90–98. .

Sharma, S.V., Bell, D.W., Settleman, J., and Haber, D.A. (2007). Epidermal growth factor receptor mutations in lung cancer. Nat. Rev. Cancer *7*, 169–181. https://doi.org/10.1038/nrc2088.

Shen, S., Park, J.W., Lu, Z., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. U. S. A. *111*, E5593-5601. https://doi.org/10.1073/pnas.1419161111.

Shi, Y., Au, J.S.-K., Thongprasert, S., Srinivasan, S., Tsai, C.-M., Khoa, M.T., Heeroma, K., Itoh, Y., Cornelio, G., and Yang, P.-C. (2014). A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer *9*, 154–162. https://doi.org/10.1097/JTO.0000000000000033.

Shigematsu, H., Lin, L., Takahashi, T., Nomura, M., Suzuki, M., Wistuba, I.I., Fong, K.M., Lee, H., Toyooka, S., Shimizu, N., et al. (2005). Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J. Natl. Cancer Inst. *97*, 339–346. https://doi.org/10.1093/jnci/dji055.

Shiroki, T., Yokoyama, M., Tanuma, N., Maejima, R., Tamai, K., Yamaguchi, K., Oikawa, T., Noguchi, T., Miura, K., Fujiya, T., et al. (2017). Enhanced expression of the M2 isoform of pyruvate kinase is involved in gastric cancer development by regulating cancer-specific metabolism. Cancer Sci. *108*, 931–940. https://doi.org/10.1111/cas.13211.

Siddiqi, S., Siddiqi, S.A., and Mansbach, C.M. (2010). Sec24C is required for docking the prechylomicron transport vesicle with the Golgi. J. Lipid Res. *51*, 1093–1100. https://doi.org/10.1194/jlr.M002758.

Siegel, R.L., Miller, K.D., and Jemal, A. (2018). Cancer statistics, 2018. CA. Cancer J. Clin. *68*, 7–30. https://doi.org/10.3322/caac.21442.

Siegelin, M.D., and Borczuk, A.C. (2014). Epidermal growth factor receptor mutations in lung adenocarcinoma. Lab. Invest. *94*, 129–137. https://doi.org/10.1038/labinvest.2013.147.

Slesina, M., Inman, E.M., Moore, A.E., Goldhaber, J.I., Rome, L.H., and Volknandt, W. (2006). Movement of vault particles visualized by GFP-tagged major vault protein. Cell Tissue Res. *324*, 403–410. https://doi.org/10.1007/s00441-006-0158-8.

Smith, C.L., Migliaccio, I., Chaubal, V., Wu, M.-F., Pace, M.C., Hartmaier, R., Jiang, S., Edwards, D.P., Gutiérrez, M.C., Hilsenbeck, S.G., et al. (2012). Elevated nuclear expression of the SMRT corepressor in breast cancer is

associated with earlier tumor recurrence. Breast Cancer Res. Treat. *136*, 253–265. https://doi.org/10.1007/s10549-012-2262-7.

Stein, U., Bergmann, S., Scheffer, G.L., Scheper, R.J., Royer, H.-D., Schlag, P.M., and Walther, W. (2005). YB-1 facilitates basal and 5-fluorouracil-inducible expression of the human major vault protein (MVP) gene. Oncogene *24*, 3606–3618. https://doi.org/10.1038/sj.onc.1208386.

Stephen, A.G., Raval-Fernandes, S., Huynh, T., Torres, M., Kickhoefer, V.A., and Rome, L.H. (2001). Assembly of vault-like particles in insect cells expressing only the major vault protein. J. Biol. Chem. *276*, 23217–23220. https://doi.org/10.1074/jbc.C100226200.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719–724. https://doi.org/10.1038/nature07943.

Sulima, S.O., Hofman, I.J.F., Keersmaecker, K.D., and Dinman, J.D. (2017). How Ribosomes Translate Cancer. Cancer Discov. *7*, 1069–1087. https://doi.org/10.1158/2159-8290.CD-17-0550.

Sun, S., Schiller, J.H., and Gazdar, A.F. (2007). Lung cancer in never smokers--a different disease. Nat. Rev. Cancer *7*, 778–790. https://doi.org/10.1038/nrc2190.

Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T., and Lehner, B. (2014). Synonymous mutations frequently act as driver mutations in human cancers. Cell *156*, 1324–1335. https://doi.org/10.1016/j.cell.2014.01.051.

Sutherland, K.D., and Berns, A. (2010). Cell of origin of lung cancer. Mol. Oncol. *4*, 397–403. https://doi.org/10.1016/j.molonc.2010.05.002.

Sutherland, K.D., Song, J.-Y., Kwon, M.C., Proost, N., Zevenhoven, J., and Berns, A. (2014). Multiple cells-of-origin of mutant K-Ras–induced mouse lung adenocarcinoma. Proc. Natl. Acad. Sci. *111*, 4952–4957. https://doi.org/10.1073/pnas.1319963111.

Swanson, M.S., and Dreyfuss, G. (1988). RNA binding specificity of hnRNP proteins: a subset bind to the 3’ end of introns. EMBO J. *7*, 3519–3529. .

Sweeney, N.P., and Vink, C.A. (2021). The impact of lentiviral vector genome size and producer cell genomic to gag-pol mRNA ratios on packaging efficiency and titre. Mol. Ther. - Methods Clin. Dev. *21*, 574–584. https://doi.org/10.1016/j.omtm.2021.04.007.

Takehara, T., Liu, X., Fujimoto, J., Friedman, S.L., and Takahashi, H. (2001). Expression and role of Bcl-xL in human hepatocellular carcinomas. Hepatol. Baltim. Md *34*, 55–61. https://doi.org/10.1053/jhep.2001.25387.

Tamborero, D., Gonzalez-Perez, A., and Lopez-Bigas, N. (2013). OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. Bioinforma. Oxf. Engl. *29*, 2238–2244. https://doi.org/10.1093/bioinformatics/btt395.

Tanaka, H., Kato, K., Yamashita, E., Sumizawa, T., Zhou, Y., Yao, M., Iwasaki, K., Yoshimura, M., and Tsukihara, T. (2009). The Structure of Rat Liver Vault at 3.5 Angstrom Resolution. Science *323*, 384–388. https://doi.org/10.1126/science.1164975.

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. *45*, W98–W102. https://doi.org/10.1093/nar/gkx247.

Tarpey, P.S., Behjati, S., Young, M.D., Martincorena, I., Alexandrov, L.B., Farndon, S.J., Guzzo, C., Hardy, C., Latimer, C., Butler, A.P., et al. (2017). The driver landscape of sporadic chordoma. Nat. Commun. *8*, 890. https://doi.org/10.1038/s41467-017-01026-0.

Teng, S., Wang, L., Srivastava, A.K., Schwartz, C.E., and Alexov, E. (2010). Structural Assessment of the Effects of Amino Acid Substitutions on Protein Stability and Protein-Protein Interaction. Int. J. Comput. Biol. Drug Des. *3*, 334–349. https://doi.org/10.1504/IJCBDD.2010.038396.

The Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543–550. https://doi.org/10.1038/nature13385.

Tokheim, C.J., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., and Karchin, R. (2016). Evaluating the evaluation of cancer driver genes. Proc. Natl. Acad. Sci. *113*, 14330–14335. https://doi.org/10.1073/pnas.1616440113.

Tomasello, C., Baldessari, C., Napolitano, M., Orsi, G., Grizzi, G., Bertolini, F., Barbieri, F., and Cascinu, S. (2018). Resistance to EGFR inhibitors in non-small cell lung cancer: Clinical management and future perspectives. Crit. Rev. Oncol. Hematol. *123*, 149–161. https://doi.org/10.1016/j.critrevonc.2018.01.013.

Tseng, C.-H., Tsuang, B.-J., Chiang, C.-J., Ku, K.-C., Tseng, J.-S., Yang, T.-Y., Hsu, K.-H., Chen, K.-C., Yu, S.-L., Lee, W.-C., et al. (2019). The Relationship Between Air Pollution and Lung Cancer in Nonsmokers in Taiwan. J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer *14*, 784–792. https://doi.org/10.1016/j.jtho.2018.12.033.

Unni, A.M., Lockwood, W.W., Zejnullahu, K., Lee-Lin, S.-Q., and Varmus, H. (2015). Evidence that synthetic lethality underlies the mutual exclusivity of oncogenic KRAS and EGFR mutations in lung adenocarcinoma. ELife *4*, e06907. https://doi.org/10.7554/eLife.06907.

Venables, J.P. (2004). Aberrant and Alternative Splicing in Cancer. Cancer Res. *64*, 7647–7654. https://doi.org/10.1158/0008-5472.CAN-04-1910.

Venables, J.P., Koh, C.-S., Froehlich, U., Lapointe, E., Couture, S., Inkel, L., Bramard, A., Paquet, E.R., Watier, V., Durand, M., et al. (2008). Multiple and specific mRNA processing targets for the major human hnRNP proteins. Mol. Cell. Biol. *28*, 6033–6043. https://doi.org/10.1128/MCB.00726-08.

Vivelo, C.A., Wat, R., Agrawal, C., Tee, H.Y., and Leung, A.K.L. (2017). ADPriboDB: The database of ADP-ribosylated proteins. Nucleic Acids Res. *45*, D204–D209. https://doi.org/10.1093/nar/gkw706.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer Genome Landscapes. Science *339*, 1546–1558. https://doi.org/10.1126/science.1235122.

Vyas, S., Chesarone-Cataldo, M., Todorova, T., Huang, Y.-H., and Chang, P. (2013). A systematic analysis of the PARP protein family identifies new functions critical for cell physiology. Nat. Commun. *4*, 2240. https://doi.org/10.1038/ncomms3240.

Vyas, S., Matic, I., Uchima, L., Rood, J., Zaja, R., Hay, R.T., Ahel, I., and Chang, P. (2014). Family-wide analysis of poly(ADP-ribose) polymerase activity. Nat. Commun. *5*, 4426. https://doi.org/10.1038/ncomms5426.

Wacker, D.A., Ruhl, D.D., Balagamwala, E.H., Hope, K.M., Zhang, T., and Kraus, W.L. (2007). The DNA binding and catalytic domains of poly(ADP-ribose) polymerase 1 cooperate in the regulation of chromatin structure and transcription. Mol. Cell. Biol. *27*, 7475–7485. https://doi.org/10.1128/MCB.01314-07.

Wagner, A.R., Scott, H.M., West, K.O., Vail, K.J., Fitzsimons, T.C., Coleman, A.K., Carter, K.E., Watson, R.O., and Patrick, K.L. (2021). Global Transcriptomics Uncovers Distinct Contributions From Splicing Regulatory Proteins to the Macrophage Innate Immune Response. Front. Immunol. *12*, 656885. https://doi.org/10.3389/fimmu.2021.656885.

Wal, L. van der, and Demmers, J.A.A. (2015). Quantitative Mass Spectrometry-based Proteomics (IntechOpen).

Wan, C., Hou, S., Ni, R., Lv, L., Ding, Z., Huang, X., Hang, Q., He, S., Wang, Y., Cheng, C., et al. (2015). MIF4G domain containing protein regulates cell cycle and hepatic carcinogenesis by antagonizing CDK2-dependent p27 stability. Oncogene *34*, 237–245. https://doi.org/10.1038/onc.2013.536.

Wang, Z., and Moult, J. (2001). SNPs, protein structure, and disease. Hum. Mutat. *17*, 263–270. https://doi.org/10.1002/humu.22.

Wang, B., Joo, J.H., Mount, R., Teubner, B.J.W., Krenzer, A., Ward, A.L., Ichhaporia, V.P., Adams, E.J., Khoriaty, R., Peters, S.T., et al. (2018a). The COPII cargo adapter SEC24C is essential for neuronal homeostasis. J. Clin. Invest. *128*, 3319–3332. https://doi.org/10.1172/JCI98194.

Wang, C., Yin, R., Dai, J., Gu, Y., Cui, S., Ma, H., Zhang, Z., Huang, J., Qin, N., Jiang, T., et al. (2018b). Whole-genome sequencing reveals genomic signatures associated with the inflammatory microenvironments in Chinese NSCLC patients. Nat. Commun. *9*, 2054. https://doi.org/10.1038/s41467-018-04492-2.

Wang, L.-H., Wu, C.-F., Rajasekaran, N., and Shin, Y.K. (2018c). Loss of Tumor Suppressor Gene Function in Human Cancer: An Overview. Cell. Physiol. Biochem. *51*, 2647–2693. https://doi.org/10.1159/000495956.

Wang, S., Cang, S., and Liu, D. (2016). Third-generation inhibitors targeting EGFR T790M mutation in advanced non-small cell lung cancer. J. Hematol. Oncol.J Hematol Oncol *9*. https://doi.org/10.1186/s13045-016-0268-z.

Wang, X., Li, J., Bian, X., Wu, C., Hua, J., Chang, S., Yu, T., Li, H., Li, Y., Hu, S., et al. (2021). CircURI1 interacts with hnRNPM to inhibit metastasis by modulating alternative splicing in gastric cancer. Proc. Natl. Acad. Sci. *118*. https://doi.org/10.1073/pnas.2012881118.

Watson, I.R., Takahashi, K., Futreal, P.A., and Chin, L. (2013). Emerging patterns of somatic mutations in cancer. Nat. Rev. Genet. *14*, 703–718. https://doi.org/10.1038/nrg3539.

Weaver, A., and Yang, E. (2013). Beyond DNA Repair: Additional Functions of PARP-1 in Cancer. Front. Oncol. *3*, 290. https://doi.org/10.3389/fonc.2013.00290.

Wee, P., and Wang, Z. (2017). Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. Cancers *9*, 52. https://doi.org/10.3390/cancers9050052.

Weir, B.A., Woo, M.S., Getz, G., Perner, S., Ding, L., Beroukhim, R., Lin, W.M., Province, M.A., Kraja, A., Johnson, L.A., et al. (2007). Characterizing the cancer genome in lung adenocarcinoma. Nature *450*, 893–898. https://doi.org/10.1038/nature06358.

West, K.O., Scott, H.M., Torres-Odio, S., West, A.P., Patrick, K.L., and Watson, R.O. (2019). The Splicing Factor hnRNP M Is a Critical Regulator of Innate Immune Gene Expression in Macrophages. Cell Rep. *29*, 1594-1609.e5. https://doi.org/10.1016/j.celrep.2019.09.078.

Wu, K., Zhang, X., Li, F., Xiao, D., Hou, Y., Zhu, S., Liu, D., Ye, X., Ye, M., Yang, J., et al. (2015). Frequent alterations in cytoskeleton remodelling genes in

primary and metastatic lung adenocarcinomas. Nat. Commun. *6*, 10131. https://doi.org/10.1038/ncomms10131.

Wu, L.C., Wang, Z.W., Tsan, J.T., Spillman, M.A., Phung, A., Xu, X.L., Yang, M.C., Hwang, L.Y., Bowcock, A.M., and Baer, R. (1996). Identification of a RING protein that can interact in vivo with the BRCA1 gene product. Nat. Genet. *14*, 430–440. https://doi.org/10.1038/ng1296-430.

Xia, W., Nagase, S., Montia, A.G., Kalachikov, S.M., Keniry, M., Su, T., Memeo, L., Hibshoosh, H., and Parsons, R. (2008). BAF180 is a critical regulator of p21 induction and a tumor suppressor mutated in breast cancer. Cancer Res. *68*, 1667–1674. https://doi.org/10.1158/0008-5472.CAN-07-5276.

Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene Set Knowledge Discovery with Enrichr. Curr. Protoc. *1*, e90. https://doi.org/10.1002/cpz1.90.

Xu, X., Rock, J.R., Lu, Y., Futtner, C., Schwab, B., Guinney, J., Hogan, B.L.M., and Onaitis, M.W. (2012). Evidence for type II cells as cells of origin of K-Ras–induced distal lung adenocarcinoma. Proc. Natl. Acad. Sci. *109*, 4910–4915. https://doi.org/10.1073/pnas.1112499109.

Xu, Y., Gao, X.D., Lee, J.-H., Huang, H., Tan, H., Ahn, J., Reinke, L.M., Peter, M.E., Feng, Y., Gius, D., et al. (2014). Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. Genes Dev. *28*, 1191–1203. https://doi.org/10.1101/gad.241968.114.

Yang, P. (2009). Epidemiology of lung cancer prognosis: quantity and quality of life. Methods Mol. Biol. Clifton NJ *471*, 469–486. https://doi.org/10.1007/978-1-59745-416-2_24.

Yang, C.-Y., Yang, J.C.-H., and Yang, P.-C. (2020). Precision Management of Advanced Non-Small Cell Lung Cancer. Annu. Rev. Med. *71*, 117–136. https://doi.org/10.1146/annurev-med-051718-013524.

Yang, T., An, Z., Zhang, C., Wang, Z., Wang, X., Liu, Y., Du, E., Liu, R., Zhang, Z., and Xu, Y. (2019). hnRNPM, a potential mediator of YY1 in promoting the epithelial-mesenchymal transition of prostate cancer cells. The Prostate *79*, 1199–1210. https://doi.org/10.1002/pros.23790.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64–69. https://doi.org/10.1038/nature10496.

Yu, K., Yau, Y.H., Sinha, A., Tan, T., Kickhoefer, V.A., Rome, L.H., Lee, H., Shochat, S.G., and Lim, S. (2017). Modulation of the Vault Protein-Protein

Interaction for Tuning of Molecular Release. Sci. Rep. *7*, 14816. https://doi.org/10.1038/s41598-017-12870-x.

Yu, Z., Fotouhi-Ardakani, N., Wu, L., Maoui, M., Wang, S., Banville, D., and Shen, S.-H. (2002). PTEN associates with the vault particles in HeLa cells. J. Biol. Chem. *277*, 40247–40252. https://doi.org/10.1074/jbc.M207608200.

Zappa, C., and Mousa, S.A. (2016). Non-small cell lung cancer: current treatment and future advances. Transl. Lung Cancer Res. *5*, 288–300. https://doi.org/10.21037/tlcr.2016.06.07.

Zarkovic, G., Belousova, E.A., Talhaoui, I., Saint-Pierre, C., Kutuzov, M.M., Matkarimov, B.T., Biard, D., Gasparutto, D., Lavrik, O.I., and Ishchenko, A.A. (2018). Characterization of DNA ADP-ribosyltransferase activities of PARP2 and PARP3: new insights into DNA ADP-ribosylation. Nucleic Acids Res. *46*, 2417–2431. https://doi.org/10.1093/nar/gkx1318.

Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L.D., and Ferretti, V. (2019a). The International Cancer Genome Consortium Data Portal. Nat. Biotechnol. *37*, 367–369. https://doi.org/10.1038/s41587-019-0055-9.

Zhang, T., Joubert, P., Ansari-Pour, N., Zhao, W., Hoang, P.H., Lokanga, R., Moye, A.L., Rosenbaum, J., Gonzalez-Perez, A., Martínez-Jiménez, F., et al. (2021). Genomic and evolutionary classification of lung cancer in never smokers. Nat. Genet. *53*, 1348–1359. https://doi.org/10.1038/s41588-021-00920-0.

Zhang, W., Neo, S.P., Gunaratne, J., Poulsen, A., Boping, L., Ong, E.H., Sangthongpitag, K., Pendharkar, V., Hill, J., and Cohen, S.M. (2015). Feedback regulation on PTEN/AKT pathway by the ER stress kinase PERK mediated by interaction with the Vault complex. Cell. Signal. *27*, 436–442. https://doi.org/10.1016/j.cellsig.2014.12.010.

Zhang, X.-C., Wang, J., Shao, G.-G., Wang, Q., Qu, X., Wang, B., Moy, C., Fan, Y., Albertyn, Z., Huang, X., et al. (2019b). Comprehensive genomic and immunological characterization of Chinese non-small cell lung cancer patients. Nat. Commun. *10*, 1772. https://doi.org/10.1038/s41467-019-09762-1.

Zhang, Y., Wang, J., Ding, M., and Yu, Y. (2013). Site-specific characterization of the Asp- and Glu-ADP-ribosylated proteome. Nat. Methods *10*, 981–984. https://doi.org/10.1038/nmeth.2603.

Zhao, M., and Zhao, Z. (2016). Concordance of copy number loss and down-regulation of tumor suppressor genes: a pan-cancer study. BMC Genomics *17*, 532. https://doi.org/10.1186/s12864-016-2904-y.

Zheng, C.-L., Sumizawa, T., Che, X.-F., Tsuyama, S., Furukawa, T., Haraguchi, M., Gao, H., Gotanda, T., Jueng, H.-C., Murata, F., et al. (2004). Characterization of MVP and VPARP assembly into vault ribonucleoprotein complexes. Biochem. Biophys. Res. Commun. *326*, 100–107. https://doi.org/10.1016/j.bbrc.2004.11.006.

Zhou, W., and Christiani, D.C. (2011). East meets West: ethnic differences in epidemiology and clinical behaviors of lung cancer between East Asians and Caucasians. Chin. J. Cancer *30*, 287–292. https://doi.org/10.5732/cjc.011.10106.

van Zon, A., Mossink, M.H., Schoester, M., Scheffer, G.L., Scheper, R.J., Sonneveld, P., and Wiemer, E.A.C. (2002). Structural domains of vault proteins: a role for the coiled coil domain in vault assembly. Biochem. Biophys. Res. Commun. *291*, 535–541. https://doi.org/10.1006/bbrc.2002.6472.

van Zon, A., Mossink, M.H., Scheper, R.J., Sonneveld, P., and Wiemer, E. a. C. (2003a). The vault complex. Cell. Mol. Life Sci. CMLS *60*, 1828–1837. https://doi.org/10.1007/s00018-003-3030-y.

van Zon, A., Mossink, M.H., Schoester, M., Houtsmuller, A.B., Scheffer, G.L., Scheper, R.J., Sonneveld, P., and Wiemer, E.A.C. (2003b). The formation of vault-tubes: a dynamic interaction between vaults and vault PARP. J. Cell Sci. *116*, 4391–4400. https://doi.org/10.1242/jcs.00749.

van Zon, A., Mossink, M.H., Houtsmuller, A.B., Schoester, M., Scheffer, G.L., Scheper, R.J., Sonneveld, P., and Wiemer, E.A.C. (2006). Vault mobility depends in part on microtubules and vaults can be recruited to the nuclear envelope. Exp. Cell Res. *312*, 245–255. https://doi.org/10.1016/j.yexcr.2005.10.016.